ADER and DeC: arbitrarily high order (explicit) methods for PDEs and ODEs

Davide Torlo

*MathLab, Mathematics Area, SISSA International School for Advanced Studies, Trieste, Italy
davidetorlo.it

Based on: Han Veiga, M., Öffner, P. & Torlo, D. *DeC and ADER: Similarities, Differences and a Unified Framework.* J Sci Comput 87, 2 (2021).
https://doi.org/10.1007/s10915-020-01397-5

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

Motivation: high order accurate explicit method

ARBITRARILY

Methods used to solve a hyperbolic PDE system for $u : \mathbb{R}^+ \times \Omega \to \mathbb{R}^D$

$$\boxed{\partial_t u + \nabla_{\mathbf{x}} \mathcal{F}(u) = 0.} \tag{1}$$

Or ODE system for $\boldsymbol{u} : \mathbb{R}^+ \to \mathbb{R}^S$

$$\boxed{\partial_t \boldsymbol{u} = F(\boldsymbol{u}).} \tag{2}$$

Applications:

- Fluids/transport
- Chemical/biological processes

How?

- Arbitrarily high order accurate
-

Motivation: high order accurate explicit method

Methods used to solve a hyperbolic PDE system for $u : \mathbb{R}^+ \times \Omega \to \mathbb{R}^D$

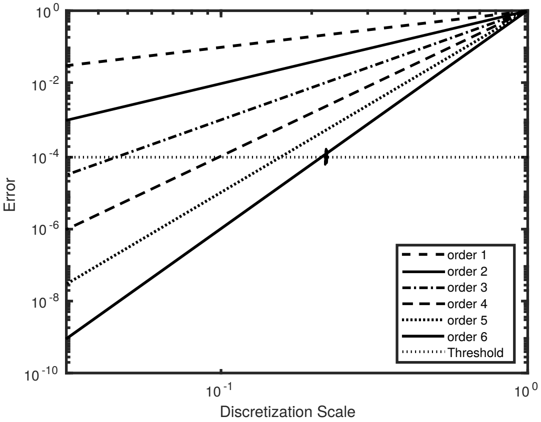
(1)

Or ODE system for $\boldsymbol{u}$:

(2)

Applications:

- Fluids/transport
- Chemical/biologica



How?

- Arbitrarily high orc
-

Motivation: high order accurate explicit method

Methods used to solve a hyperbolic PDE system for $u : \mathbb{R}^+ \times \Omega \to \mathbb{R}^D$

$$\partial_t u + \nabla_{\mathbf{x}} \mathcal{F}(u) = 0. \tag{1}$$

Or ODE system for $\boldsymbol{u} : \mathbb{R}^+ \to \mathbb{R}^S$

$$\partial_t \boldsymbol{u} = F(\boldsymbol{u}). \tag{2}$$

Applications:

- Fluids/transport
- Chemical/biological processes

How?

- Arbitrarily high order accurate
- Explicit (if nonstiff problem)

Classical time integration: Runge–Kutta

$$
\begin{cases}
\boldsymbol{u}^{(1)} := \boldsymbol{u}^n, & (3)\\
\boldsymbol{u}^{(k)} := \boldsymbol{u}^n + \overset{\Delta t}{\vee}\displaystyle\sum_{s=1}^{K} a_{ks} F\left(t^n + c_s \Delta t, \boldsymbol{u}^{(s)}\right), \quad \text{for } k = 2, \dots, K, & (4)\\
\boldsymbol{u}^{n+1} := \boldsymbol{u}^n + \overset{\Delta t}{\vee}\displaystyle\sum_{s=1}^{K} b_s F\left(t^n + c_s \Delta t, \boldsymbol{u}^{(s)}\right). & (5)
\end{cases}
$$

Classical time integration: Explicit Runge–Kutta

$$\boldsymbol{u}^{(k)} := \boldsymbol{u}^n + \sum_{s=1}^{k-1} a_{ks} F\left(t^n + c_s \Delta t, \boldsymbol{u}^{(s)}\right), \quad \text{for } k = 2, \dots, K.$$

- Easy to solve
- High orders involved:
  - Order conditions: system of many equations
  - Stages $K \geq d$ order of accuracy (e.g. RK44, RK65)

Classical time integration: Implicit Runge–Kutta

$$\boldsymbol{u}^{(k)} := \boldsymbol{u}^n + \sum_{s=1}^{K} a_{ks} F\left(t^n + c_s \Delta t, \boldsymbol{u}^{(s)}\right), \quad \text{for } k = 2, \dots, K.$$

- More complicated to solve for nonlinear systems
- High orders easily done:
  - Take a high order quadrature rule on $[t^n, t^{n+1}]$
  - Compute the coefficients accordingly, see Gauss–Legendre ($2S$) or Gauss–Lobatto ($2S-2$) polynomials
  - Order up to $d = 2K$

ADER and DeC

Two iterative explicit arbitrarily high order accurate methods.

- ADER[1] for hyperbolic PDE, after a first analytic more complicated approach.
- Deferred Correction (DeC): introduced for explicit ODE[2], extended to implicit ODE[3] and to hyperbolic PDE[4].

[1] M. Dumbser, D. S. Balsara, E. F. Toro, and C.-D. Munz. A unified framework for the construction of one-step finite volume and discontinuous galerkin schemes on unstructured meshes. Journal of Computational Physics, 227(18):8209–8253, 2008.

[2] A. Dutt, L. Greengard, and V. Rokhlin. Spectral Deferred Correction Methods for Ordinary Differential Equations. BIT Numerical Mathematics, 40(2):241–266, 2000.

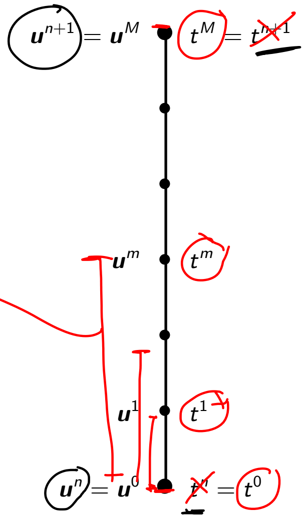
[3] M. L. Minion. Semi-implicit spectral deferred correction methods for ordinary differential equations. Commun. Math. Sci., 1(3):471–500, 09 2003.

[4] R. Abgrall. High order schemes for hyperbolic problems using globally continuous approximation and avoiding mass matrices. Journal of Scientific Computing, 73(2):461–494, Dec 2017.

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

DeC high order time discretization: $\mathcal{L}^2$



High order in time: we discretize our variable on $[t^n, t^{n+1}]$ in M substeps ($\boldsymbol{u}^m$).

$$\boxed{\partial_t \boldsymbol{u} = F(\boldsymbol{u}(t)).}$$

Thanks to Picard–Lindelöf theorem, we can rewrite

INTEGRAL FORM

$$\boldsymbol{u}^m = \boldsymbol{u}^0 + \int_{t^0}^{t^m} F(\boldsymbol{u}(t))dt.$$

and if we want to reach order $r+1$ we need $M = r$.

- EQUIDISTANT POINTS
- GAUSS-LOBATTO

DeC high order time discretization: $\mathcal{L}^2$

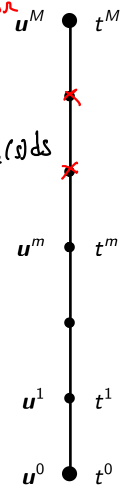
φ_r LAGRANGIAN POLYS

$u^m = u^0 + \int_{t^0}^{t^m} F(u(t))\,dt$ $\qquad \varphi_r(t^m) = \delta_{mr}$

More precisely, we want to solve $\mathcal{L}^2(\boldsymbol{u}^{n,0},\dots,\boldsymbol{u}^{n,M})=0$, where

$$\mathcal{L}^2(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) = \begin{pmatrix} \boldsymbol{u}^M - \boldsymbol{u}^0 + \sum_{r=0}^M \int_{t^0}^{t^M} F(\boldsymbol{u}^r)\underline{\varphi_r(s)\mathrm{d}s} \\ \vdots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 + \sum_{r=0}^M \int_{t^0}^{t^1} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \end{pmatrix}$$

$\theta_r^m = \int_0^{t^m} \varphi_r(s)\,ds$



$U \in \mathbb{R}^S$

- $\mathcal{L}^2 = 0$ is a system of $\boxed{M}\times\boxed{S}$ coupled (non)linear equations
- $\mathcal{L}^2$ is an implicit method (collocation method: Gauss, LobattoIIIA)
- Not easy to solve directly $\mathcal{L}^2(\underline{\boldsymbol{u}}^*)=0$
- High order ($\geq M+1$), depending on points distribution

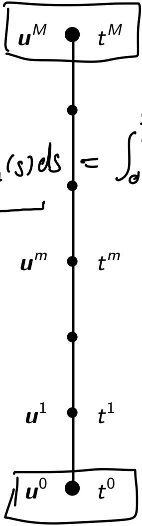
DeC high order time discretization: $\mathcal{L}^2$

More precisely, for each σ we want to solve $\mathcal{L}^2(\boldsymbol{u}^{n,0},\dots,\boldsymbol{u}^{n,M})=0$, where

$$\mathcal{L}^2(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M)=\begin{pmatrix}\boldsymbol{u}^M-\boldsymbol{u}^0+\Delta t\sum_{r=0}^M\theta_r^M F(\boldsymbol{u}^r)\\ \vdots\\ \boldsymbol{u}^1-\boldsymbol{u}^0+\Delta t\sum_{r=0}^M\theta_r^1 F(\boldsymbol{u}^r)\end{pmatrix}=0$$

$$\theta_r^m=\underbrace{\frac{1}{\Delta t}\int_{t^0}^{t^m}\varphi_r(s)\,ds}=\int_0^{\beta^m}\tilde{\varphi}_r(s)\,ds$$

$$u^m=u^0+\Delta t\sum\theta_r^m F(u^r)\quad\forall m$$



- $\mathcal{L}^2=0$ is a system of $\underline{M}\times\underline{S}$ coupled (non)linear equations
- $\mathcal{L}^2$ is an implicit method (collocation method: Gauss, LobattoIIIA)
- Not easy to solve directly $\underline{\mathcal{L}^2(\boldsymbol{u}^*)=0}$ $\quad u^*$ UNKNOWN
- High order ($\geq \underline{M+1}$), depending on points distribution

2M LOBATTO

DeC low order time discretization: $\mathcal{L}^1$

$\partial_t u = F(u)$ $\qquad$ $\partial_t u + F(u) = 0$

Instead of solving the implicit system directly (difficult), we introduce a first order scheme $\mathcal{L}^1(\boldsymbol{u}^{n,0},\dots,\boldsymbol{u}^{n,M})$:

EXPLICIT EULER

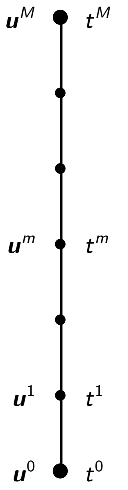
$$\mathcal{L}^1(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) = \begin{pmatrix} \boxed{\boldsymbol{u}^M - \boldsymbol{u}^0 + \Delta t \beta^M F(\boldsymbol{u}^0)} \\ \vdots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 + \Delta t \beta^1 F(\boldsymbol{u}^0) \end{pmatrix}$$

- First order approximation
- Explicit Euler
- Easy to solve $\mathcal{L}^1(\underline{\boldsymbol{u}}) = 0$

$\mathcal{L}^1(\underline{u}) = 0$

$\beta^m \Delta t = t^m - t^0$

$\beta^m = \dfrac{t^m - t^0}{\Delta t}$



β^m

Δt

Deferred Correction[5]

How to combine two methods keeping the accuracy of the second and the stability and simplicity of the first one?

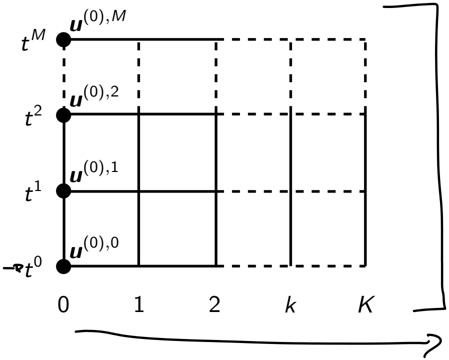
- $\mathcal{L}^1(\underline{\boldsymbol{u}}) = 0$, first order accuracy, easily invertible.
- $\mathcal{L}^2(\underline{\boldsymbol{u}}) = 0$, high order $M+1$.

$$\boxed{\boldsymbol{u}^{0,(k)} := \boldsymbol{u}(t^n),} \qquad k = 0, \dots, K,$$
$$\boxed{\boldsymbol{u}^{m,(0)} := \boldsymbol{u}(t^n),} \qquad m = 1, \dots, M$$
$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}) \text{ with } k = 1, \dots, K.$$

Theorem (Convergence DeC)

- $\mathcal{L}^2(\underline{\boldsymbol{u}}^*) = 0$
- *If* $\mathcal{L}^1$ *coercive with constant* C_1
- *If* $\mathcal{L}^1 - \mathcal{L}^2$ *Lipschitz with constant* $C_2 \Delta t$

Then $\|\underline{\boldsymbol{u}}^{(K)} - \underline{\boldsymbol{u}}^*\| \leq C(\Delta t)^K$



[5] A. Dutt, L. Greengard, and V. Rokhlin. BIT Numerical Mathematics, 40(2):241–266, 2000.

Deferred Correction[5]

How to combine two methods keeping the accuracy of the second and the stability and simplicity of the first one?

- $\mathcal{L}^1(\underline{\boldsymbol{u}}) = 0$, first order accuracy, easily invertible.
- $\mathcal{L}^2(\underline{\boldsymbol{u}}) = 0$, high order $M+1$.

$$\boldsymbol{u}^{0,(k)} := \boldsymbol{u}(t^n), \quad k = 0, \dots, K,$$

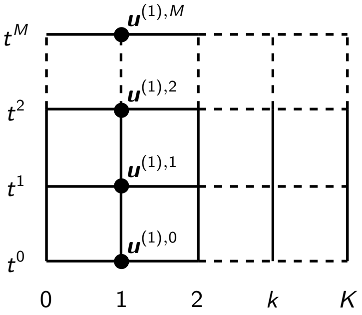
$$\boldsymbol{u}^{m,(0)} := \boldsymbol{u}(t^n), \quad m = 1, \dots, M$$

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}) \text{ with } k = 1, \dots, K.$$

Theorem (Convergence DeC)

- $\mathcal{L}^2(\underline{\boldsymbol{u}}^*) = 0$
- *If $\mathcal{L}^1$ coercive with constant C_1*
- *If $\mathcal{L}^1 - \mathcal{L}^2$ Lipschitz with constant $C_2 \Delta t$*

Then $\|\underline{\boldsymbol{u}}^{(K)} - \underline{\boldsymbol{u}}^*\| \leq C(\Delta t)^K$



[5] A. Dutt, L. Greengard, and V. Rokhlin. BIT Numerical Mathematics, 40(2):241–266, 2000.

Deferred Correction[5]

How to combine two methods keeping the accuracy of the second and the stability and simplicity of the first one?

- $\mathcal{L}^1(\underline{\boldsymbol{u}}) = 0$, first order accuracy, easily invertible.
- $\mathcal{L}^2(\underline{\boldsymbol{u}}) = 0$, high order $M+1$.

$$\boldsymbol{u}^{0,(k)} := \boldsymbol{u}(t^n), \quad k = 0, \dots, K,$$

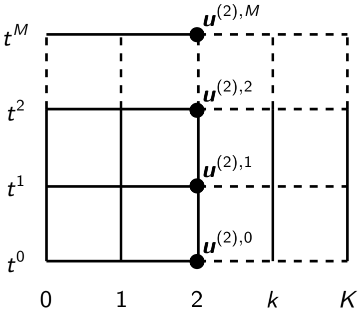
$$\boldsymbol{u}^{m,(0)} := \boldsymbol{u}(t^n), \quad m = 1, \dots, M$$

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}) \text{ with } k = 1, \dots, K.$$

Theorem (Convergence DeC)

- $\mathcal{L}^2(\underline{\boldsymbol{u}}^*) = 0$
- *If $\mathcal{L}^1$ coercive with constant C_1*
- *If $\mathcal{L}^1 - \mathcal{L}^2$ Lipschitz with constant $C_2 \Delta t$*

Then $\|\underline{\boldsymbol{u}}^{(K)} - \underline{\boldsymbol{u}}^*\| \leq C(\Delta t)^K$



[5] A. Dutt, L. Greengard, and V. Rokhlin. BIT Numerical Mathematics, 40(2):241–266, 2000.

Deferred Correction[5]

How to combine two methods keeping the accuracy of the second and the stability and simplicity of the first one?

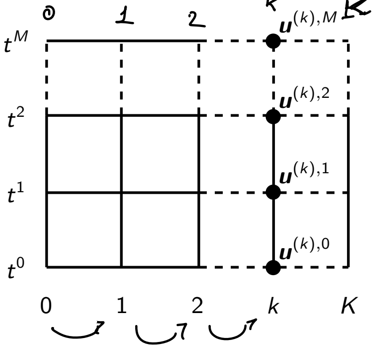
$$u^{0,(k)} := u(t^n), \quad k = 0, \dots, K,$$

$$u^{m,(0)} := u(t^n), \quad m = 1, \dots, M$$

$$\mathcal{L}^1(\underline{u}^{(k)}) = \boxed{\mathcal{L}^1(\underline{u}^{(k-1)}) - \mathcal{L}^2(\underline{u}^{(k-1)})} \text{ with } k = 1, \dots, K.$$

- $\mathcal{L}^1(\underline{u}) = 0$, first order accuracy, easily invertible.
- $\boxed{\mathcal{L}^2(\underline{u}) = 0}$ high order $\underbrace{M+1}_{K}$.

$K = M+1$



Theorem (Convergence DeC)

- $\mathcal{L}^2(\underline{u}^*) = 0$
- *If* $\mathcal{L}^1$ *coercive with constant* C_1
- *If* $\mathcal{L}^1 - \mathcal{L}^2$ *Lipschitz with constant* $C_2 \Delta t$

Then $\|\underline{u}^{(K)} - \underline{u}^*\| \leq C(\Delta t)^K$

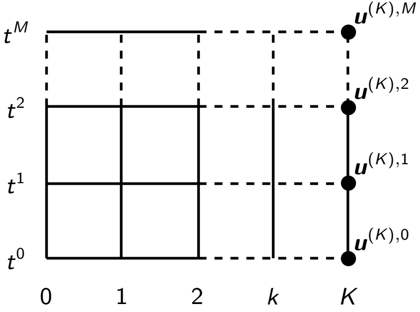
[5] A. Dutt, L. Greengard, and V. Rokhlin. BIT Numerical Mathematics, 40(2):241–266, 2000.

Deferred Correction[5]

How to combine two methods keeping the accuracy of the second and the stability and simplicity of the first one?

- $\mathcal{L}^1(\underline{\boldsymbol{u}}) = 0$, first order accuracy, easily invertible.
- $\mathcal{L}^2(\underline{\boldsymbol{u}}) = 0$, high order $M+1$.

$$\boldsymbol{u}^{0,(k)} := \boldsymbol{u}(t^n), \quad k = 0, \dots, K,$$
$$\boldsymbol{u}^{m,(0)} := \boldsymbol{u}(t^n), \quad m = 1, \dots, M$$
$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}) \text{ with } k = 1, \dots, K.$$



Theorem (Convergence DeC)

- $\mathcal{L}^2(\underline{\boldsymbol{u}}^*) = 0$
- *If $\mathcal{L}^1$ coercive with constant C_1*
- *If $\mathcal{L}^1 - \mathcal{L}^2$ Lipschitz with constant $C_2 \Delta t$*

Then $\|\underline{\boldsymbol{u}}^{(K)} - \underline{\boldsymbol{u}}^*\| \leq C(\Delta t)^K$

[5] A. Dutt, L. Greengard, and V. Rokhlin. BIT Numerical Mathematics, 40(2):241–266, 2000.

DeC – Proof

THESIS: $\| u^{(K)} - u^* \| \leq (C\Delta t)^K \cdot \| u^{(0)} - u^* \|$

HYP: 1) $\exists ! u^* : \mathcal{L}^2(u^*) = 0$ 2) $\| \mathcal{L}^1(u) - \mathcal{L}^1(v) \| \geq C_1 \| u - v \|$ 3) $\| \mathcal{L}^1(u) - \mathcal{L}^2(u) - (\mathcal{L}^1(v) - \mathcal{L}^2(v)) \| \leq \| u - v \| \cdot C_2 \Delta t$

Proof.

Let u^* be the solution of $\mathcal{L}^2(\underline{u}^*) = 0$. We know that $\mathcal{L}^1(\underline{u}^*) = \mathcal{L}^1(\underline{u}^*) - \mathcal{L}^2(\underline{u}^*)$, so that

$$\| u^{(K)} - u^* \| \leq \frac{1}{C_1} \| \mathcal{L}^1(u^{(K)}) - \mathcal{L}^1(u^*) \| = \frac{1}{C_1} \| \mathcal{L}^1(u^{(K-1)}) - \mathcal{L}^2(u^{(K-1)}) - \mathcal{L}^1(u^*) + \mathcal{L}^2(u^*) \|$$

$$\underset{\text{LIP.}}{\leq} \underbrace{\frac{C_2 \cdot \Delta t}{C_1}}_{C\Delta t} \| u^{(K-1)} - u^* \| \leq (C \Delta t)^2 \| u^{(K-2)} - u^* \| \leq \dots \leq (C\Delta t)^K \| u^{(0)} - u^* \|$$

$\square$

ITERATIVE PROCESS

$$\mathcal{L}^1(u^{(K)}) = \mathcal{L}^1(u^{(K-1)}) - \mathcal{L}^2(u^{(K-1)})$$

DeC – Proof

Proof.

Let f^* be the solution of $\mathcal{L}^2(\underline{\boldsymbol{u}}^*)=0$. We know that $\mathcal{L}^1(\underline{\boldsymbol{u}}^*)=\mathcal{L}^1(\underline{\boldsymbol{u}}^*)-\mathcal{L}^2(\underline{\boldsymbol{u}}^*)$, so that

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k+1)})-\mathcal{L}^1(\underline{\boldsymbol{u}}^*)=\left(\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)})-\mathcal{L}^2(\underline{\boldsymbol{u}}^{(k)})\right)-\left(\mathcal{L}^1(\underline{\boldsymbol{u}}^*)-\mathcal{L}^2(\underline{\boldsymbol{u}}^*)\right)$$

$$\begin{aligned}
C_1\|\underline{\boldsymbol{u}}^{(k+1)}-\underline{\boldsymbol{u}}^*\| \leq & \|\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k+1)})-\mathcal{L}^1(\underline{\boldsymbol{u}}^*)\| = \\
= & \|\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)})-\mathcal{L}^2(\underline{\boldsymbol{u}}^{(k)})-(\mathcal{L}^1(\underline{\boldsymbol{u}}^*)-\mathcal{L}^2(\underline{\boldsymbol{u}}^*))\| \leq \\
\leq & C_2\Delta\|\underline{\boldsymbol{u}}^{(k)}-\underline{\boldsymbol{u}}^*\|.
\end{aligned}$$

$$\|\underline{\boldsymbol{u}}^{(k+1)}-\underline{\boldsymbol{u}}^*\| \leq \left(\frac{C_2}{C_1}\Delta\right)\|\underline{\boldsymbol{u}}^{(k)}-\underline{\boldsymbol{u}}^*\| \leq \left(\frac{C_2}{C_1}\Delta\right)^{k+1}\|\underline{\boldsymbol{u}}^{(0)}-\underline{\boldsymbol{u}}^*\|.$$

After K iteration we have an error at most of $\left(\frac{C_2}{C_1}\Delta\right)^K\|\underline{\boldsymbol{u}}^{(0)}-\underline{\boldsymbol{u}}^*\|$. □

~~DeC: Second order example~~

COERCIVITY: $||\mathcal{L}^2(\underline{u}) - \mathcal{L}^2(\underline{v})|| \geq C_1 ||\underline{u} - \underline{v}||$

$$\mathcal{L}^1(u)-\mathcal{L}^1(v) = \begin{pmatrix} u^n - \cancel{u^0} + \cancel{\beta^n \Delta t F(u^0)} \\ \vdots \\ u^1 - \cancel{u^0} + \cancel{\beta^1 \Delta t F(u^0)} \end{pmatrix} - \begin{pmatrix} v^n - \cancel{u^0} + \cancel{\beta^n \Delta t F(u^0)} \\ v^1 - \cancel{u^0} + \cancel{\beta^1 \Delta t F(u^0)} \end{pmatrix}$$

$$= \begin{pmatrix} u^n - v^n \\ u^1 - v^1 \end{pmatrix} = \underline{u} - \underline{v} \qquad C_1 = 1$$

LIPSCHITZ CONT

$\mathcal{L}^1$ EXPL EUL APPROX

$\mathcal{L}^2$ H.O. IMPLICIT RK METH

$\underbrace{\mathcal{L}^1(u^{ex}) = O(\Delta t^2)}$

$\mathcal{L}^2(u^{ex}) = O(\Delta t^{P+1})$

$|\mathcal{L}^1 - \mathcal{L}^2| = O(\Delta t^2)$

SKETCH

PROOF WITH DETAILS

EFFICIENT DeC

MICALIZZI, TORLO

DeC: Second order example

t^1, t^0

$$\mathcal{L}^2(u) = \left(u^1 - u^0 + \Delta t \, \frac{1}{2}\left(F(u^0) - F(u^1)\right)\right)$$

$$\mathcal{L}^1(u) = u^1 - u^0 + \Delta t \, F(u^0)$$

ITERATIVE PROCESS

$$u^{(0),0} = u^{(0),1} = u(t^n) = u^0 \qquad \boxed{k=0}$$

$k=1$ $$\mathcal{L}^1(u^{(1)}) = \mathcal{L}^1(u^{(0)}) - \mathcal{L}^2(u^{(0)})$$

$$u^{(1),1} - \cancel{u^0} + \cancel{\Delta t F(u^0)} = \cancel{u^0} - \cancel{u^0} + \cancel{\Delta t F(u^0)} - \left(\cancel{u^0} - u^0 + \frac{\Delta t}{2}\left[F(u^0) + F(u^0)\right]\right)$$

$$u^{(1),1} = u^0 - \Delta t F(u^0) \qquad \text{(EXPLICIT EULER 1}^{st}\text{ ORDER)}$$

$\underline{k=2}$ $$\mathcal{L}^1(u^{(2)}) = \mathcal{L}^1(u^{(1)}) - \mathcal{L}^2(u^{(1)})$$

$$u^{(2),1} - \cancel{u^0} + \cancel{\Delta t F(u^0)} = \cancel{u^{(1),1}} - \cancel{u^0} + \cancel{\Delta t F(u^0)} - \left[\cancel{u^{(1),1}} - u^0 + \frac{\Delta t}{2}\left(F(u^0) + F(u^{(1),1})\right)\right]$$

DeC: Second order example

$$U^{(2),1} = U^0 - \frac{\Delta t}{2}\left(F(U^0) + F(U^{(1),1}) \right)$$

2^{nd} ORDER ACCURATE

DeC: Second order example

Simplification of DeC for ODE

In practice

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

For $m = 1, \dots, M$

$$\boldsymbol{u}^{(k),m} - \boldsymbol{u}^0 - \beta^m \Delta t F(\boldsymbol{u}^0) - \boldsymbol{u}^{(k-1),m} + \boldsymbol{u}^0 + \beta^m \Delta t F(\boldsymbol{u}^0)$$

$$+ \boldsymbol{u}^{(k-1),m} - \boldsymbol{u}^0 - \Delta t \sum_{r=0}^{M} \theta_r^m F(\boldsymbol{u}^{(k-1),r}) = 0$$

Simplification of DeC for ODE

In practice

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

For $m = 1, \dots, M$

$$\boldsymbol{u}^{(k),m} \cancel{- \boldsymbol{u}^0 - \beta^m \Delta t F(\boldsymbol{u}^0)} - \boldsymbol{u}^{(k-1),m} \cancel{+ \boldsymbol{u}^0 + \beta^m \Delta t F(\boldsymbol{u}^0)}$$

$$+ \boldsymbol{u}^{(k-1),m} - \boldsymbol{u}^0 - \Delta t \sum_{r=0}^{M} \theta_r^m F(\boldsymbol{u}^{(k-1),r}) = 0$$

Simplification of DeC for ODE

In practice

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

For $m = 1, \dots, M$

$$\boldsymbol{u}^{(k),m} - \cancel{\boldsymbol{u}^0 - \beta^m \Delta t F(\boldsymbol{u}^0)} - \cancel{\boldsymbol{u}^{(k-1),m}} + \cancel{\boldsymbol{u}^0 + \beta^m \Delta t F(\boldsymbol{u}^0)}$$

$$+ \cancel{\boldsymbol{u}^{(k-1),m}} - \boldsymbol{u}^0 - \Delta t \sum_{r=0}^{M} \theta_r^m F(\boldsymbol{u}^{(k-1),r}) = 0$$

Simplification of DeC for ODE

In practice

$$\underbrace{\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)})} = \mathcal{L}^1(\underbrace{\underline{\boldsymbol{u}}^{(k-1)}}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

For $m = 1, \dots, M$

$$\boldsymbol{u}^{(k),m} \underbrace{- \boldsymbol{u}^0 \cancel{- \beta^m \Delta t F(\boldsymbol{u}^0)}} - \underbrace{\cancel{\boldsymbol{u}^{(k-1),m}}} \underbrace{+ \boldsymbol{u}^0 \cancel{+ \beta^m \Delta t F(\boldsymbol{u}^0)}}$$
$$\underbrace{+ \cancel{\boldsymbol{u}^{(k-1),m}}} - \boldsymbol{u}^0 - \Delta t \sum_{r=0}^{M} \theta_r^m F(\boldsymbol{u}^{(k-1),r}) = 0$$

$$\boxed{\boldsymbol{u}^{(k),m} - \boldsymbol{u}^0 - \Delta t \sum_{r=0}^{M} \theta_r^m F(\boldsymbol{u}^{(k-1),r}) = 0.}$$

$\forall k \quad \forall m$

DeC and residual distribution

n $\partial_t u = F(u)$ $\mathcal{L}^2(\square^n\, \partial_t u + \dots$ implic$)$

Deferred Correction + Residual distribution

- Residual distribution (FV $\Rightarrow$ FE) $\Rightarrow$ High order in space
- Prediction/correction/iterations $\Rightarrow$ High order in time
- Subtimesteps $\Rightarrow$ High order in time

$\mathcal{L}^1 = (\backslash\; \partial_t u +$ expl$)$

$$U_\xi^{m,(k+1)} = U_\xi^{m,(k)} - |C_p|^{-1} \sum_{\mathrm{E}|\xi\in\mathrm{E}} \left(\int_{\mathrm{E}} \Phi_\xi \left(U^{m,(k)} - U^{n,0} \right) \mathrm{d}\mathbf{x} + \Delta t \sum_{r=0}^{M} \theta_r^m \mathcal{R}_\xi^{\mathrm{E}}(U^{r,(k)}) \right),$$

EXPLICIT

with

$$\sum_{\xi\in\mathrm{E}} \mathcal{R}_\xi^{\mathrm{E}}(u) = \int_{\mathrm{E}} \nabla_{\mathbf{x}} F(u) \mathrm{d}\mathbf{x}.$$

- The $\mathcal{L}^2$ operator contains also the complications of the spatial discretization (e.g. mass matrix)
- $\mathcal{L}^1$ operator further simplified up to a first order approximation (e.g. **mass lumping**)

$\mathcal{L}^1$ with mass lumping

Implicit simple DeC (Rosenbrock)

Define $\mathcal{L}^1$ as

$$\mathcal{L}^1(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) = \begin{pmatrix} \boldsymbol{u}^M - \boldsymbol{u}^0 - \Delta t \beta^M F(\boldsymbol{u}^0) \\ \vdots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 - \Delta t \beta^1 F(\boldsymbol{u}^0) \end{pmatrix}$$

Implicit simple DeC (Rosenbrock)

Define $\mathcal{L}^1$ as

$$\mathcal{L}^1(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) = \begin{pmatrix} \boldsymbol{u}^M - \boldsymbol{u}^0 - \Delta t \beta^M \left(F(\boldsymbol{u}^0) + \partial_{\boldsymbol{u}} F(\boldsymbol{u}^0)(\boldsymbol{u}^M - \boldsymbol{u}^0) \right) \\ \vdots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 - \Delta t \beta^1 \left(F(\boldsymbol{u}^0) + \partial_{\boldsymbol{u}} F(\boldsymbol{u}^0)(\boldsymbol{u}^1 - \boldsymbol{u}^0) \right) \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{u}^M - \boldsymbol{u}^0 - \Delta t \beta^M \partial_{\boldsymbol{u}} F(\boldsymbol{u}^0) \boldsymbol{u}^M \\ \vdots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 - \Delta t \beta^1 \partial_{\boldsymbol{u}} F(\boldsymbol{u}^0) \boldsymbol{u}^1 \end{pmatrix}$$

$$\underbrace{\left(I - \Delta t \beta^m \boxed{\partial_u F(u^0)}\right)}\, u^m$$

Implicit simple DeC (Rosenbrock)

$$\mathcal{L}^{1,m}(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) = \boldsymbol{u}^m - \boldsymbol{u}^0 - \Delta t \beta^m \partial_{\boldsymbol{u}} F(\boldsymbol{u}^0)\boldsymbol{u}^m$$

$$\mathcal{L}^{2,m}(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) = \boldsymbol{u}^m - \boldsymbol{u}^0 - \Delta t \sum_r \theta_r^m F(\boldsymbol{u}^r)$$

ITERATION PROCESS

$$\mathcal{L}^{1,m}(U^{(k)}) - \mathcal{L}^{1,m}(U^{(k-1)}) + \mathcal{L}^{2,m}(U^{(k-1)}) = 0$$

$$U^{m,(k)} - \cancel{U^0} - \Delta t \beta^m \partial_u F(U^0) \cdot U^{m,(k)} - U^{m,(k-1)} + \cancel{U^0} + \Delta t \beta^m \partial_u F(U^0) U^{m,(k-1)}$$

$$+ \underbrace{U^{m,(k)} - U^0 - \Delta t \theta_r^m F(U^r)} = 0$$

$$\underbrace{\left[I - \Delta t \beta^m \boxed{\partial_u F(U^0)}\right]}(U^{m,(k)} - U^{m,(k-1)}) = \mathcal{L}^{2,m}(U^{(k-1)}) \qquad k = 1, \dots, K$$

$$\forall m$$

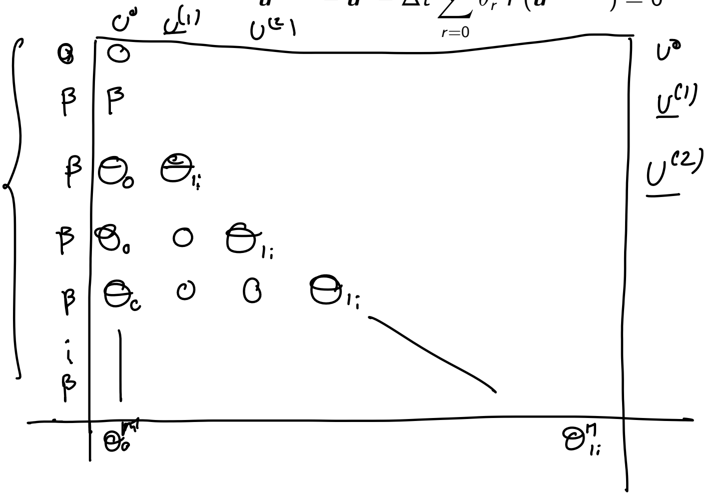
Implicit simple DeC (Rosenbrock)

DeC as RK

$$\forall k = 1, \dots, K$$
$$\forall m = 1, \dots, M$$

$$\boldsymbol{u}^{(k),m} - \boldsymbol{u}^0 - \Delta t \sum_{r=0}^{M} \theta_r^m F(\boldsymbol{u}^{(k-1),r}) = 0$$

$$S \equiv K \cdot M$$



$$U^0$$
$$\underline{U}^{(1)} = \begin{pmatrix} U^{(1),1} \\ U^{(1),M} \end{pmatrix}$$
$$\underline{U}^{(2)}$$

DeC as RK

DeC as RK

We can write DeC as RK defining $\underline{\theta}_0 = \{\theta_0^m\}_{m=1}^M$, $\underline{\theta}^M = \theta_r^M$ with $r \in 1, \dots, M$, denoting the vector $\underline{\theta}_r^{M,T} = (\theta_1^M, \dots, \theta_M^M)$. The Butcher tableau for an arbitrarily high order DeC approach is given by:

$$
\begin{array}{c|ccccccc}
0 & 0 & & & & & & \\
\underline{\beta} & \underline{\beta} & & & & & & \\
\underline{\beta} & \underline{\theta}_0 & \underline{\underline{\tilde{\theta}}} & & & & & \\
\vdots & \underline{\theta}_0 & \underline{\underline{0}} & \underline{\underline{\tilde{\theta}}} & & & & \\
\vdots & \underline{\theta}_0 & \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{\tilde{\theta}}} & & & \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \\
\underline{\beta} & \underline{\theta}_0 & \underline{\underline{0}} & \dots & \dots & \underline{\underline{0}} & \underline{\underline{\tilde{\theta}}} & \\
\hline
 & \theta_0^M & \underline{0}^T & \dots & & \dots & \underline{0}^T & \underline{\theta}_r^{M,T}
\end{array}
. \tag{6}
$$

Stability of (explicit) DeC

Idea: study the RK version!

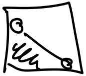
$$\boxed{u' = \lambda u} \qquad \boxed{\Re(\lambda) < 0.} \tag{7}$$

$$u_{n+1} = R(\lambda \Delta t) u_n, \qquad R(z) = 1 + z b^T (I - zA)^{-1} \mathbf{1}, \qquad z = \lambda \Delta t \tag{8}$$

Goal: find $z \in \mathbb{C}$ such that $|R(z)| < 1$.
Recall: stability function for explicit RK methods is a polynomial, indeed the inverse of $(I - zA)$ can be written in Taylor expansion as

$$(I - zA)^{-1} = \sum_{r=0}^{\infty} z^r A^s = I + zA + z^2 A^2 + \dots, \tag{9}$$

and, since A is strictly lower triangular, it is nilpotent. Hence, $R(z)$ is a polynomial in z with degree at most equal to S.



$A^S = 0$

Stability of (explicit) DeC

Theorem HAIRER BOOK

If the RK method is of order P, then

$$R(z) = \underbrace{1 + z + \frac{z^2}{2!} + \dots + \frac{z^P}{P!}}_{} + \underbrace{O(z^{P+1})}_{}. \tag{10}$$

The first $P+1$ terms of the stability functions $R(\cdot)$ for explicit DeCs of order P are known.

Theorem

The stability function of any explicit DeC of order P (with P iterations) is

$$R(z) = \sum_{r=0}^{P} \frac{z^r}{r!} = \underbrace{1 + z + \frac{z^2}{2!} + \dots + \frac{z^P}{P!}}_{} \tag{11}$$

and does not depend on the distribution of the subtimenodes.

Proof (1/3)

$$A = \begin{array}{cc} \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ \star & 0 & 0 & \dots & 0 & 0 \\ \star & \star & 0 & \dots & 0 & 0 \\ \star & 0 & \star & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \star & 0 & 0 & \cdots & \star & 0 \end{pmatrix} & \begin{matrix} 0 \\ 1 \\ \\ , \\ \\ K-1 \end{matrix} \end{array}$$

Block structure of the matrix A

$\star$ are some non-zero block matrices and the 0 are some zero block matrices.

The number of blocks in each line and row of these matrices is P, the order of the scheme.

$A^P = 0$

Stability of (explicit) DeC

Proof (2/3)

By induction, A^k has zeros in the upper triangular part, in the main block diagonal, and in all the $k-1$ block diagonals below the main diagonal, i.e.,

$$(A^k)_{i,j} = 0 \quad , \text{if } i < j+k,$$

where the indexes here refer to the blocks. Indeed, it is true that $A_{i,j} = 0$ if $i < j+1$. Now, let us consider the entry $(A^{k+1})_{i,j}$ with $i < j+k+1$, i.e., $i-k < j+1$. It is defined as

$$(A^{k+1})_{i,j} = \sum_w (A^k)_{i,w} A_{w,j}. \tag{12}$$

Now, we can prove that all the terms of the sum are 0. Let $w < j+1$, then $A_{w,j} = 0$ because of the structure of A; while, if $w \geq j+1 > i-k$, we have that $i < w+k$, so $(A^k)_{i,w} = 0$ by induction.

Stability of (explicit) DeC

Proof (3/3)

In particular, this means that $A^P = \underline{\underline{0}}$, because i is always smaller than $j+P$ as P is the number of the block matrices that we have. Hence,

$$\underbrace{(I-zA)^{-1}} = \sum_{r=0}^{\infty} z^r A^s = \sum_{r=0}^{P-1} z^r A^s = \underbrace{I + zA + z^2A^2 + \cdots + \boxed{z^{P-1}}A^{P-1}}. \tag{13}$$

Plugging this result into $R(z) = 1 + zb^T(I-zA)^{-1}\mathbf{1}$, the stability function $R(z)$ is a polynomial of degree P, the order of the scheme. All terms of order lower or equal to P must agree with the expansion of the exponential function, so it must be

$$R(z) = \sum_{r=0}^{P} \frac{z^r}{r!} = \underbrace{1 + z + \frac{z^2}{2!} + \cdots + \frac{z^P}{P!}}. \tag{14}$$

Note: no assumption on the distribution of the subtimenodes.

CODE

- Choice of iterations (P) and order
- Choice of point distributions $t^0, \dots, t^M$
- Computation of θ
- Loop for timesteps
- Loop for correction
- Loop for subtimesteps

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

ADER   Arbitrary DERivative

- Cauchy–Kovalevskaya theorem
- Modern automatic version 2008
- Space/time DG
- Prediction/Correction
- Fixed-point iteration process

Modern approach is DG in space time for hyperbolic problem

FV   FEM

$$\boxed{\partial_t u(x,t) + \nabla \cdot F(u(x,t)) = 0}, \quad x \in \Omega \subset \mathbb{R}^d, \;\; t > 0. \quad (15)$$

$\theta_{rs}(x,t)$

WEAK FORMULATION

Prediction: iterative procedure

TEST

$$\sum_{p,q} \int_{T^n \times V_i} \theta_{rs}(x,t) \partial_t \theta_{pq}(x,t) z^{pq} \mathrm{d}x\mathrm{d}t + \int_{T^n \times V_i} \theta_{rs}(x,t) \nabla_{\mathbf{x}} \cdot F(\underbrace{\theta_{pq}(x,t) z^{pq}}_{\text{RECONSTR}}) \mathrm{d}x\mathrm{d}t = 0. \quad \forall r,s$$

Correction step: communication between cells

$$\int_{V_i} \Phi_r \left(u(t^{n+1}) - u(t^n) \right) \mathrm{d}x + \int_{T^n \times \partial V_i} \Phi_r(x) \mathcal{G}(z^-, z^+) \cdot \mathbf{n} \, \mathrm{d}S \, \mathrm{d}t - \int_{T^n \times V_i} \nabla_{\mathbf{x}} \Phi_r \cdot F(z) \, \mathrm{d}x \, \mathrm{d}t = 0,$$

ADER: space-time discretization

Defining $\theta_{rs}(x,t) = \Phi_r(x)\phi_s(t)$ basis functions in space and time

$$\int_{T^n \times V_i} \theta_{rs}(x,t)\partial_t \theta_{pq}(x,t) u^{pq} \mathrm{d}x\mathrm{d}t + \int_{T^n \times V_i} \theta_{rs}(x,t) \nabla \cdot F(\theta_{pq}(x,t)u^{pq}) \mathrm{d}x\mathrm{d}t = 0. \tag{16}$$

ADER: space-time discretization

Defining $\theta_{rs}(x,t) = \Phi_r(x)\phi_s(t)$ basis functions in space and time

$$\int_{T^n \times V_i} \theta_{rs}(x,t)\partial_t \theta_{pq}(x,t) u^{pq} \mathrm{d}x\mathrm{d}t + \int_{T^n \times V_i} \theta_{rs}(x,t) \nabla \cdot F(\theta_{pq}(x,t)u^{pq}) \mathrm{d}x\mathrm{d}t = 0. \tag{16}$$

This leads to

$$\underline{\underline{\underline{\underline{\mathrm{M}}}}}_{rspq} u^{pq} = \underline{\underline{r}}(\underline{\underline{\mathbf{u}}})_{rs}, \tag{17}$$

solved with fixed point iteration method.

$+$ Correction step where cells communication is allowed (derived from (16)).

ADER: time integration method

ODE $\partial_t u = F(u)$

Simplify! Take $\boldsymbol{u}(t) = \sum_{m=0}^{M} \phi_m(t) \boldsymbol{u}^m = \underline{\phi}(t)^T \underline{\boldsymbol{u}}$

WEAK FORM OF ODE
$$\int_{T^n} \psi(t) \partial_t \boldsymbol{u}(t) dt - \int_{T^n} \psi(t) F(\boldsymbol{u}(t)) dt = 0, \quad \forall \psi : \underbrace{T^n = [t^n, t^{n+1}]}\to \mathbb{R}.$$

$$\mathcal{L}^2(\underline{\boldsymbol{u}}) := \cancel{\int_{T^n} \underline{\phi}(t) \partial_t \underline{\phi}(t)^T \underline{\boldsymbol{u}} dt - \int_{T^n} \underline{\phi}(t) F(\underline{\phi}(t)^T \underline{\boldsymbol{u}}) dt = 0}$$

$\underline{\phi}(t) = (\phi_0(t), \dots, \phi_M(t))^T$

$$\boxed{\int_{T^n} \phi_i(t) \partial_t \phi_j(t) U^j \, dt} - \int_{T^n} \phi_i F(\phi_j(t) U^j) dt$$

Quadrature...

INTEGRATION BY PARTS

$$= \phi_i(t^{n+1}) \phi_j(t^{n+1}) U^j - \phi_i(t^n) \underbrace{\phi_j(t^n) U^j}_{U(t^n)}$$

$$- \int_{t^n}^{t^{n+1}} \partial_t \phi_i(t) \cdot \phi_j(t) U^j dt$$

$$- \int_{T^n} \phi_i F(\phi_j U^j) dt$$

$$\mathcal{L}^2(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - \underline{r}(\underline{\boldsymbol{u}}) = 0 \Longleftrightarrow \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} = \underline{r}(\underline{\boldsymbol{u}}). \quad (18)$$

Nonlinear system of $M \times S$ equations

$$\phi_i(t^{n+1}) \phi_j(t^{n+1}) U^j - \int_{t^n}^{t^{n+1}} \phi_i'(t) \phi_j(t) dt \cdot U^j$$

$$- \underbrace{\phi_i(t^n) U(t^n)}_{\text{EXPLICIT}} - \int_{t^n}^{t^{n+1}} \phi_i F(\phi_j(t) U^j) dt = \mathcal{L}^{2,i}$$

ADER: Mass matrix

What goes into the mass matrix? Use of the integration by parts

WEAK FORMULATION

$$\mathcal{L}^2(\underline{\mathbf{u}}) := \int_{T^n} \underline{\phi}(t)\partial_t \underline{\phi}(t)^T \underline{\mathbf{u}} dt + \int_{T^n} \underline{\phi}(t) F(\underline{\phi}(t)^T \underline{\mathbf{u}}) dt =$$

I.B.P. IN TIME $\Rightarrow$

$$\underline{\phi}(t^{n+1})\underline{\phi}(t^{n+1})^T \underline{\mathbf{u}} - \underbrace{\underline{\phi}(t^n)\mathbf{u}^n}_{\text{KNOWN}} - \int_{T^n} \partial_t \underline{\phi}(t)\underline{\phi}(t)^T \underline{\mathbf{u}} - \int_{T^n} \underline{\phi}(t) F(\underline{\phi}(t)^T \underline{\mathbf{u}}) dt$$

$$\underline{\underline{\mathrm{M}}} = \underline{\phi}(t^{n+1})\underline{\phi}(t^{n+1})^T - \int_{T^n} \partial_t \underline{\phi}(t)\underline{\phi}(t)^T$$

$$\underline{r}(\underline{\mathbf{u}}) = \underline{\phi}(t^n)\mathbf{u}^n + \int_{T^n} \underline{\phi}(t) F(\underline{\phi}(t)^T \underline{\mathbf{u}}) dt$$

$$\underline{\underline{\mathrm{M}}}\underline{\mathbf{u}} = \underline{r}(\underline{\mathbf{u}})$$

NONLINEAR SYS

LINEAR NONLINEAR DIM $(M \times S)$

ADER: Fixed point iteration

Iterative procedure to solve the problem for each time step

$$\underline{\boldsymbol{u}}^{(k)} = \underline{\underline{\mathrm{M}}}^{-1}\underline{\boldsymbol{r}}(\underline{\boldsymbol{u}}^{(k-1)}), \quad k = 1, \dots, \text{convergence} \tag{19}$$

with $\underline{\boldsymbol{u}}^{(0)} = \boldsymbol{u}(t^n)$.
Reconstruction step

$$\boldsymbol{u}(t^{n+1}) = \boldsymbol{u}(t^n) - \int_{T^n} F(\boldsymbol{u}^{(K)}(t))dt.$$

- Convergence?
- How many steps K?
- Accuracy $\mathcal{L}^2$? $\longrightarrow$ IMPLICIT RK

ADER 2nd order

POLY LAGR	QUAD
GAUSS	GAUSS
LOBATTO	LOBATTO
EQUI	GAUSS

Example with 2 Gauss Legendre points, Lagrange polynomials and 2 iterations
Let us consider the timestep interval $[t^n, t^{n+1}]$, rescaled to $[0, 1]$.
Gauss-Legendre points quadrature and interpolation (in the interval $[0, 1]$)

$$\underline{t}_q = (t_q^0, t_q^1) = (t^0, t^1) = \left(\frac{\sqrt{3}-1}{2\sqrt{3}}, \frac{\sqrt{3}+1}{2\sqrt{3}} \right), \quad \underline{w} = (1/2, 1/2).$$

$$\underline{\phi}(t) = (\phi_0(t), \phi_1(t)) = \left(\frac{t-t^1}{t^0-t^1}, \frac{t-t^0}{t^1-t^0} \right).$$

p points

Then, the mass matrix is given by

$$\phi_m(t^{n+1})\phi_l(t^{n+1}) - \int_{t^n}^{t^{n+1}} \phi_m'(t)\cdot\phi_l(t)\,dt$$

$$\underbrace{\underline{\underline{\mathrm{M}}}_{m,l} = \phi_m(1)\phi_l(1) - \phi_m'(t^l)w_l,} \quad m, l = 0, 1,$$

$$\mathbb{P}^{P-2}$$

$$-\int_0^1 \phi_m' \phi_l \,dt = -w_l \cdot \phi_m'(t^l)$$

$$\underline{\underline{\mathrm{M}}} = \begin{pmatrix} 1 & \frac{\sqrt{3}-1}{2} \\ -\frac{\sqrt{3}+1}{2} & 1 \end{pmatrix}.$$

QUAD = LAGR POLY

ADER 2nd order

$$\sum_J \phi_J U^J$$

The right hand side is given

$$\phi_m U^n + \Delta t \int_0^1 \phi_m F(\overbrace{u(t)}^{\sum_J \phi_J U^J})\, dt \approx \underline{\phi_m \cdot U^n + \Delta t\, w_m \cdot F(u^m)}$$

$$r(\underline{\boldsymbol{u}})_m = \overset{U^n}{\cancel{\alpha(0)}}\phi_m(0) + \underline{\Delta} t F(\cancel{\alpha}(t^m)) w_m, \quad m = 0, 1.$$

$$\phi_J(t^m) = \delta_{Jm}$$

$$\underbrace{\underline{r}(\underline{\boldsymbol{u}}) = \overset{U^n}{\cancel{\alpha(0)}}\underline{\phi}(0) + \Delta t \begin{pmatrix} F(\cancel{\alpha}(t^1)) w_1 \\ F(\cancel{\alpha}(t^2)) w_2. \end{pmatrix}}.$$

Then, the coefficients $\underline{\boldsymbol{u}}$ are given by

$$\boxed{\underline{\boldsymbol{u}}^{(k+1)} = \underline{\underline{\mathrm{M}}}^{-1} \underline{r}(\underline{\boldsymbol{u}}^{(k)}).} = \overbrace{\mathrm{M}^{-1} \phi(1)^T U^n}^{= U^n} + \Delta t \underbrace{\underline{\underline{\mathrm{M}}}^{-1} \underline{\underline{R}}}_{\text{eval mat}} \overbrace{F(\underline{U}^{(k)})}^{\text{RHS}}$$

Finally, use $\underline{\boldsymbol{u}}^{(k+1)}$ to reconstruct the solution at the time step t^{n+1}:

$$\sum \phi_J(1) U^{J,(k+1)}$$

$$\boldsymbol{u}^{n+1} = \underline{\phi}(1)^T \underline{\boldsymbol{u}}^{(k+1)} = \boldsymbol{u}^n + \underbrace{\int_{T^n} \underline{\phi}(t)^T dt\, F(\underline{\boldsymbol{u}}^{(k)})}.$$

CODE

- Choice: ϕ Lagrangian basis functions (TRIAL = TEST)
- Different subtimesteps: Gauss-Legendre, Gauss–Lobatto, equispaced
  QUAD | GAUSS LOB GAUSS
- Precompute $\underline{\underline{M}}$
- Precompute the rhs vector part using quadratures after a further approximation

$$\underline{r}(\underline{\boldsymbol{u}}) = \underbrace{\underline{\phi}(t^n)}_{\underline{\phi}(0)}\boldsymbol{u}^n + \int_{T^n} \underline{\phi}(t) F(\underline{\phi}(t)^T \underline{\boldsymbol{u}}) dt \approx \underline{\phi}(t^n)\boldsymbol{u}^n + \underbrace{\int_{T^n} \underline{\phi}(t)\underline{\phi}(t)^T dt}_{\text{Can be stored}} F(\underline{\boldsymbol{u}})$$

- Precompute the reconstruction coefficients $\underline{\phi}(1)^T$

$$\underline{\underline{R}}_{ij} = \int \varphi_i \varphi_j \, dt$$

IF QUAD = LAGR POINTS $= w_i \delta_{ij}$

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

ADER[6] and DeC[7]: immediate similarities

ARBITRARILY

- High order time(space) discretization
- Start from a well known space discretization (FE/DG/FV)
- FE reconstruction in time
- System in time, with M equations $\mathcal{L}^2 = 0$
- Iterative method / K corrections

[6] M. Dumbser, D. S. Balsara, E. F. Toro, and C.-D. Munz. A unified framework for the construction of one-step finite volume and discontinuous galerkin schemes on unstructured meshes. Journal of Computational Physics, 227(18):8209–8253, 2008.

[7] R. Abgrall. High order schemes for hyperbolic problems using globally continuous approximation and avoiding mass matrices. Journal of Scientific Computing, 73(2):461–494, Dec 2017.

ADER[6] and DeC[7]: immediate similarities

- High order time-space discretization
- Start from a well known space discretization (FE/DG/FV)
- FE reconstruction in time
- System in time, with M equations
- Iterative method / K corrections

- Both high order explicit time integration methods (neglecting spatial discretization)

[6] M. Dumbser, D. S. Balsara, E. F. Toro, and C.-D. Munz. A unified framework for the construction of one-step finite volume and discontinuous galerkin schemes on unstructured meshes. Journal of Computational Physics, 227(18):8209–8253, 2008.

[7] R. Abgrall. High order schemes for hyperbolic problems using globally continuous approximation and avoiding mass matrices. Journal of Scientific Computing, 73(2):461–494, Dec 2017.

ADER as DeC

ADER $\underline{\underline{M}}\, \underline{u}^{(k)} = \underline{r}(\underline{u}^{(k-1)})$ GOAL

$$\begin{cases} \mathcal{L}^2(\underline{u}) = \underline{\underline{M}}\,\underline{u} - \underline{r}(\underline{u}) \\ \mathcal{L}^1(\underline{u}) = \underline{\underline{M}}\,\underline{u} - \underbrace{\underline{r}(\underline{u}^n)}_{\text{explicit}} \end{cases}$$

FIRST APPROXIMATION

$$\mathcal{L}^1(\underline{u}^{(k)}) = \mathcal{L}^1(\underline{u}^{(k-1)}) - \mathcal{L}^2(\underline{u}^{(k-1)})$$

$$\underline{\underline{M}}\,\underline{u}^{(k)} - \cancel{\underline{r}(\underline{u}^n)} = \cancel{\underline{\underline{M}}\,\underline{u}^{(k-1)}} - \cancel{\underline{r}(\underline{u}^n)} - \left(\cancel{\underline{\underline{M}}\,\underline{u}^{(k-1)}} - \underline{r}(\underline{u}^{(k-1)})\right)$$

$$\underline{\underline{M}}\,\underline{u}^{(k)} = \underline{r}(\underline{u}^{(k-1)}) \quad \checkmark$$

DeC $\mathcal{L}^{1,m}(\underline{U}) = U^m - U^0 + \Delta t F(U^0)\beta^m$

$\mathcal{L}^{2,m}(\underline{U}) = U^m - U^0 + \Delta t \sum_{r=0} \theta_r^m F(U^r)$

$\Rightarrow \mathcal{L}^1(U^{(k+1)}) = \mathcal{L}^1(U^{(k)}) - \mathcal{L}^2(U^{(k)})$

ADER as DeC

✓ $\mathcal{L}^1$ is coercive ✓ $\mathcal{L}^1-\mathcal{L}^2$ is Lipschitz cont. const $\Delta t \cdot C_2$

$\cdot\ \exists !_{u^*}\ \mathcal{L}^2(u^*)=0 \quad \Rightarrow \quad \|u^{(k)}-u^*\| \le (C\Delta t)^k \|u^{(0)}-u^*\|$

$$\boxed{\begin{cases} \mathcal{L}^2(\underline{u}) = \underline{\underline{M}}\,\underline{u} - \underline{R}(\underline{u}) \\ \mathcal{L}^1(\underline{u}) = \underline{\underline{M}}\,\underline{u} - \underline{R}(u^n) \end{cases}}$$

(c) $\|\mathcal{L}^1(u) - \mathcal{L}^1(v)\| \ge C_1 \|u-v\|$

$\|\underline{\underline{M}}\,\underline{u} - \cancel{R(u^n)} - \underline{\underline{M}}\,v - \cancel{R(u^n)}\|$

$= \|\underline{\underline{M}}(u-v)\| > C_1(\underline{\underline{M}})\,\|\underline{u}-\underline{v}\|$

$(\|M^{-1}\|)$

(L) $\|\mathcal{L}^1(u) - \mathcal{L}^2(u) - \mathcal{L}^1(v) + \mathcal{L}^2(v)\| = \|R(\underline{u}) - \cancel{R(u^n)} - R(\underline{v}) + \cancel{R(u^n)}\|$

$= \|R(u) - R(v)\| \le \|\cancel{u^n} + \Delta t \underline{R} F(\underline{u}) - \cancel{u^n} + \Delta t \underline{R} F(\underline{v})\| \le \underbrace{\Delta t}\cdot \underbrace{\|\underline{R}\| \, L}_{C_2} \|\underline{u}-\underline{v}\|$

ADER as DeC

$$\mathcal{L}^2(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\underline{\boldsymbol{u}}),$$
$$\mathcal{L}^1(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\boldsymbol{u}(t^n)).$$

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

$$\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k)} - r(\boldsymbol{u}^{(k),0}) - \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k-1)} + r(\boldsymbol{u}^{(k-1),0}) + \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k-1)} - r(\underline{\boldsymbol{u}}^{(k-1)}) = 0$$

ADER as DeC

$$\mathcal{L}^2(\underline{\mathbf{u}}) := \underline{\underline{\mathrm{M}}}\underline{\mathbf{u}} - r(\underline{\mathbf{u}}),$$
$$\mathcal{L}^1(\underline{\mathbf{u}}) := \underline{\underline{\mathrm{M}}}\underline{\mathbf{u}} - r(\mathbf{u}(t^n)).$$

$$\mathcal{L}^1(\underline{\mathbf{u}}^{(k)}) = \mathcal{L}^1(\underline{\mathbf{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\mathbf{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

$$\underline{\underline{\mathrm{M}}}\underline{\mathbf{u}}^{(k)} - \cancel{r(\mathbf{u}^{(k),0})} - \underline{\underline{\mathrm{M}}}\underline{\mathbf{u}}^{(k-1)} + \cancel{r(\mathbf{u}^{(k-1),0})} + \underline{\underline{\mathrm{M}}}\underline{\mathbf{u}}^{(k-1)} - r(\underline{\mathbf{u}}^{(k-1)}) = 0$$

ADER as DeC

$$\mathcal{L}^2(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\underline{\boldsymbol{u}}),$$
$$\mathcal{L}^1(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\boldsymbol{u}(t^n)).$$

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

$$\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k)} - \cancel{r(\boldsymbol{u}^{(k),0})} - \cancel{\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k-1)}} + \cancel{r(\boldsymbol{u}^{(k-1),0})} + \cancel{\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k-1)}} - r(\underline{\boldsymbol{u}}^{(k-1)}) = 0$$

ADER as DeC

$$\mathcal{L}^2(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\underline{\boldsymbol{u}}),$$
$$\mathcal{L}^1(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\boldsymbol{u}(t^n)).$$

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(k)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(k-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(k-1)}), \qquad k = 1, \dots, K,$$

$$\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k)} - \cancel{r(\boldsymbol{u}^{(k),0})} - \cancel{\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k-1)}} + \cancel{r(\boldsymbol{u}^{(k-1),0})} + \cancel{\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k-1)}} - r(\underline{\boldsymbol{u}}^{(k-1)}) = 0$$
$$\underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}}^{(k)} - r(\underline{\boldsymbol{u}}^{(k-1)}) = 0.$$

ADER as DeC

$$\mathcal{L}^2(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\underline{\boldsymbol{u}}),$$
$$\mathcal{L}^1(\underline{\boldsymbol{u}}) := \underline{\underline{\mathrm{M}}}\underline{\boldsymbol{u}} - r(\boldsymbol{u}(t^n)).$$

Apply the DeC Convergence theorem!

- $\mathcal{L}^1$ is coercive because $\underline{\underline{\mathrm{M}}}$ is always invertible
- $\mathcal{L}^1 - \mathcal{L}^2$ is Lipschitz with constant $C\Delta t$ because they are consistent approx of the same problem
- Hence, after K iterations we obtain a Kth order accurate approximation of $\underline{\boldsymbol{u}}^*$

$$\|u^{(K)} - u^*\| \le (C\Delta t)^K \|u^{(0)} - u^*\|$$

DeC as ADER

$$\mathcal{L}^2(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) := \begin{cases} \boldsymbol{u}^M - \boldsymbol{u}^0 - \sum_{r=0}^M \int_{t^0}^{t^M} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \\ \dots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 - \sum_{r=0}^M \int_{t^0}^{t^1} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \end{cases}.$$

? INTO WEAK FORM ?

DeC as ADER

DeC as ADER

DeC as ADER

$$\mathcal{L}^2(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) := \begin{cases} \boldsymbol{u}^M - \boldsymbol{u}^0 - \sum_{r=0}^M \int_{t^0}^{t^M} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \\ \dots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 - \sum_{r=0}^M \int_{t^0}^{t^1} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \end{cases}.$$

DeC as ADER

$$\mathcal{L}^2(\boldsymbol{u}^0,\dots,\boldsymbol{u}^M) := \begin{cases} \boldsymbol{u}^M - \boldsymbol{u}^0 - \sum_{r=0}^M \int_{t^0}^{t^M} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \\ \dots \\ \boldsymbol{u}^1 - \boldsymbol{u}^0 - \sum_{r=0}^M \int_{t^0}^{t^1} F(\boldsymbol{u}^r)\varphi_r(s)\mathrm{d}s \end{cases} .$$

$$\chi_{[t^0,t^m]}(t^m)\boldsymbol{u}^m - \chi_{[t^0,t^m]}(t_0)\boldsymbol{u}^0 - \int_{t^0}^{t^m} \chi_{[t^0,t^m]}(t)\sum_{r=0}^M F(\boldsymbol{u}^r)\varphi_r(t)\mathrm{d}t = 0$$

$$\int_{t^0}^{t^M} \chi_{[t^0,t^m]}(t)\partial_t\left(\boldsymbol{u}(t)\right)\mathrm{d}t - \int_{t^0}^{t^M} \chi_{[t^0,t^m]}(t)\sum_{r=0}^M F(\boldsymbol{u}^r)\varphi_r(t)\mathrm{d}t = 0,$$

$$\int_{T^n} \psi_m(t)\partial_t\boldsymbol{u}(t)\mathrm{d}t - \int_{T^n} \psi_m(t)F(\boldsymbol{u}(t))\mathrm{d}t = 0.$$

Runge Kutta vs DeC–ADER

Classical Runge Kutta (RK)

- One step method
- Internal stages

Explicit Runge Kutta

+ Simple to code
- Not easily generalizable to arbitrary order
- Stages > order

Implicit Runge Kutta

+ Arbitrarily high order
- Require nonlinear solvers for nonlinear systems
- May not converge

DeC – ADER

- One step method
- Internal subtimesteps + ITERATIONS
- Can be rewritten as explicit RK (for ODE)

+ Explicit
+ Simple to code
+ Iterations = order
+ Arbitrarily high order
- Large memory storage

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

Stability

Since ADER can be written as a DeC, the stability functions are given by the same formula as for DeC and the stability regions are the following.

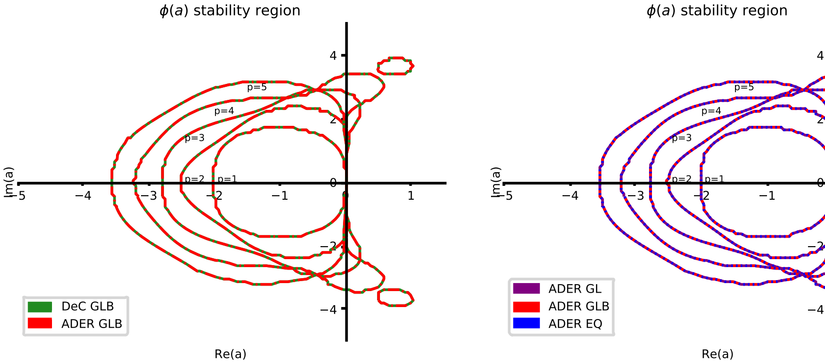


Figure: Stability region

Accuracy of ADER $\mathcal{L}^2$ operators

$\mathcal{L}^2 = \underline{\underline{M}}\, \underline{u} - \underline{r}(\underline{u})$? ORDER (IMPLICIT RK)

The two things that determine the accuracy of the ADER method are the iterations P and the accuracy of $\mathcal{L}^2$.

Accuracy of ADER $\mathcal{L}^2$ for different distributions

- Equispaced: boring, minimum accuracy possible $M+1$ nodes $p = M+1$
- Guass–Lobatto: this generates the LobattoIIIC methods, $M+1$ nodes $p = 2M$ — S stages $2S-2$ order
- Gauss–Legendre: this does not generate Gauss methods, $M+1$ nodes $p = 2M+1$ — S stages $2S-1$ order

ADER EXPLICIT

GLG $M+1$ NODES $\Rightarrow \mathcal{L}^2$ ORDER $2M+1$ $\Rightarrow K = 2M+1$ ✓

GLB $M+1$ NODES $\Rightarrow \mathcal{L}^2$ $2M$ $\Rightarrow K = 2M$ ✓

$\Rightarrow p = K$

$\mathcal{L}^2$ ADER as RK

EINSTEIN NOTATION

Here, we see $\mathcal{L}^2$ as an implicit RK

$$\mathcal{L}^{2,m}(\underline{\boldsymbol{u}}) \overset{\sum_j}{=} \underline{\underline{\mathrm{M}}}^m_j \boldsymbol{u}^{(j)} - \underline{\phi}^m(t^n)\boldsymbol{u}^n - \underbrace{\int_{T^n} \underline{\phi}^m(t)\underline{\phi}(t)_j dt}_{\Delta t \underline{\underline{\mathrm{R}}}^m_j} F(\boldsymbol{u}^{(j)}) = 0$$

$$(M^{-1})^z_m$$

$$\tilde{\mathcal{L}}^{2,z}(\underline{\boldsymbol{u}}) = \boldsymbol{u}^{(z)} - \overbrace{\underbrace{(\underline{\underline{\mathrm{M}}}^{-1})^z_m \underline{\phi}^m(t^n)}}^{1}\boldsymbol{u}^n - \Delta t\underbrace{(\underline{\underline{\mathrm{M}}}^{-1})^z_m \underline{\underline{\mathrm{R}}}^m_j} F(\boldsymbol{u}^{(j)}) = 0$$

RK $\quad \underline{\underline{\boldsymbol{u}}}^{(z)} = \underbrace{\boldsymbol{u}^n} + \Delta t \underbrace{a_{z,j}} F(\boldsymbol{u}^{(j)}) \quad \Rightarrow 1$

- $a_{mj} := (\underline{\underline{\mathrm{M}}}^{-1})^z_m \underline{\underline{\mathrm{R}}}^m_j$ ✓
- Prove that $\underbrace{(\underline{\underline{\mathrm{M}}}^{-1})^z_m \underline{\phi}^m(t^n) = 1}$ for every z ✓
- $c^m = \sum_r \underbrace{a_{mr}} = \underbrace{t^m}$?
- $b_r = \frac{1}{\Delta t}\int_{T^m} \phi_r(t)dt = w_r$ quadrature weights ✓

$$U^{n+1} = U^n + \frac{1}{\Delta t}\underbrace{\int_0^1 \phi_r(t)dt}\cdot F(U^r)$$

$$= w_r = b_r$$

$\mathcal{L}^2$ ADER as RK

$t^0, \dots, t^n$ => QUAD POINTS

- $(M^{-1})^z_j \phi^j(0) \stackrel{?}{=} 1^z \quad \forall z = 0, \dots, n$

$$\Longleftrightarrow \quad \underbrace{M^m_z (M^{-1})^z_j}_{I^m_j} \phi^j(0) \stackrel{?}{=} M^m_z 1^z \Longleftrightarrow \phi^m(0) \stackrel{?}{=} \underbrace{M^m_z \cdot 1^z}$$

$$\sum_z M^m_z \cdot 1^z = \sum_{z=0}^n M^m_z = \sum_{z=0}^n \phi_m(1)\phi_z(1) - \int_0^1 \underbrace{\phi'_m(t)}_{\mathbb{P}^{n-1}} \cdot \underbrace{\phi_z(t)}_{\mathbb{P}^n} dt = \quad \mathbb{P}^{n-1} \times \mathbb{P}^n \in \mathbb{P}^{2n-1}$$

RECALL THAT GLG, GLB QUADRATURE RULE IS EXACT POLY DEGREE $2S-3$

GLG: $2S-1$, GLB: $2S-3$; $2S-3 = 2(n+1)-3 = 2n-1$

RECALL LAGR. BASIS FUNCTION $\sum_{j=0}^n \phi_j(t) \equiv 1 \quad \forall t$

$$= \phi_m(1) \cdot 1 - \int_0^1 \underbrace{\phi'_m(t)}_{\mathbb{P}^{n-1}} dt = \phi_m(1) - [\phi_m(t)]_0^1 = \cancel{\phi_m(1)} - \cancel{\phi_m(1)} + \phi_m(0)$$

$\boxtimes$

$\mathcal{L}^2$ ADER as RK

$\sum_k a_{zk} = c_z = t^z$

2) $\sum_k (M^{-1})^z_j R^j_k \cdot 1^k \overset{?}{=} t^z \iff R^j_k \cdot 1^k \overset{?}{=} M^j_z \cdot t^z$

$$R^j_k \cdot 1^k = \sum_k \int_0^1 \phi_j(t) \underbrace{\phi_k(t)}_{=1} = \int_0^1 \phi_j(t)\, dt = \underline{w_j}$$

$$M^j_z t^z = \phi_j(1) \underbrace{\phi_z(1)}_{t=1} \cdot t^z - \int_0^1 \phi'_j(t) \cdot \overbrace{\phi_z(t) \cdot t^z}^{} \underset{\text{EXACT INTER}}{=} \phi_j(1) \cdot 1 - \int_0^1 \underbrace{\phi'_j(t) \cdot t}_{\in \mathbb{P}^M}\, dt$$

$$\sum_z \phi_z(t) \cdot t^z = \underline{t}$$

INTERPOLATING t NODES t^z $t \in \mathbb{P}$

I.B.P.

$$\underset{\text{EXACT QUAD}}{=} \phi_j(1) - [\phi_j(t) \cdot t]_0^1 + \int_0^1 \underbrace{\phi_j(t) \cdot 1}_{(t)^1}\, dt = \cancel{\phi_j(1)} - \cancel{\phi_j(1)} - 0 + w_j$$

BCD conditions (Butcher 1964)

$S \to 2S-2$ LOBATTO

$S \to 2S-1$ GAUSS-LEGENDRE

Define the conditions

B(2S-2) GLB
B(2S) GLG

$$B(p): \quad \sum_{i=1}^{s} b_i c_i^{z-1} = \frac{1}{z}, \qquad z = 1, \dots, p; \tag{20}$$

~~C(S)~~ C(S-1)

$$C(\eta): \quad \sum_{j=1}^{s} a_{ij} c_j^{z-1} = \frac{c_i^z}{z}, \qquad i = 1, \dots, s, \; z = 1, \dots, \eta; \tag{21}$$

D(S-1)

$$D(\zeta): \quad \sum_{i=1}^{s} b_i c_i^{z-1} a_{ij} = \frac{b_j}{z}(1 - c_j^z), \qquad j = 1, \dots, s, \; z = 1, \dots, \zeta. \tag{22}$$

Theorem (Butcher 1964)

If the coefficients b_i, c_i, a_{ij} of a RK scheme satisfy $B(p)$, $C(\eta)$ and $D(\zeta)$ with $p \leq \eta + \zeta + 1$ and $p \leq 2\eta + 2$, then the method is of order p.

$C(s-1)$ $D(s-1)$ S = # POINTS QUAD = LAGRANGE POINTS

Lemma

$\mathcal{L}^2$ operator of ADER defined by Gauss–Lobatto or Gauss–Legendre points and quadrature (they coincide) with $s = M+1$ stages satisfies $\underbrace{C(s-1)}$ and $\underbrace{D(s-1)}$.

Proof (1/4).

- Interpolation with ϕ^j is exact for polynomials of degree $s-1$. ✓
- The quadrature is exact for polynomials of degree $2s-3$.

$\underline{\underline{A}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}$

Recall that $\underline{\underline{A}} = \underline{\underline{M}}^{-1}\underline{\underline{R}}$, Condition $C(s-1)$ reads

$$\underline{\underline{A}}\,\underline{c}^{z-1} \overset{?}{=} \frac{1}{z}\underline{c}^z \iff \underline{\underline{R}}\,\underline{c}^{z-1} \overset{?}{=} \frac{1}{z}\underline{\underline{M}}\,\underline{c}^z \iff \underline{\mathcal{X}} := \underline{\underline{R}}\,\underline{c}^{z-1} - \frac{1}{z}\underline{\underline{M}}\,\underline{c}^z \overset{?}{=} \underline{0}, \qquad z = 1,\dots,s-1.$$

- Recall $\underline{c}_m = t^m$, $\underline{b}_m = w_m$, $\underline{\underline{R}}_{i,j} = \delta_{i,j}w_i$ and the definition of $\underline{\underline{M}}_{ij} = \phi_i(1)\phi_j(1) - \underbrace{\int_0^1 \phi_i' \phi_j}_{\text{exact}}$

$$\mathcal{X}_m := w_m\underbrace{(t^m)^{z-1}} - \frac{1}{z}\underbrace{\left(\phi^m(1)\underbrace{\phi^j(1)(t^j)^z} - \int_0^1 \frac{d}{d\xi}\phi^m(\xi)\underbrace{\phi^j(\xi)(t^j)^z}d\xi\right)}.$$

$$\sum_j \phi^j(\xi)\cdot(t^j)^z = \xi^z \in \mathbb{P}^{s-1} \checkmark$$

Proof (2/4).

Now, the interpolation of t^z with $z \leq s-1$ with basis functions ϕ^j is exact. Hence, we can substitute $\phi^j(\xi)(t^j)^z = \xi^z$ for all $z = 1, \dots, s-1$, obtaining

$$\mathcal{X}_m = w_m (t^m)^{z-1} - \frac{1}{z} \left(\phi^m(1)1^z - \int_0^1 \overbrace{\frac{d}{d\xi}\phi^m(\xi)}^{\mathbb{P}^{s-2}} \overbrace{\xi^z}^{\mathbb{P}^{s-1}} d\xi \right).$$

$[\phi^m \xi^z]_0^1$

Using the exactness of the quadrature for polynomials of degree $2s-3$, both true for Gauss–Lobatto and Gauss–Legendre, we know that the previous integral is exactly computed as $\frac{d}{d\xi}\phi^m(\xi)$ is of degree at most $s-2$ and ξ^z is at most $s-1$. So, we can use integration by parts and obtain

$$\mathcal{X}_m = w_m (t^m)^{z-1} - \frac{1}{z} \left(\cancel{\phi^m(0)0^z} + \int_0^1 \phi^m(\xi) \frac{d}{d\xi} \xi^z d\xi \right) = w_m (t^m)^{z-1} - \cancel{\frac{z}{z}} \overbrace{\int_0^1 \phi^m(\xi) \xi^{z-1} d\xi}^{w_m (\xi^m)^{z-1}} = 0 \checkmark$$

by the exactness of the quadrature rule and the definition of w_m. Note that the condition is sharp, since the interpolation is not anymore exact for $z = s$, hence $C(s)$ is not satisfied.

$C(s-1)$ ✓

$C(s-1)$ $D(s-1)$

Proof (3/4).

To prove $D(s-1)$, we write explicitly the condition in matricial form, for all $z=1,\dots,s-1$

$$\underline{\underbrace{bc^{z-1}}}\,\underline{\underline{A}} = \frac{1}{z}\underline{b(1-c^z)} \iff \underline{bc^{z-1}}\,\underline{\underline{\mathrm{M}}}^{-1}\,\underline{\underline{\mathrm{R}}} = \frac{1}{z}\underline{b(1-c^z)} \iff \underbrace{\underline{bc^{z-1}}} = \underbrace{\frac{1}{z}}\underbrace{\underline{b(1-c^z)}}\,\underline{\underline{\mathrm{R}}}^{-1}\,\underbrace{\underline{\underline{\mathrm{M}}}}.$$

Note that $b^m = w_m$ and $\underline{\underline{\mathrm{R}}}^m_r = w_m\delta^m_r$, so $\underline{\cancel{b}(1-c^z)}\,\cancel{\underline{\underline{\mathrm{R}}}^{-1}} = \underline{(1-c^z)}$. It is left to prove that

$$\mathcal{Y} := \underline{bc^{z-1}} - \frac{1}{z}\underline{(1-\overset{\phi}{c^z})\underline{\underline{\mathrm{M}}}} = \underline{0}.$$

$$\mathcal{Y}_m = w_m(t^m)^{z-1} - \frac{1}{z}\sum_{j=1}^{s}\underbrace{(1-(t^j)^z)\Big(\phi^j(1)\phi^m(1)}_{1-\xi^z \in \mathbb{P}^{s-1}} - \int_0^1 \underbrace{\frac{d}{d\xi}\phi^j(\xi)}_{\phi}\phi^m(\xi)d\xi\Big).$$

$\frac{d}{d\xi}(1-\xi^z) \in \mathbb{P}^{s-2}$

$C(s-1)$ $D(s-1)$

Proof (4/4).

Let us observe that, since $z \leq s-1$, the polynomial is exactly represented by the Lagrangian interpolation $t^z = \sum_{j=1}^s \phi(t)(t^m)^z$. Hence, using the exactness of the quadrature for polynomials of degree at most $2s-3$, we have

$$\begin{aligned}
\mathcal{Y}_m =& w_m(t^m)^{z-1} - \frac{1}{z}\overset{0}{\cancel{\left(1-(1)^z\right)}}\phi^m(1) + \frac{1}{z}\int_0^1 \frac{d}{d\xi}\left(1-(\xi)^z\right)\phi^m(\xi)d\xi \\
=& w_m(t^m)^{z-1} - \frac{1}{\cancel{z}}\int_0^1 \cancel{z}\,\xi^{z-1}\phi^m(\xi)d\xi = w_m(t^m)^{z-1} - w_m(t^m)^{z-1} = 0.
\end{aligned}$$

Hence, ADER-Legendre and ADER-Lobatto satisfy $D(s-1)$. Note that the condition is sharp, since the interpolation is not anymore exact for $z=s$, hence $D(s)$ is not satisfied.

ADER Gauss–Legendre $\mathcal{L}^2$

Remark (ADER-Legendre is no collocation method)

From the proof of previous Lemma, we can observe that ADER-Legendre methods do not satisfy $C(s)$, hence, the methods are not collocation methods and they do not coincide with Gauss–Legendre implicit RK methods.

ORDER 2S => C(S) => GLG ORDER < 2S

Theorem

$\mathcal{L}^2$ of ADER with Gauss–Legendre is of order $2s-1$.

Proof.

ADER-Legendre with $s = M+1$ stages satisfies $B(2s)$ for the quadrature rule and, hence, it satisfies $B(2s-1)$. For previous Lemma it also satisfies $C(s-1)$ and $D(s-1)$. Hence, Butcher's (1964) Theorem ($p \leq \eta + \zeta + 1$ and $p \leq 2\eta + 2$) guarantees that the method is of order $\underbrace{2s-1}_{p}$, since it is satisfied with $p = 2s-1$ and $\eta = \zeta = s-1$. □

$B(p) \quad C(\eta) \quad D(\zeta)$

$2S-1 \leq S-1 + S-1 + 1 = 2S-1$ ✓

$P = 2S-1 \leq 2(S-1)+2 = 2S$ ✓

ADER Gauss–Lobatto $\mathcal{L}^2$

Theorem

$\mathcal{L}^2$ of ADER with Gauss-Lobatto is of order $2s-2$.

Proof.

The condition for $B(2s-2)$ is satisfied as (c, b) is the Gauss–Lobatto quadrature with order $2s-2$. Previous Lemma guarantees that ADER-Lobatto satisfies $B(2s-2)$, $C(s-1)$ and $D(s-1)$, so Butcher's (1964) Theorem ($p \leq \eta + \zeta + 1$ and $p \leq 2\eta + 2$) is satisfied for order $p = 2s-2$ and $\eta = \zeta = s-1$. $\square$

$p = 2s-2 \leq s-1+s-1+1 = 2s-1$ ✓

$p = 2s-2 \leq 2(s-1)+2 = 2s$ ✓

ADER Gauss–Lobatto $\mathcal{L}^2$

Theorem

$\mathcal{L}^2$ of ADER with Gauss-Lobatto is LobattoIIIC.

The Lobatto IIIC method is defined using the condition

$$\longrightarrow \quad a_{i1} = \overset{w_1}{b_1} \qquad \text{for } i = 1, \dots, s. \quad + C(s-1) \quad \checkmark \tag{23}$$

Lemma

$\mathcal{L}^2$ of ADER with Gauss-Lobatto satisfies (23).

Theorem (Chipman 1971)

Lobatto IIIC schemes (in particular RK a_{ij}) are uniquely determined by Gauss–Lobatto quadrature rule (c, b), condition (23) *and by $C(s-1)$.*

ADER Gauss–Lobatto $\mathcal{L}^2$

Lemma

$\mathcal{L}^2$ of ADER with Gauss-Lobatto satisfies (23).

Proof.

$$a_{i1} = \sum_j (\underline{\underline{M}}^{-1})_{ij}\mathbb{R}_{j1} = b_1 = w_1 \iff$$

$$\sum_{i,j} \underline{\underline{M}}_{ki}(\underline{\underline{M}}^{-1})_{ij}\mathbb{R}_{j1} = \sum_i \underline{\underline{M}}_{ki} w_1 \iff$$

$$\underbrace{\delta_{k1} w_1} = \mathbb{R}_{k1} = \underbrace{\sum_i \underline{\underline{M}}_{ki} w_1}$$

$$\sum_i \underline{\underline{M}}_{ki} w_1 = \phi^m(1) w_1 - \int_0^1 \frac{d}{dt}\phi^m(\xi) w_1 dt = w_1 \phi^m(0) = w_1 \delta_{m,1}.$$

$\square$

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

Applications

DeC ADER

Usages

- Hyperbolic PDEs as explicit iterative methods (ADER: Toro, Dumbser, Klingenberg, Boscheri; DeC: Abgrall, Ricchiuto)
- IMEX solvers for hyperbolic with stiff sources (ADER: Dumbser, Boscheri; DeC: Abgrall, Torlo)
- IMEX solvers for hyperbolic with viscosity (treated implicitly) as compressible Navier Stokes (DeC: Minion, Dumbser, Zeifang)

IMEX

$\partial_t u = F(u) + S(u)$
$S(u)$ stiff to be treated implicitly

Advantages

- Arbitrary high order
- Unique framework to have matching between implicit and explicit terms
- Easy to code
- Iterative solver automatically included

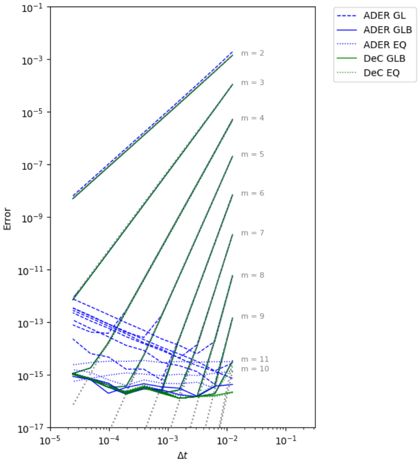
Disadvantages

- Explicit solver: many many stages
- Implicit: many stages
- Explicit: not amazing stability property (wrt SSP RK e.g.)

Convergence

$$
\begin{aligned}
&y'(t) = -|y(t)|y(t), \\
&y(0) = 1, \\
&t \in [0, 0.1].
\end{aligned} \tag{24}
$$

Convergence curves for ADER and DeC, varying the approximation order and collocation of nodes for the subtimesteps for a scalar nonlinear ODE



Lotka–Volterra

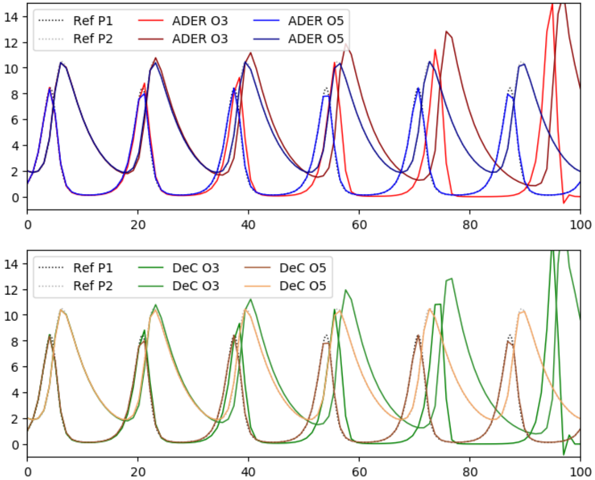
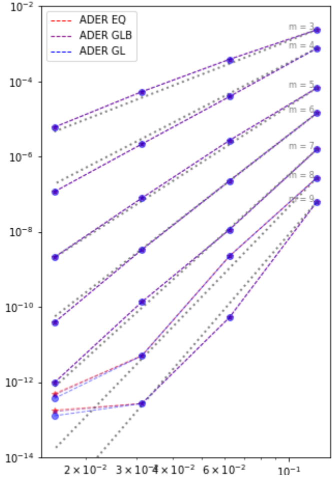


Figure: Numerical solution of the Lotka-Volterra system using ADER (top) and DeC (bottom) with Gauss-Lobatto nodes with timestep $\Delta T = 1$.

PDE: Burgers with spectral difference



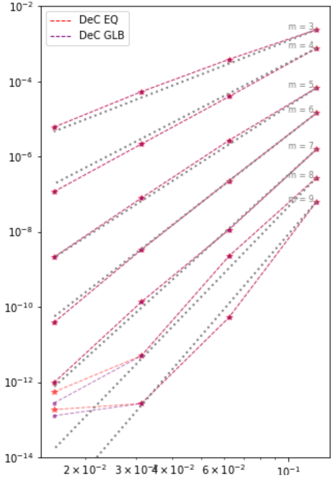


Figure: Convergence error for Burgers equations: Left ADER right DeC. Space

Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. **Efficient DeC (ADER)**
8. An efficient Deferred Correction

Reduce computational cost for explicit DeC

Literature

- *Micalizzi, L., Torlo, D. A new efficient explicit Deferred Correction framework: analysis and applications to hyperbolic PDEs and adaptivity.* arxiv.org/abs/2210.02976
- *Micalizzi, L., Torlo, D., Boscheri, W. Efficient iterative arbitrary high order methods: an adaptive bridge between low and high order.* arxiv.org/abs/2212.07783

Goal

Reduce computational costs of explicit DeC.

DeC as RK for ODEs

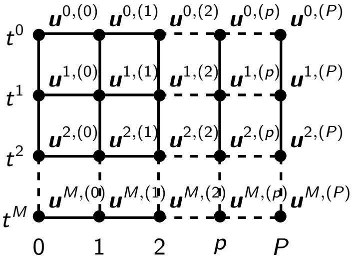
$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(p)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(p-1)}) \text{ with } p = 1, \dots, P.$$

$$\boldsymbol{u}^{m,(p)} = \boldsymbol{u}^0 + \sum_{r=0}^{M} \theta_r^m F(t^r, \boldsymbol{u}^{r,(p-1)}), \qquad \forall m = 1, \dots, M,\ p = 1, \dots, P$$

DeC as RK for ODEs

$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(p)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(p-1)}) \text{ with } p = 1, \dots, P.$$

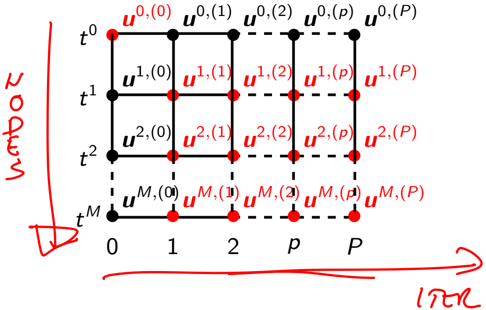
$$\boldsymbol{u}^{m,(p)} = \boldsymbol{u}^0 + \sum_{r=0}^{M} \theta_r^m F(t^r, \boldsymbol{u}^{r,(p-1)}), \qquad \forall m = 1, \dots, M, \ p = 1, \dots, P$$



DeC as RK for ODEs

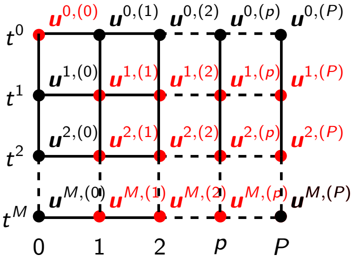
$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(p)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(p-1)}) \text{ with } p = 1, \dots, P.$$

$$\boldsymbol{u}^{m,(p)} = \boldsymbol{u}^0 + \sum_{r=0}^{M} \theta_r^m F(t^r, \boldsymbol{u}^{r,(p-1)}), \qquad \forall m = 1, \dots, M, \; p = 1, \dots, P$$



$$\mathcal{L}^1(\underline{\boldsymbol{u}}^{(p)}) = \mathcal{L}^1(\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}^2(\underline{\boldsymbol{u}}^{(p-1)}) \text{ with } p = 1, \dots, P.$$

$$\boldsymbol{u}^{m,(p)} = \boldsymbol{u}^0 + \sum_{r=0}^{M} \theta_r^m F(t^r, \boldsymbol{u}^{r,(p-1)}), \qquad \forall m = 1, \dots, M,\ p = 1, \dots, P$$



$\underline{c}$	$\mathbf{u}^0$	$\mathbf{u}^{(1)}$	$\mathbf{u}^{(2)}$	$\mathbf{u}^{(3)}$	$\cdots$	$\mathbf{u}^{(M-1)}$	$\mathbf{u}^{(M)}$	A
0	0							$\mathbf{u}^0$
$\underline{\beta}_{1:}$	$\underline{\beta}_{1:}$	$\underline{\underline{0}}$						$\mathbf{u}^{(1)}$
$\underline{\beta}_{1:}$	$\Theta_{1:,0}$	$\Theta_{1:,1:}$	$\underline{\underline{0}}$					$\mathbf{u}^{(2)}$
$\underline{\beta}_{1:}$	$\Theta_{1:,0}$	$\underline{\underline{0}}$	$\Theta_{1:,1:}$	$\underline{\underline{0}}$				$\mathbf{u}^{(3)}$
	$\vdots$	$\vdots$		$\ddots$	$\ddots$			$\vdots$
	$\vdots$	$\vdots$			$\ddots$	$\ddots$		$\vdots$
$\underline{\beta}_{1:}$	$\Theta_{1:,0}$	$\underline{\underline{0}}$	$\cdots$	$\cdots$	$\underline{\underline{0}}$	$\Theta_{1:,1:}$	$\underline{\underline{0}}$	$\mathbf{u}^{(M)}$
$\underline{b}$	$\Theta_{M,0}$	$\underline{0}$	$\cdots$	$\cdots$	$\cdots$	$\underline{0}$	$\Theta_{M,1:}$	$\mathbf{u}^{M,(M+1)}$

Costs

Large costs!

Costs

Large costs!

- DeC $S = M \cdot (P-1) + 1$
  - DeC equi $S = (P-1)^2 + 1$
  - DeC GLB $S = \left\lceil \frac{P}{2} \right\rceil (P-1) + 1$

Equispaced

P	M	DeC
2	1	2
3	2	5
4	3	10
5	4	17
6	5	26
7	6	37
8	7	50
9	8	65
10	9	82

Gauss–Lobatto

P	M	DeC
2	1	2
3	2	5
4	2	7
5	3	13
6	3	16
7	4	25
8	4	29
9	5	41
10	5	46

Costs

Large costs!

- DeC $S = M \cdot (P-1)+1$
  - DeC equi $S=(P-1)^2+1$
  - DeC GLB $S = \left\lceil \frac{P}{2} \right\rceil (P-1)+1$

Equispaced

P	M	DeC
2	1	2
3	2	5
4	3	10
5	4	17
6	5	26
7	6	37
8	7	50
9	8	65
10	9	82

Gauss–Lobatto

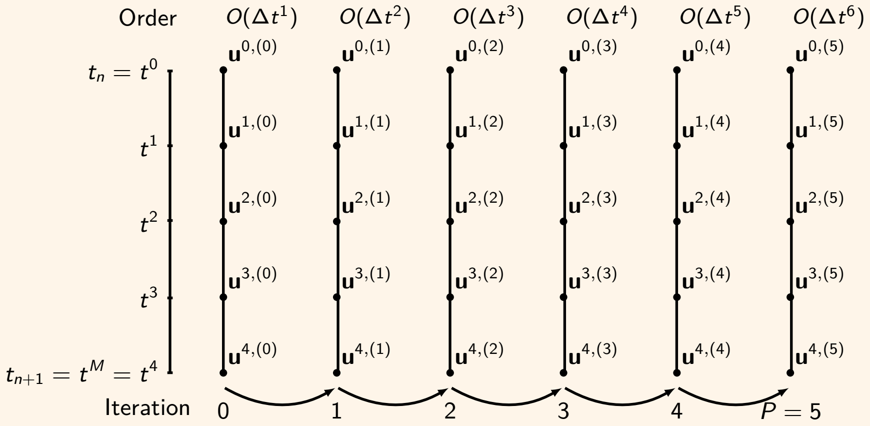
P	M	DeC
2	1	2
3	2	5
4	2	7
5	3	13
6	3	16
7	4	25
8	4	29
9	5	41
10	5	46

How can we save computational time?

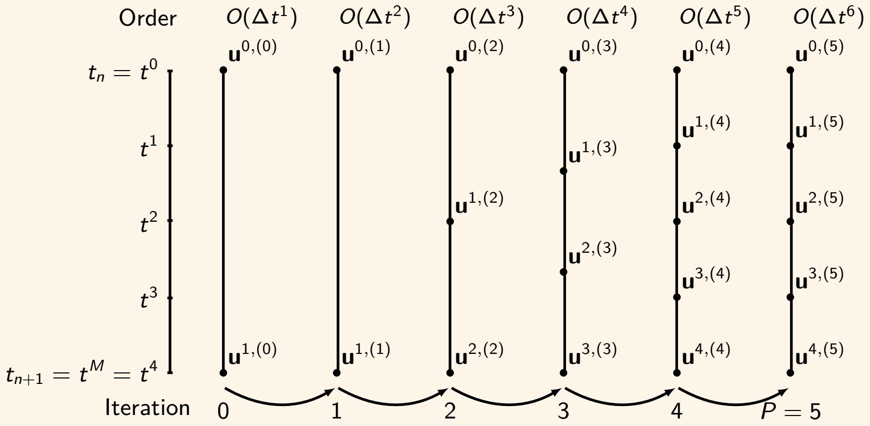
Outline

1. Motivation
2. DeC
3. ADER
4. Similarities
5. ADER stability and accuracy
6. Simulations
7. Efficient DeC (ADER)
8. An efficient Deferred Correction

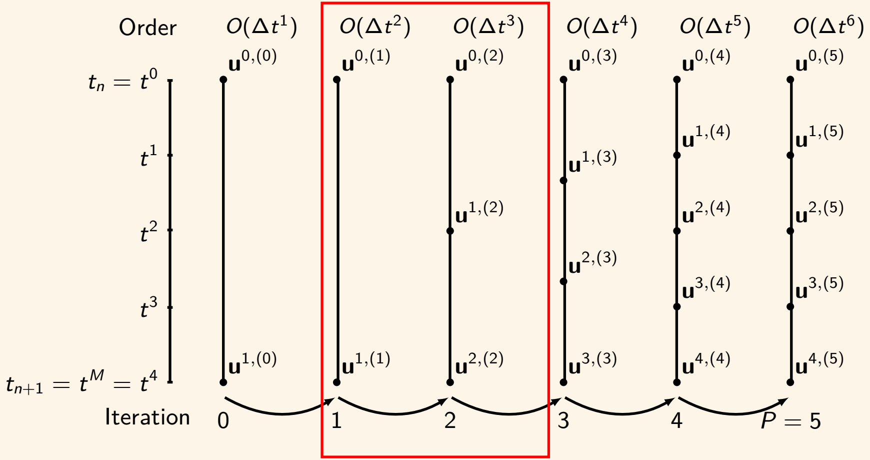
Idea for reduction of stages



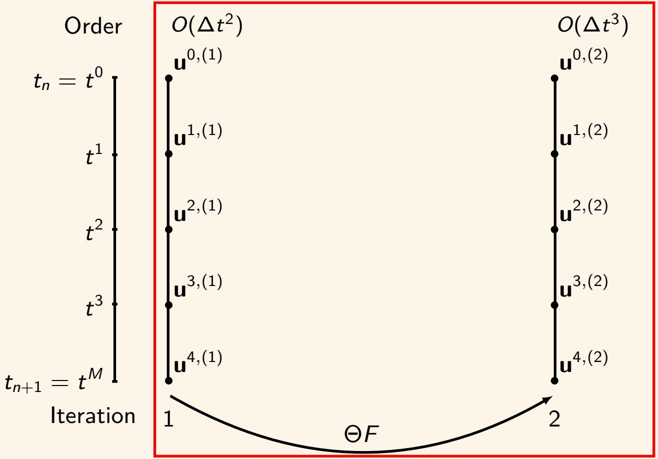
Idea for reduction of stages



Idea for reduction of stages



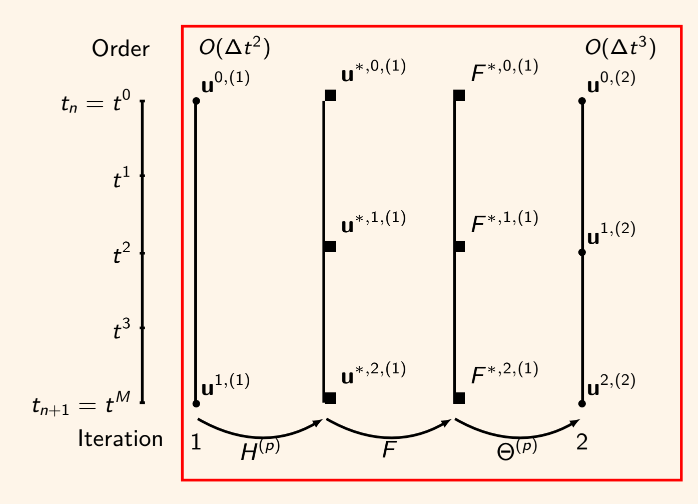
How to communicate between iterations?



DeC

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^{0} + \Delta t \Theta F(\underline{\boldsymbol{u}}^{(p-1)})$$

How to communicate between iterations?



DeC

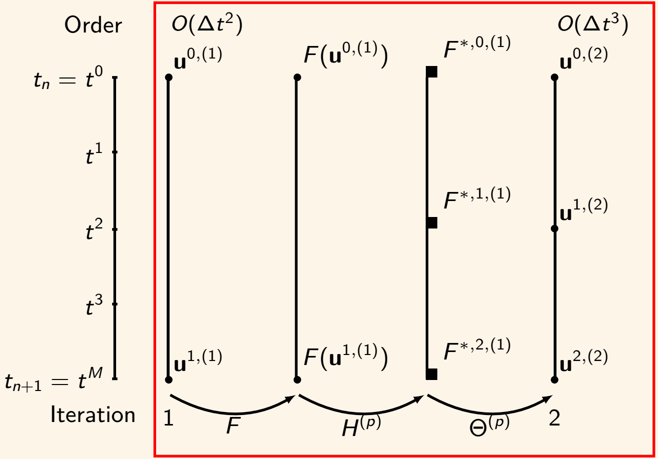
$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^{0} + \Delta t \Theta F(\underline{\boldsymbol{u}}^{(p-1)})$$

DeCu

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^{0} + \Delta t \Theta^{(p)} F(H^{(p)} \underline{\boldsymbol{u}}^{(p-1)})$$

$$H^{(p)}_{ij} = \phi_j^{(p-1)}(t^{i,(p)})$$

How to communicate between iterations?



DeC

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^0 + \Delta t \Theta F(\underline{\boldsymbol{u}}^{(p-1)})$$

DeCu

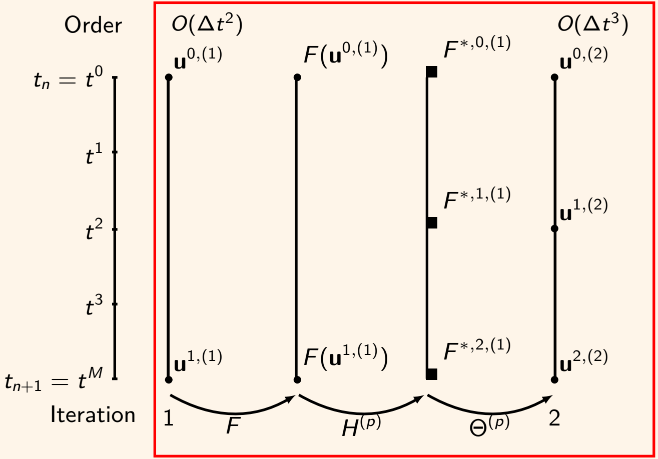
$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^0 + \Delta t \Theta^{(p)} F(H^{(p)} \underline{\boldsymbol{u}}^{(p-1)})$$

DeCdu

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^0 + \Delta t \Theta^{(p)} H^{(p)} F(\underline{\boldsymbol{u}}^{(p-1)})$$

$$H^{(p)}_{ij} = \phi_j^{(p-1)}(t^{i,(p)})$$

How to communicate between iterations?



DeC

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^0 + \Delta t \Theta F(\underline{\boldsymbol{u}}^{(p-1)})$$

DeCu

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^0 + \Delta t \Theta^{(p)} F(H^{(p)} \underline{\boldsymbol{u}}^{(p-1)})$$

$$\underline{\boldsymbol{u}}^{*(p)} = \underline{\boldsymbol{u}}^0 + \Delta t H^{(p)} \Theta^{*(p-1)} F(\underline{\boldsymbol{u}}^{*(p-1)})$$

DeCdu

$$\underline{\boldsymbol{u}}^{(p)} = \underline{\boldsymbol{u}}^0 + \Delta t \Theta^{(p)} H^{(p)} F(\underline{\boldsymbol{u}}^{(p-1)})$$

$$H_{ij}^{(p)} = \phi_j^{(p-1)}(t^{i,(p)})$$

DeC $\quad S = M \cdot (P-1) + 1$

$\underline{c}$	$\mathbf{u}^0$	$\mathbf{u}^{(1)}$	$\mathbf{u}^{(2)}$	$\mathbf{u}^{(3)}$	$\dots$	$\mathbf{u}^{(M-1)}$	$\mathbf{u}^{(M)}$	A	dim
0	0							$\mathbf{u}^0$	1
$\underline{\beta}_{1:}$	$\underline{\beta}_{1:}$	$\underline{\underline{0}}$						$\mathbf{u}^{(1)}$	M
$\underline{\beta}_{1:}$	$\Theta_{1:,0}$	$\Theta_{1:,1:}$	$\underline{\underline{0}}$					$\mathbf{u}^{(2)}$	M
$\underline{\beta}_{1:}$	$\Theta_{1:,0}$	$\underline{\underline{0}}$	$\Theta_{1:,1:}$	$\underline{\underline{0}}$				$\mathbf{u}^{(3)}$	M
	$\vdots$	$\vdots$		$\ddots$	$\ddots$			$\vdots$	M
	$\vdots$	$\vdots$			$\ddots$	$\ddots$		$\vdots$	M
$\underline{\beta}_{1:}$	$\Theta_{1:,0}$	$\underline{\underline{0}}$	$\cdots$	$\cdots$	$\underline{\underline{0}}$	$\Theta_{1:,1:}$	$\underline{\underline{0}}$	$\mathbf{u}^{(M)}$	M
$\underline{b}$	$\Theta_{M,0}$	$\underline{0}$	$\cdots$	$\cdots$	$\cdots$	$\underline{0}$	$\Theta_{M,1:}$	$\mathbf{u}^{M,(M+1)}$	

Efficient DeC into RK framework

$$\textbf{DeCu} \qquad S = M\cdot(P-1)+1-\frac{(M-1)(M-2)}{2}$$

$\underline{c}$	$\mathbf{u}^0$	$\mathbf{u}^{*(1)}$	$\mathbf{u}^{*(2)}$	$\mathbf{u}^{*(3)}$	$\dots$	$\mathbf{u}^{*(M-2)}$	$\mathbf{u}^{*(M-1)}$	$\mathbf{u}^{(M)}$	A	dim
0	0								$\mathbf{u}^0$	1
$\underline{\beta}^{(2)}_{1:}$	$\underline{\beta}^{(2)}_{1:}$	$\underline{\underline{0}}$							$\mathbf{u}^{*(1)}$	2
$\underline{\beta}^{(3)}_{1:}$	$W^{(2)}_{1:,0}$	$W^{(2)}_{1:,1:}$	$\underline{\underline{0}}$						$\mathbf{u}^{*(2)}$	3
$\underline{\beta}^{(4)}_{1:}$	$W^{(3)}_{1:,0}$	$\underline{\underline{0}}$	$W^{(3)}_{1:,1:}$	$\underline{\underline{0}}$					$\mathbf{u}^{*(3)}$	4
	$\vdots$	$\vdots$		$\ddots$	$\ddots$				$\vdots$	$\vdots$
	$\vdots$	$\vdots$			$\ddots$	$\ddots$			$\vdots$	$\vdots$
$\underline{\beta}^{(M)}_{1:}$	$W^{(M-1)}_{1:,0}$	$\underline{\underline{0}}$	$\cdots$	$\cdots$	$\underline{\underline{0}}$	$W^{(M-1)}_{1:,1:}$	$\underline{\underline{0}}$	$\underline{\underline{0}}$	$\mathbf{u}^{*(M-1)}$	M
$\underline{\beta}^{(M)}_{1:}$	$W^{(M)}_{1:,0}$	$\underline{\underline{0}}$	$\cdots$	$\cdots$	$\cdots$	$\underline{\underline{0}}$	$W^{(M)}_{1:,1:}$	$\underline{\underline{0}}$	$\mathbf{u}^{(M)}$	M
$\underline{b}$	$W^{(M+1)}_{M,0}$	$\underline{0}$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\underline{0}$	$W^{(M+1)}_{M,1:}$	$\mathbf{u}^{M,(M+1)}$	

$$W^{(p)} := \begin{cases} H^{(p)}\Theta^{(p)} \in \mathbb{R}^{(p+2)\times(p+1)}, & \text{if } p = 2,\dots,M-1, \\ \Theta^{(M)} \in \mathbb{R}^{(M+1)\times(M+1)}, & \text{if } p \geq M. \end{cases}$$

DeCdu $S = M \cdot (P-1) + 1 - \frac{M(M-1)}{2}$

$\underline{c}$	$\mathbf{u}^0$	$\mathbf{u}^{(1)}$	$\mathbf{u}^{(2)}$	$\mathbf{u}^{(3)}$	$\dots$	$\mathbf{u}^{(M-2)}$	$\mathbf{u}^{(M-1)}$	$\mathbf{u}^{(M)}$	A	dim
0	0								$\mathbf{u}^0$	1
$\underline{\beta}^{(1)}_{1:}$	$\underline{\beta}^{(1)}_{1:}$	$\underline{\underline{0}}$							$\mathbf{u}^{(1)}$	1
$\underline{\beta}^{(2)}_{1:}$	$Z^{(2)}_{1:,0}$	$Z^{(2)}_{1:,1:}$	$\underline{\underline{0}}$						$\mathbf{u}^{(2)}$	2
$\underline{\beta}^{(3)}_{1:}$	$Z^{(3)}_{1:,0}$	$\underline{\underline{0}}$	$Z^{(3)}_{1:,1:}$	$\underline{\underline{0}}$					$\mathbf{u}^{(3)}$	3
	$\vdots$	$\vdots$		$\ddots$	$\ddots$				$\vdots$	$\vdots$
	$\vdots$	$\vdots$			$\ddots$	$\ddots$			$\vdots$	$\vdots$
$\underline{\beta}^{(M-1)}_{1:}$	$Z^{(M-1)}_{1:,0}$	$\underline{\underline{0}}$	$\dots$	$\dots$	$\underline{\underline{0}}$	$Z^{(M-1)}_{1:,1:}$	$\underline{\underline{0}}$	$\underline{\underline{0}}$	$\mathbf{u}^{(M-1)}$	$M-1$
$\underline{\beta}^{(M)}_{1:}$	$Z^{(M)}_{1:,0}$	$\underline{\underline{0}}$	$\dots$	$\dots$	$\dots$	$\underline{\underline{0}}$	$Z^{(M)}_{1:,1:}$	$\underline{\underline{0}}$	$\mathbf{u}^{(M)}$	M
$\underline{b}$	$Z^{(M+1)}_{M,0}$	$\underline{0}$	$\dots$	$\dots$	$\dots$	$\dots$	$\underline{0}$	$Z^{(M+1)}_{M,1:}$	$\mathbf{u}^{M,(M+1)}$	

$$Z^{(p)} := \begin{cases} \Theta^{(p)} H^{(p-1)} \in \mathbb{R}^{(p+1)\times p}, & \text{if } p = 1, \dots, M, \\ \Theta^{(M)} \in \mathbb{R}^{(M+1)\times(M+1)}, & \text{if } p > M. \end{cases}$$

Computational costs reduction: RK stages

Equispaced

P	M	DeC	DeCu	DeCdu
2	1	2	2	2
3	2	5	5	4
4	3	10	9	7
5	4	17	14	11
6	5	26	20	16
7	6	37	27	22
8	7	50	35	29
9	8	65	44	37
10	9	82	54	46
11	10	101	65	56
12	11	122	77	67
13	12	145	90	79

Gauss-Lobatto

P	M	DeC	DeCu	DeCdu
2	1	2	2	2
3	2	5	5	4
4	2	7	7	6
5	3	13	12	10
6	3	16	15	13
7	4	25	22	19
8	4	29	26	23
9	5	41	35	31
10	5	46	40	36
11	6	61	51	46
12	6	67	57	52
13	7	85	70	64

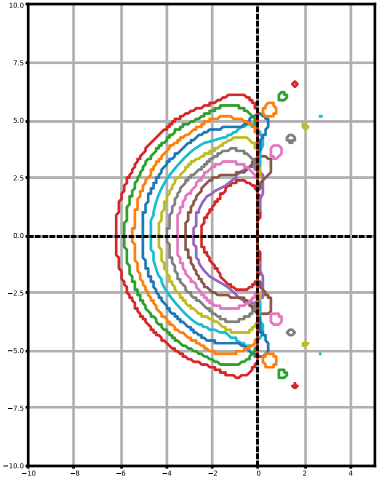
Stability Properties

DeC-DeCu-DeCdu

The **stability function** of DeC, DeCu, DeCdu of order P for any nodes distribution is

$$R(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^P}{P!}.$$

DeC, DeCu, DeCdu



Exercise

Efficient DeC

- Code DeCu or DeCdu
- Check order of accuracy
- Write a code to obtain its RK matrix
- Check the stability function with nodepy
- Compare computational costs with original DeC