Invariant-domain preserving Runge–Kutta methods

Alexandre Ern

ENPC and INRIA, Paris, France
joint work with Jean-Luc Guermond (TAMU)

FORTH Workshop, 09/2023

Outline

- Setting
- Beyond strong stability preserving (SSP) RK schemes
- New perspective on explicit RK schemes
- New perspective on implicit-explicit (IMEX) schemes

Outline

- Setting
- Beyond strong stability preserving (SSP) RK schemes
- New perspective on explicit RK schemes
- New perspective on implicit-explicit (IMEX) schemes

Warning: space discretization is hidden in the background, but is important!

Outline

- Setting
- Beyond strong stability preserving (SSP) RK schemes
- New perspective on explicit RK schemes
- New perspective on implicit-explicit (IMEX) schemes

Warning: space discretization is hidden in the background, but is important!

Main references:

- [AE & JLG, SISC 22] for ERK
- [AE & JLG, SISC 23] for IMEX

Setting

Cauchy problem

- Cauchy problem

$$\begin{cases} \partial_t \boldsymbol{u} + \nabla\cdot\boldsymbol{f}(\boldsymbol{u}) + \nabla\cdot\boldsymbol{g}(\boldsymbol{u}, \nabla\boldsymbol{u}) = \boldsymbol{S}(\boldsymbol{u}) & \text{in } D \times \mathbb{R}_+ \\ \boldsymbol{u}(\cdot, 0) = \boldsymbol{u}_0 & \text{in } D \end{cases}$$

$D \subset \mathbb{R}^d$ (open Lipschitz polytope)

Cauchy problem

- Cauchy problem

$$\begin{cases} \partial_t \boldsymbol{u} + \nabla\cdot\boldsymbol{f}(\boldsymbol{u}) + \nabla\cdot\boldsymbol{g}(\boldsymbol{u}, \nabla\boldsymbol{u}) = \boldsymbol{S}(\boldsymbol{u}) & \text{in } D \times \mathbb{R}_+ \\ \boldsymbol{u}(\cdot, 0) = \boldsymbol{u}_0 & \text{in } D \end{cases}$$

$D \subset \mathbb{R}^d$ (open Lipschitz polytope)

- Field $\boldsymbol{u}$ takes values in $\mathbb{R}^m$, i.e., $\boldsymbol{u} : D \times \mathbb{R}_+ \to \mathbb{R}^m$
- $\boldsymbol{f} \in C^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$: hyperbolic flux
- $\boldsymbol{g} \in C^1(\mathbb{R}^m \times \mathbb{R}^{m\times d}; \mathbb{R}^{m\times d})$: parabolic/diffusive flux
- $\boldsymbol{S} \in C^1(\mathbb{R}^m; \mathbb{R}^m)$: source/relaxation term

- Cauchy problem

$$\begin{cases} \partial_t \boldsymbol{u} + \nabla\cdot \boldsymbol{f}(\boldsymbol{u}) + \nabla\cdot \boldsymbol{g}(\boldsymbol{u}, \nabla \boldsymbol{u}) = \boldsymbol{S}(\boldsymbol{u}) & \text{in } D \times \mathbb{R}_+ \\ \boldsymbol{u}(\cdot, 0) = \boldsymbol{u}_0 & \text{in } D \end{cases}$$

$D \subset \mathbb{R}^d$ (open Lipschitz polytope)

- Field $\boldsymbol{u}$ takes values in $\mathbb{R}^m$, i.e., $\boldsymbol{u} : D \times \mathbb{R}_+ \to \mathbb{R}^m$
- $\boldsymbol{f} \in C^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$: hyperbolic flux
- $\boldsymbol{g} \in C^1(\mathbb{R}^m \times \mathbb{R}^{m\times d}; \mathbb{R}^{m\times d})$: parabolic/diffusive flux
- $\boldsymbol{S} \in C^1(\mathbb{R}^m; \mathbb{R}^m)$: source/relaxation term
- $\boldsymbol{u}_0$: admissible initial data
- BCs not discussed herein

Exemple 1: Scalar advection-diffusion-reaction

- Find $u : D \times \mathbb{R}_+ \longrightarrow \mathbb{R}$ such that

$$\partial_t u + \nabla\cdot \boldsymbol{f}(\boldsymbol{x}, u) + \nabla\cdot(\kappa(u)\nabla u) = S(u), \quad u(\cdot, 0) = u_0$$

- Hyperbolic and parabolic fluxes

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) := \begin{cases} (f_1(u), \dots, f_d(u)) \\ \boldsymbol{\beta}(\boldsymbol{x})u \text{ with } \nabla\cdot\boldsymbol{\beta} = 0 \end{cases} \qquad \boldsymbol{g}(u, \nabla u) := \kappa(u)\nabla u$$

- Source term, for instance

$$S(u) := \mu\phi(u)u(1-u), \qquad \phi \in C^0(\mathbb{R}; [-1, 1]), \qquad \mu \geq 0$$

Exemple 2: compressible Navier–Stokes

- Find $\boldsymbol{u} := (\rho, \boldsymbol{m}^\mathsf{T}, E)^\mathsf{T} : D \times \mathbb{R}_+ \longrightarrow \mathbb{R}^{d+2}$ such that

$$\begin{cases} \partial_t \rho + \nabla\cdot(\boldsymbol{v}\rho) = 0 \\ \partial_t \boldsymbol{m} + \nabla\cdot(\boldsymbol{v} \otimes \boldsymbol{m} + p(\boldsymbol{u})\mathbb{I} - \boldsymbol{s}(\boldsymbol{v})) = \boldsymbol{0} \\ \partial_t E + \nabla\cdot(\boldsymbol{v}(E + p(\boldsymbol{u})) - \boldsymbol{v}\cdot\boldsymbol{s}(\boldsymbol{v}) + \boldsymbol{q}(\boldsymbol{u})) = 0 \end{cases}$$

with velocity $\boldsymbol{v} := \boldsymbol{m}/\rho$ and pressure $p(\boldsymbol{u})$

Exemple 2: compressible Navier–Stokes

- Find $\boldsymbol{u} := (\rho, \boldsymbol{m}^\mathsf{T}, E)^\mathsf{T} : D \times \mathbb{R}_+ \to \mathbb{R}^{d+2}$ such that

$$\begin{cases} \partial_t \rho + \nabla\cdot(\boldsymbol{v}\rho) = 0 \\ \partial_t \boldsymbol{m} + \nabla\cdot(\boldsymbol{v} \otimes \boldsymbol{m} + p(\boldsymbol{u})\mathbb{I} - \boldsymbol{s}(\boldsymbol{v})) = \boldsymbol{0} \\ \partial_t E + \nabla\cdot(\boldsymbol{v}(E + p(\boldsymbol{u})) - \boldsymbol{v}\cdot\boldsymbol{s}(\boldsymbol{v}) + \boldsymbol{q}(\boldsymbol{u})) = 0 \end{cases}$$

  with velocity $\boldsymbol{v} := \boldsymbol{m}/\rho$ and pressure $p(\boldsymbol{u})$

- Hyperbolic (Euler) and parabolic fluxes

$$\boldsymbol{f}(\boldsymbol{u}) := \begin{pmatrix} \boldsymbol{v}\rho \\ \boldsymbol{v} \otimes \boldsymbol{m} + p(\boldsymbol{u})\mathbb{I} \\ \boldsymbol{v}(E + p(\boldsymbol{u})) \end{pmatrix}, \qquad \boldsymbol{g}(\boldsymbol{u}, \nabla \boldsymbol{u}) := \begin{pmatrix} 0 \\ -\boldsymbol{s}(\boldsymbol{v}) \\ -\boldsymbol{v}\cdot\boldsymbol{s}(\boldsymbol{v}) + \boldsymbol{q}(\boldsymbol{u}) \end{pmatrix}$$

  with viscous stress tensor and heat flux such that

$$\boldsymbol{s}(\boldsymbol{v}) = 2\mu \boldsymbol{e}(\boldsymbol{v}) + (\lambda - \tfrac{2}{3}\mu)(\nabla\cdot\boldsymbol{v})\mathbb{I}, \qquad \boldsymbol{q}(\boldsymbol{u}) = -\kappa \nabla T(\boldsymbol{u})$$

  with (linearized) strain tensor $\boldsymbol{e}(\boldsymbol{v}) := \frac{1}{2}(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^\mathsf{T})$ and temperature $T(\boldsymbol{u})$ (from specific internal energy $e(\boldsymbol{u}) := E/\rho - \frac{1}{2}\|\boldsymbol{v}\|_{\ell^2}^2$)

- μ, λ, κ constant for simplicity

Invariant domain

- Key assumption: There exists a convex subset $\mathcal{A} \subsetneq \mathbb{R}^m$ (depending on the initial data $\boldsymbol{u}_0$) s.t. the entropy/admissible solution to the Cauchy problem takes values in $\mathcal{A}$ for a.e. $(\boldsymbol{x}, t) \in D \times \mathbb{R}_+$

$$\left\{\boldsymbol{u}_0(\boldsymbol{x}) \in \mathcal{A} \text{ for a.e. } \boldsymbol{x} \in D\right\} \implies \left\{\boldsymbol{u}(\boldsymbol{x}, t) \in \mathcal{A} \text{ for a.e. } (\boldsymbol{x}, t) \in D \in \mathbb{R}_+\right\}$$

Invariant domain

- Key assumption: There exists a convex subset $\mathcal{A} \subsetneq \mathbb{R}^m$ (depending on the initial data $\boldsymbol{u}_0$) s.t. the entropy/admissible solution to the Cauchy problem takes values in $\mathcal{A}$ for a.e. $(\boldsymbol{x}, t) \in D \times \mathbb{R}_+$

$$\left\{\boldsymbol{u}_0(\boldsymbol{x}) \in \mathcal{A} \text{ for a.e. } \boldsymbol{x} \in D\right\} \implies \left\{\boldsymbol{u}(\boldsymbol{x}, t) \in \mathcal{A} \text{ for a.e. } (\boldsymbol{x}, t) \in D \in \mathbb{R}_+\right\}$$

- This is a generalization of the maximum principle

Invariant domain

- Key assumption: There exists a convex subset $\mathcal{A} \subsetneq \mathbb{R}^m$ (depending on the initial data $\boldsymbol{u}_0$) s.t. the entropy/admissible solution to the Cauchy problem takes values in $\mathcal{A}$ for a.e. $(\boldsymbol{x}, t) \in D \times \mathbb{R}_+$

$$\{\boldsymbol{u}_0(\boldsymbol{x}) \in \mathcal{A} \text{ for a.e. } \boldsymbol{x} \in D\} \implies \{\boldsymbol{u}(\boldsymbol{x}, t) \in \mathcal{A} \text{ for a.e. } (\boldsymbol{x}, t) \in D \in \mathbb{R}_+\}$$

- This is a generalization of the maximum principle

- Scalar conservation equations without reaction

$$\mathcal{A} := [\operatorname{ess\,inf} u_0, \operatorname{ess\,sup} u_0]$$

- Scalar conservation equations with $S(u) := \mu\phi(u)u(1-u)$

$$\mathcal{A} := [0, 1]$$

Invariant domain

- Key assumption: There exists a convex subset $\mathcal{A} \subsetneq \mathbb{R}^m$ (depending on the initial data $\boldsymbol{u}_0$) s.t. the entropy/admissible solution to the Cauchy problem takes values in $\mathcal{A}$ for a.e. $(\boldsymbol{x}, t) \in D \times \mathbb{R}_+$

$$\left\{\boldsymbol{u}_0(\boldsymbol{x}) \in \mathcal{A} \text{ for a.e. } \boldsymbol{x} \in D\right\} \implies \left\{\boldsymbol{u}(\boldsymbol{x}, t) \in \mathcal{A} \text{ for a.e. } (\boldsymbol{x}, t) \in D \in \mathbb{R}_+\right\}$$

- This is a generalization of the maximum principle

- Scalar conservation equations without reaction

$$\mathcal{A} := [\operatorname{ess\,inf} u_0, \operatorname{ess\,sup} u_0]$$

- Scalar conservation equations with $S(u) := \mu\phi(u)u(1-u)$

$$\mathcal{A} := [0, 1]$$

- Navier–Stokes and Euler equations ($s(\boldsymbol{u})$: specific entropy)

$$\mathcal{A}_{\text{NS}} := \{(\rho, \boldsymbol{m}^\mathsf{T}, E)^\mathsf{T} \in \mathbb{R}^m \mid 0 < \rho,\ 0 < T(\boldsymbol{u})\}$$
$$\mathcal{A}_{\text{Eu}} := \{(\rho, \boldsymbol{m}^\mathsf{T}, E)^\mathsf{T} \in \mathbb{R}^m \mid 0 < \rho,\ 0 < T(\boldsymbol{u}),\ \operatorname{ess\,inf} s(\boldsymbol{u}_0) \le s(\boldsymbol{u})\}$$

Invariant-domain preserving (IDP) approximation methods

- Approximation methods that preserve invariant domains are called Invariant domain preserving (IDP)

Invariant-domain preserving (IDP) approximation methods

- Approximation methods that preserve invariant domains are called Invariant domain preserving (IDP)
- Space semi-discrete problem: Find $\mathbf{U} \in C^1(\mathbb{R}_+; (\mathbb{R}^m)^I)$ s.t.

$$\mathbb{M}\partial_t \mathbf{U} = \mathbf{F}(\mathbf{U}) + \mathbf{G}(\mathbf{U}), \qquad \mathbf{U}(0) = \mathbf{U}_0$$

  - I: #dofs for space approximation (C^0-FEM, dG, FV, FD, ...)
  - $\mathbb{M}$: mass matrix (invertible)
  - $\boldsymbol{F} : (\mathbb{R}^m)^I \rightarrow (\mathbb{R}^m)^I$: space approximation of $-\nabla\cdot \boldsymbol{f}(\boldsymbol{u})$
  - $\boldsymbol{G} : (\mathbb{R}^m)^I \rightarrow (\mathbb{R}^m)^I$: space approximation of $-\nabla\cdot \boldsymbol{g}(\boldsymbol{u}, \nabla \boldsymbol{u}) + \boldsymbol{S}(\boldsymbol{u})$
  - $\mathbf{U}_i$ approximates $\boldsymbol{u}$ at some point $\boldsymbol{x}_i \in D \Longrightarrow$ natural requirement is $\mathbf{U} \in \mathcal{A}^I$

Invariant-domain preserving (IDP) approximation methods

- Approximation methods that preserve invariant domains are called Invariant domain preserving (IDP)

- Space semi-discrete problem: Find $\mathsf{U} \in C^1(\mathbb{R}_+; (\mathbb{R}^m)^I)$ s.t.

$$\mathbb{M}\partial_t \mathsf{U} = \mathsf{F}(\mathsf{U}) + \mathsf{G}(\mathsf{U}), \qquad \mathsf{U}(0) = \mathsf{U}_0$$

  - I: #dofs for space approximation (C^0-FEM, dG, FV, FD, ...)
  - $\mathbb{M}$: mass matrix (invertible)
  - $\boldsymbol{F} : (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$: space approximation of $-\nabla\cdot \boldsymbol{f}(\boldsymbol{u})$
  - $\boldsymbol{G} : (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$: space approximation of $-\nabla\cdot \boldsymbol{g}(\boldsymbol{u}, \nabla \boldsymbol{u}) + \boldsymbol{S}(\boldsymbol{u})$
  - U_i approximates $\boldsymbol{u}$ at some point $\boldsymbol{x}_i \in D \Longrightarrow$ natural requirement is $\mathsf{U} \in \mathcal{A}^I$

- Time-stepping scheme produces a sequence $\mathsf{U}^0, \mathsf{U}^1, \ldots, \mathsf{U}^n, \ldots$

- Time-stepping scheme is IDP if

$$\{\mathsf{U}_0 \in \mathcal{A}^I\} \implies \{\mathsf{U}^n \in \mathcal{A}^I \ \forall n \geq 0\}$$

How to achieve this goal?

SSP paradigm for hyperbolic problems

- Let us focus first on hyperbolic problems
- Key idea: [Shu & Osher 88] SSPRK are ERK methods where all updates are convex combinations of previous updates computed with forward Euler method (recall $\mathcal{A}$ convex)

SSP paradigm for hyperbolic problems

- Let us focus first on hyperbolic problems
- Key idea: [Shu & Osher 88] SSPRK are ERK methods where all updates are convex combinations of previous updates computed with forward Euler method (recall $\mathcal{A}$ convex)
- Key assumption: $\exists \tau^* > 0$ s.t. $\forall \tau \in (0, \tau^*]$,

$$\{\mathsf{V} \in \mathcal{A}^I\} \implies \{\mathsf{V} + \tau(\mathbb{M})^{-1}\mathsf{F}(\mathsf{V}) \in \mathcal{A}^I\}$$

In other words, $\mathcal{A}^I$ is invariant under the forward Euler method with CFL condition $\tau \in (0, \tau^*]$

SSP paradigm for hyperbolic problems

- Let us focus first on hyperbolic problems
- Key idea: [Shu & Osher 88] SSPRK are ERK methods where all updates are convex combinations of previous updates computed with forward Euler method (recall $\mathcal{A}$ convex)
- Key assumption: $\exists \tau^* > 0$ s.t. $\forall \tau \in (0, \tau^*]$,

$$\left\{ \mathsf{V} \in \mathcal{A}^I \right\} \implies \left\{ \mathsf{V} + \tau (\mathbb{M})^{-1} \mathsf{F}(\mathsf{V}) \in \mathcal{A}^I \right\}$$

  In other words, $\mathcal{A}^I$ is invariant under the forward Euler method with CFL condition $\tau \in (0, \tau^*]$

- Theory of SSP methods applied to ODEs is well understood [Kraaijevanger 91;S Ruuth & Spiteri 02; Ferracina & Spijker 05; Higueras 05]

Examples (for $\partial_t u = L(t, u)$)

- **Notation:** SSPRK(s, p) for s-stage, pth-order method
- SSPRK(2,2) (two-stage, second-order) [Heun's second-order method]

$$
\begin{aligned}
w^{(1)} &= u^n + \tau L(t^n, u^n) \\
u^{n+1} &= \tfrac{1}{2} u^n + \tfrac{1}{2}\left(w^{(1)} + \tau L(t^{n+1}, w^{(1)})\right)
\end{aligned}
$$

- SSPRK(3,3) (three-stage, third-order) [Fehlberg's method]

$$
\begin{aligned}
w^{(1)} &= u^n + \tau L(t^n, u^n) \\
w^{(2)} &= \tfrac{3}{4} u^n + \tfrac{1}{4}\left(w^{(1)} + \tau L(t^{n+1}, w^{(1)})\right) \\
u^{n+1} &= \tfrac{1}{3} u^n + \tfrac{2}{3}\left(w^{(2)} + \tau L(t^{n+\frac{1}{2}}, w^{(2)})\right)
\end{aligned}
$$

- SSPRK(4,3) and SSPRK(5,4) also available

Why (and how to) go beyond SSP?

Limitations of SSP paradigm (1/2)

- Restriction in accuracy: SSPRK are restricted to fourth-order (if one insists on never stepping backward in time) [Ruuth & Spiteri 02]

Limitations of SSP paradigm (1/2)

- Restriction in accuracy: SSPRK are restricted to fourth-order (if one insists on never stepping backward in time) [Ruuth & Spiteri 02]
- Difficult to accommodate implicit and explicit substeps
  - implicit RK schemes of order ≥ 2 cannot be SSP [Gottlieb, Shu, Tadmor 01]
  - explicit methods suffer from parabolic CFL restriction $\tau \leq ch^2$

Limitations of SSP paradigm (2/2)

- **Definition:** efficiency ratio of any s-stage ERK method
  - τ^*: maximal time step that makes forward Euler method IDP
  - $\tilde{\tau}$: maximal time step that makes s-stage ERK method IDP

$$c_{\text{eff}} := \frac{\tilde{\tau}}{s\tau^*} \qquad \text{(usually, } c_{\text{eff}} \leq 1\text{)}$$

Limitations of SSP paradigm (2/2)

- **Definition:** efficiency ratio of any s-stage ERK method
  - τ^*: maximal time step that makes forward Euler method IDP
  - $\tilde{\tau}$: maximal time step that makes s-stage ERK method IDP

$$c_{\text{eff}} := \frac{\tilde{\tau}}{s\tau^*} \qquad \text{(usually, } c_{\text{eff}} \leq 1\text{)}$$

- Do we care? Under the same CFL constraint, # flux evaluations to reach some T for s-stage ERK is $\frac{1}{c_{\text{eff}}}\times$ that for forward Euler method

Limitations of SSP paradigm (2/2)

- **Definition:** efficiency ratio of any s-stage ERK method
  - τ^*: maximal time step that makes forward Euler method IDP
  - $\tilde{\tau}$: maximal time step that makes s-stage ERK method IDP

$$c_{\text{eff}} := \frac{\tilde{\tau}}{s\tau^*} \qquad \text{(usually, } c_{\text{eff}} \leq 1\text{)}$$

- Do we care? Under the same CFL constraint, # flux evaluations to reach some T for s-stage ERK is $\frac{1}{c_{\text{eff}}}\times$ that for forward Euler method
- SSPRK methods are usually inefficient!
  - $c_{\text{eff}} = \frac{1}{2}$ for SSPRK(2,2)
  - $c_{\text{eff}} = \frac{1}{3}$ for SSPRK(3,3)
  - $c_{\text{eff}} = \frac{1}{2}$ for SSPRK(4,3)

Limitations of SSP paradigm (2/2)

- **Definition:** efficiency ratio of any s-stage ERK method
  - τ^*: maximal time step that makes forward Euler method IDP
  - $\tilde{\tau}$: maximal time step that makes s-stage ERK method IDP

$$c_{\text{eff}} := \frac{\tilde{\tau}}{s\tau^*} \qquad \text{(usually, } c_{\text{eff}} \leq 1\text{)}$$

- Do we care? Under the same CFL constraint, # flux evaluations to reach some T for s-stage ERK is $\frac{1}{c_{\text{eff}}}\times$ that for forward Euler method
- SSPRK methods are usually inefficient!
  - $c_{\text{eff}} = \frac{1}{2}$ for SSPRK(2,2)
  - $c_{\text{eff}} = \frac{1}{3}$ for SSPRK(3,3)
  - $c_{\text{eff}} = \frac{1}{2}$ for SSPRK(4,3)
- **Notation:** RK$(s,p;e)$ for s-stage, pth-order method, efficiency ratio e

SSPRK$(2,2;\frac{1}{2})$ SSPRK$(3,3;\frac{1}{3})$ SSPRK$(4,3;\frac{1}{2})$

Our contribution

- Introduce a new methodology that makes any ERK scheme IDP

Our contribution

- Introduce a new methodology that makes any ERK scheme IDP
- Introduce a new methodology that makes any IMEX scheme IDP

Our contribution

- Introduce a new methodology that makes any ERK scheme IDP
- Introduce a new methodology that makes any IMEX scheme IDP
- Benefits
  - employ optimally efficient methods
  - break order barriers
  - introduce IDP-IMEX schemes of order $p \geq 2$

Examples of optimally efficient ERK methods

- We will see that for an ERK-IDP scheme, maximal efficiency with $c_{\text{eff}} = 1$ is reached for equi-distributed sub-stages

Examples of optimally efficient ERK methods

- We will see that for an ERK-IDP scheme, maximal efficiency with $c_{\text{eff}} = 1$ is reached for equi-distributed sub-stages

- RK(2,2;1) (midpoint), RK(3,3;1) (Heun), RK(4,3;1) [fourth-order on linear pb.]

$$
\begin{array}{c|cc}
0 & 0 & \\
\frac{1}{2} & \frac{1}{2} & 0 \\
\hline
1 & 0 & 1
\end{array}
\qquad
\begin{array}{c|ccc}
0 & 0 & & \\
\frac{1}{3} & \frac{1}{3} & 0 & \\
\frac{2}{3} & 0 & \frac{2}{3} & 0 \\
\hline
1 & \frac{1}{4} & 0 & \frac{3}{4}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{4} & \frac{1}{4} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
\frac{3}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 \\
\hline
1 & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}
\end{array}
$$

- RK(5,4;1), RK(6,4;1) [fifth-order on linear pb.] and RK(7,5;1) [AE & JLG 22]

IDP ERK schemes

Peep under the hood of SSP (1/3)

- The beauty of SSP is that the forward Euler substep is a black box

Peep under the hood of SSP (1/3)

- The beauty of SSP is that the forward Euler substep is a black box
- This black box involves two fluxes (not just one as one might think)
  - low-order in space: flux $\mathbf{F}^{\mathrm{L}}$ and mass matrix $\mathbb{M}^{\mathrm{L}}$
  - high-order in space: flux $\mathbf{F}^{\mathrm{H}}$ and mass matrix $\mathbb{M}^{\mathrm{H}}$

Peep under the hood of SSP (1/3)

- The beauty of SSP is that the forward Euler substep is a black box
- This black box involves two fluxes (not just one as one might think)
  - low-order in space: flux $\mathbf{F}^{\mathrm{L}}$ and mass matrix $\mathbb{M}^{\mathrm{L}}$
  - high-order in space: flux $\mathbf{F}^{\mathrm{H}}$ and mass matrix $\mathbb{M}^{\mathrm{H}}$
- Some details

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},n+1} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n} + \tau\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n})$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},n+1} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^{n} + \tau\mathbf{F}^{\mathrm{H}}(\mathbf{U}^{n})$$

Peep under the hood of SSP (1/3)

- The beauty of SSP is that the forward Euler substep is a black box
- This black box involves two fluxes (not just one as one might think)
  - low-order in space: flux $\mathbf{F}^{\mathrm{L}}$ and mass matrix $\mathbb{M}^{\mathrm{L}}$
  - high-order in space: flux $\mathbf{F}^{\mathrm{H}}$ and mass matrix $\mathbb{M}^{\mathrm{H}}$
- Some details

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},n+1} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n} + \tau\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n})$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},n+1} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^{n} + \tau\mathbf{F}^{\mathrm{H}}(\mathbf{U}^{n})$$

Starting from $\mathbf{U}^n \in \mathcal{A}^I$,

- $\mathbf{U}^{\mathrm{L},n+1} \in \mathcal{A}^I$ under CFL, but is low-order accurate ...
- $\mathbf{U}^{\mathrm{H},n+1}$ departs from $\mathcal{A}^I$ but is high-order accurate ...

$\Longrightarrow$ employ a limiter to construct new update $\mathbf{U}^{n+1} \in \mathcal{A}^I$ as close as possible to $\mathbf{U}^{\mathrm{H},n+1}$

Peep under the hood of SSP (2/3)

- Let us formalize a little bit
- Assumption 1. [forward Euler with low-order flux is IDP under CFL condition] $\exists\tau^*$ s.t. $\forall\tau \in (0,\tau^*]$ and all $\mathbf{V} \in (\mathbb{R}^m)^I$,

$$\left\{\mathbf{V} \in \mathcal{A}^I\right\} \implies \left\{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\right\}$$

Peep under the hood of SSP (2/3)

- Let us formalize a little bit
- Assumption 1. [forward Euler with low-order flux is IDP under CFL condition]
  $\exists\tau^*$ s.t. $\forall\tau \in (0,\tau^*]$ and all $\mathbf{V} \in (\mathbb{R}^m)^I$,

$$\left\{\mathbf{V} \in \mathcal{A}^I\right\} \implies \left\{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\right\}$$

- Assumption 2. [nonlinear limiting operator]
  $\ell : \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$ s.t. for all $(\mathbf{V}, \mathbf{F}^{\mathrm{L}}, \mathbf{F}^{\mathrm{H}})$,

$$\left\{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\right\} \implies \left\{\ell(\mathbf{V}, \mathbf{F}^{\mathrm{L}}, \mathbf{F}^{\mathrm{H}}) \in \mathcal{A}^I\right\}$$

  Key idea: $\ell(\mathbf{V}, \mathbf{F}^{\mathrm{L}}, \mathbf{F}^{\mathrm{H}})$ is built as a convex combination of $\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}$ and $\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{H}}$

Peep under the hood of SSP (2/3)

- Let us formalize a little bit
- Assumption 1. [forward Euler with low-order flux is IDP under CFL condition]
  $\exists \tau^*$ s.t. $\forall \tau \in (0, \tau^*]$ and all $\mathbf{V} \in (\mathbb{R}^m)^I$,

$$\left\{\mathbf{V} \in \mathcal{A}^I\right\} \implies \left\{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\right\}$$

- Assumption 2. [nonlinear limiting operator]
  $\ell : \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$ s.t. for all $(\mathbf{V}, \mathbf{F}^{\mathrm{L}}, \mathbf{F}^{\mathrm{H}})$,

$$\left\{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\right\} \implies \left\{\ell(\mathbf{V}, \mathbf{F}^{\mathrm{L}}, \mathbf{F}^{\mathrm{H}}) \in \mathcal{A}^I\right\}$$

  Key idea: $\ell(\mathbf{V}, \mathbf{F}^{\mathrm{L}}, \mathbf{F}^{\mathrm{H}})$ is built as a convex combination of $\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}$ and $\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{H}}$

- Notice that both low/high-order updates start from the same vector $\mathbf{V}$

Peep under the hood of SSP (3/3)

- Given $\mathbf{U}^n$ in the invariant set $\mathcal{A}^I$
- The forward Euler step proceeds as follows:
  - compute low-order flux $\mathbf{F}^{\mathrm{L}}(\mathbf{U}^n)$
  - compute high-order flux $\mathbf{F}^{\mathrm{H}}(\mathbf{U}^n)$
  - compute update by limiting

$$\mathbf{U}^{n+1} := \ell(\mathbf{U}^n, \mathbf{F}^{\mathrm{L}}(\mathbf{U}^n), \mathbf{F}^{\mathrm{H}}(\mathbf{U}^n))$$

Peep under the hood of SSP (3/3)

- Given $\mathbf{U}^n$ in the invariant set $\mathcal{A}^I$
- The forward Euler step proceeds as follows:
  - compute low-order flux $\mathbf{F}^{\mathrm{L}}(\mathbf{U}^n)$
  - compute high-order flux $\mathbf{F}^{\mathrm{H}}(\mathbf{U}^n)$
  - compute update by limiting

$$\mathbf{U}^{n+1} := \ell(\mathbf{U}^n, \mathbf{F}^{\mathrm{L}}(\mathbf{U}^n), \mathbf{F}^{\mathrm{H}}(\mathbf{U}^n))$$

- **(Well-known) Proposition.** [Forward Euler is IDP]
  Let Assumptions 1 and 2 be met. Assume $\mathbf{U}^n \in \mathcal{A}^I$. Then, $\mathbf{U}^{n+1} \in \mathcal{A}^I$ for all $\tau \in (0, \tau^*]$

The two key ideas of IDP-ERK

- We are now ready to go high-order in time!

The two key ideas of IDP-ERK

- We are now ready to go high-order in time!
- Externalize the limiting process at each ERK stage

The two key ideas of IDP-ERK

- We are now ready to go high-order in time!
- Externalize the limiting process at each ERK stage
- Rewrite ERK scheme in incremental form: at each stage,
    - compute low/high-order updates using a common previous IDP-update
    - apply limiter

The two key ideas of IDP-ERK

- We are now ready to go high-order in time!
- Externalize the limiting process at each ERK stage
- Rewrite ERK scheme in incremental form: at each stage,
  - compute low/high-order updates using a common previous IDP-update
  - apply limiter
- Literature:
  - idea of externalizing the limiter proposed independently in [Kuzmin, Quezada de Luna, Ketcheson, Grüll, 22] for ERK and in [Quezada de Luna, Ketcheson 22] for DIRK
  - central idea of writing scheme in incremental form and maximizing efficiency only in [AE, JLG 22]
  - schemes with two time-derivatives [Gottlieb, Grant, Hu, Shu 22]

Butcher tableau of s-stage ERK method

- Generic form of Butcher tableau

$$
\begin{array}{c|ccccc}
c_1 & 0 & & & & \\
c_2 & a_{2,1} & 0 & & & \\
c_3 & a_{3,1} & a_{3,2} & 0 & & \\
\vdots & \vdots & & \ddots & \ddots & \\
c_s & a_{s,1} & a_{s,2} & \cdots & a_{s,s-1} & 0 \\
\hline
 & b_1 & b_2 & \cdots & b_{s-1} & b_s
\end{array}
$$

Butcher tableau of s-stage ERK method

- Generic form of Butcher tableau

$$
\begin{array}{c|ccccc}
c_1 & 0 & & & & \\
c_2 & a_{2,1} & 0 & & & \\
c_3 & a_{3,1} & a_{3,2} & 0 & & \\
\vdots & \vdots & & \ddots & \ddots & \\
c_s & a_{s,1} & a_{s,2} & \cdots & a_{s,s-1} & 0 \\
\hline
 & b_1 & b_2 & \cdots & b_{s-1} & b_s
\end{array}
$$

- Rename last line, set $c_1 := 0$ and $c_{s+1} := 1$

$$
\begin{array}{c|ccccc}
0 & 0 & & & & \\
c_2 & a_{2,1} & 0 & & & \\
c_3 & a_{3,1} & a_{3,2} & 0 & & \\
\vdots & \vdots & & \ddots & \ddots & \\
c_s & a_{s,1} & a_{s,2} & \cdots & a_{s,s-1} & 0 \\
\hline
1 & a_{s+1,1} & a_{s+1,2} & \cdots & a_{s+1,s-1} & a_{s+1,s}
\end{array}
$$

Butcher tableau of s-stage ERK method

- Generic form of Butcher tableau

$$\begin{array}{c|ccccc}
c_1 & 0 & & & & \\
c_2 & a_{2,1} & 0 & & & \\
c_3 & a_{3,1} & a_{3,2} & 0 & & \\
\vdots & \vdots & & \ddots & \ddots & \\
c_s & a_{s,1} & a_{s,2} & \cdots & a_{s,s-1} & 0 \\
\hline
 & b_1 & b_2 & \cdots & b_{s-1} & b_s
\end{array}$$

- Rename last line, set $c_1 := 0$ and $c_{s+1} := 1$

$$\begin{array}{c|ccccc}
\color{red}{0} & 0 & & & & \\
c_2 & a_{2,1} & 0 & & & \\
c_3 & a_{3,1} & a_{3,2} & 0 & & \\
\vdots & \vdots & & \ddots & \ddots & \\
c_s & a_{s,1} & a_{s,2} & \cdots & a_{s,s-1} & 0 \\
\hline
\color{red}{1} & \color{red}{a_{s+1,1}} & \color{red}{a_{s+1,2}} & \cdots & \color{red}{a_{s+1,s-1}} & \color{red}{a_{s+1,s}}
\end{array}$$

- Assume $c_k \geq 0$ for all $k \in \{1{:}s+1\}$
- For all $l \in \{2{:}s+1\}$, set

$$l'(l) := \max\{k < l \mid c_k \leq c_l\}$$

Think of $l'(l) := l - 1$ if sequence $(c_l)_{l \in \{1:s+1\}}$ is increasing

Details

- Let $\mathbf{U}^n \in \mathcal{A}^I$ and set $\mathbf{U}^{n,1} := \mathbf{U}^n$

Details

- Let $\mathbf{U}^n \in \mathcal{A}^I$ and set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- Loop over $l \in \{2{:}s+1\}$ (stage index)

Details

- Let $\mathbf{U}^n \in \mathcal{A}^I$ and set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- Loop over $l \in \{2{:}s+1\}$ (stage index)
- Compute low-order update starting from $\mathbf{U}^{n,l'}$ (think of $l' = l - 1$)

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},l} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n,l'} + \tau \underbrace{(c_l - c_{l'})\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n,l'})}_{:=\mathbf{\Phi}^{\mathrm{L}}}$$

Details

- Let $\mathbf{U}^n \in \mathcal{A}^I$ and set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- Loop over $l \in \{2{:}s+1\}$ (stage index)
- Compute low-order update starting from $\mathbf{U}^{n,l'}$ (think of $l' = l-1$)

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},l} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n,l'} + \tau \underbrace{(c_l - c_{l'})\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n,l'})}_{:=\mathbf{\Phi}^{\mathrm{L}}}$$

- Compute high-order update using same starting point $\mathbf{U}^{n,l'}$ (incremental form)

$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},l} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^{n,l'} + \tau \underbrace{\sum_{k\in\{1:l-1\}} (a_{l,k} - a_{l',k})\mathbf{F}^{\mathrm{H}}(\mathbf{U}^{n,k})}_{:=\mathbf{\Phi}^{\mathrm{H}}}$$

Details

- Let $\mathbf{U}^n \in \mathcal{A}^I$ and set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- Loop over $l \in \{2{:}s+1\}$ (stage index)
- Compute low-order update starting from $\mathbf{U}^{n,l'}$ (think of $l' = l - 1$)

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},l} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n,l'} + \tau \underbrace{(c_l - c_{l'})\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n,l'})}_{:=\boldsymbol{\Phi}^{\mathrm{L}}}$$

- Compute high-order update using same starting point $\mathbf{U}^{n,l'}$ (incremental form)

$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},l} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^{n,l'} + \tau \underbrace{\sum_{k\in\{1:l-1\}} (a_{l,k} - a_{l',k})\mathbf{F}^{\mathrm{H}}(\mathbf{U}^{n,k})}_{:=\boldsymbol{\Phi}^{\mathrm{H}}}$$

- Apply limiter: $\mathbf{U}^{n,l} := \ell(\mathbf{U}^{n,l'}, \boldsymbol{\Phi}^{\mathrm{L}}, \boldsymbol{\Phi}^{\mathrm{H}})$

Details

- Let $\mathbf{U}^n \in \mathcal{A}^I$ and set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- Loop over $l \in \{2{:}s+1\}$ (stage index)
- Compute low-order update starting from $\mathbf{U}^{n,l'}$ (think of $l' = l-1$)

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},l} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n,l'} + \tau \underbrace{(c_l - c_{l'})\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n,l'})}_{:=\mathbf{\Phi}^{\mathrm{L}}}$$

- Compute high-order update using same starting point $\mathbf{U}^{n,l'}$ (incremental form)

$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},l} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^{n,l'} + \tau \underbrace{\sum_{k\in\{1:l-1\}} (a_{l,k} - a_{l',k})\mathbf{F}^{\mathrm{H}}(\mathbf{U}^{n,k})}_{:=\mathbf{\Phi}^{\mathrm{H}}}$$

- Apply limiter: $\mathbf{U}^{n,l} := \ell(\mathbf{U}^{n,l'}, \mathbf{\Phi}^{\mathrm{L}}, \mathbf{\Phi}^{\mathrm{H}})$
- End of loop: return $\mathbf{U}^{n+1} := \mathbf{U}^{n,s+1}$

Main results

- **Theorem.** [IDP-ERK scheme]
  Let Assumptions 1 and 2 be met. Assume $\mathbf{U}^n \in \mathcal{A}^I$. Then, $\mathbf{U}^{n+1} \in \mathcal{A}^I$ (as well as all intermediate stages) for all

$$\tau \in (0, \tau^* / \max_{l \in \{2:s+1\}} (c_l - c_{l'})]$$

Main results

- **Theorem.** [IDP-ERK scheme]
  Let Assumptions 1 and 2 be met. Assume $\mathbf{U}^n \in \mathcal{A}^I$. Then, $\mathbf{U}^{n+1} \in \mathcal{A}^I$ (as well as all intermediate stages) for all

$$\tau \in (0, \tau^*/\max_{l\in\{2:s+1\}}(c_l - c_{l'})]$$

- **Corollary.** [Optimal efficiency]
  - $c_{\text{eff}} = 1/(s\max_{l\in\{2:s+1\}}(c_l - c_{l'}))$
  - optimal efficiency (with $c_{\text{eff}} = 1$) reached when points $(c_l)_{l\in\{1:s+1\}}$ are equi-distributed in $[0, 1]$

Examples: second- and third-order methods

- Some optimal methods: RK(2,2;1), RK(3,3;1), RK(4,3;1)

$$
\begin{array}{c|cc}
0 & 0 & \\
\frac{1}{2} & \frac{1}{2} & 0 \\
\hline
1 & 0 & 1
\end{array}
\qquad
\begin{array}{c|ccc}
0 & 0 & & \\
\frac{1}{3} & \frac{1}{3} & 0 & \\
\frac{2}{3} & 0 & \frac{2}{3} & 0 \\
\hline
1 & \frac{1}{4} & 0 & \frac{3}{4}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{4} & \frac{1}{4} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
\frac{3}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 \\
\hline
1 & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}
\end{array}
$$

Examples: second- and third-order methods

- Some optimal methods: RK(2,2;1), RK(3,3;1), RK(4,3;1)

$$
\begin{array}{c|cc}
0 & 0 & \\
\frac{1}{2} & \frac{1}{2} & 0 \\
\hline
1 & 0 & 1
\end{array}
\qquad
\begin{array}{c|ccc}
0 & 0 & & \\
\frac{1}{3} & \frac{1}{3} & 0 & \\
\frac{2}{3} & 0 & \frac{2}{3} & 0 \\
\hline
1 & \frac{1}{4} & 0 & \frac{3}{4}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{4} & \frac{1}{4} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
\frac{3}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 \\
\hline
1 & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}
\end{array}
$$

- Some non-optimal methods: SSPRK(2,2;$\frac{1}{2}$), SSPRK(3,3;$\frac{1}{3}$)

$$
\begin{array}{c|cc}
0 & 0 & \\
1 & 1 & 0 \\
\hline
 & \frac{1}{2} & \frac{1}{2}
\end{array}
\qquad
\begin{array}{c|ccc}
0 & 0 & & \\
1 & 1 & 0 & \\
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\
\hline
 & \frac{1}{6} & \frac{1}{6} & \frac{2}{3}
\end{array}
$$

Examples: fourth-order methods

- Two popular but sub-optimal methods: RK(4,4;$\frac{1}{2}$) and RK(4,4;$\frac{3}{4}$)

$$
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{2} & \frac{1}{2} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
1 & 0 & 0 & 1 & 0 \\
\hline
1 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & -\frac{1}{3} & 1 & 0 & \\
1 & 1 & -1 & 1 & 0 \\
\hline
1 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
$$

Examples: fourth-order methods

- Two popular but sub-optimal methods: RK(4,4;$\frac{1}{2}$) and RK(4,4;$\frac{3}{4}$)

$$
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{2} & \frac{1}{2} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
1 & 0 & 0 & 1 & 0 \\
\hline
1 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & -\frac{1}{3} & 1 & 0 & \\
1 & 1 & -1 & 1 & 0 \\
\hline
1 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
$$

- Optimal RK(5,4;1) and RK(6,4;1) devised in [AE & JLG 22]

  [both can be used within an IMEX scheme]

  RK(6,4;1) is fifth-order accurate on linear problems

Examples: fifth-order methods

- Butcher's method RK(6,5;$\frac{2}{3}$) (requires $c_6 = 1$)

$$
\begin{array}{c|cccccc}
0 & 0 & & & & & \\
\frac{1}{4} & \frac{1}{4} & 0 & & & & \\
\frac{1}{4} & \frac{1}{8} & \frac{1}{8} & 0 & & & \\
\frac{1}{2} & 0 & -\frac{1}{2} & 1 & 0 & & \\
\frac{3}{4} & \frac{3}{16} & 0 & 0 & \frac{9}{16} & 0 & \\
1 & -\frac{3}{7} & \frac{2}{7} & \frac{12}{7} & -\frac{12}{7} & \frac{8}{7} & 0 \\
\hline
1 & \frac{7}{90} & 0 & \frac{32}{90} & \frac{12}{90} & \frac{32}{90} & \frac{7}{90}
\end{array}
$$

Examples: fifth-order methods

- Butcher's method RK(6,5;$\frac{2}{3}$) (requires $c_6 = 1$)

$$
\begin{array}{c|cccccc}
0 & 0 \\
\frac{1}{4} & \frac{1}{4} & 0 \\
\frac{1}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\
\frac{1}{2} & 0 & -\frac{1}{2} & 1 & 0 \\
\frac{3}{4} & \frac{3}{16} & 0 & 0 & \frac{9}{16} & 0 \\
1 & -\frac{3}{7} & \frac{2}{7} & \frac{12}{7} & -\frac{12}{7} & \frac{8}{7} & 0 \\
\hline
1 & \frac{7}{90} & 0 & \frac{32}{90} & \frac{12}{90} & \frac{32}{90} & \frac{7}{90}
\end{array}
$$

- Novel RK(7,5;1) method [AE & JLG 22]

0	0				
$\frac{1}{7}$	0.1428571428571428	0			
$\frac{2}{7}$	0.0107112392440216	0.2750030464702641	0		
$\frac{3}{7}$	0.4812641640977338	−0.9634955610240432	0.9108028254977381	0	
$\frac{4}{7}$	0.3718168921589701	−0.5615016072648120	0.5590150320681445	0.2020982544662687	0
$\frac{5}{7}$	0.2210152091353413	0.3526985345185138	−0.8940286416537777	0.8097519357352928	...
$\frac{6}{7}$	0.2038005573304709	−0.4759394836772968	1.0938423462712870	−0.2853403360392873	...
1	0.0979996468518433	−0.0044680013474903	0.3592897484042552	0.0225280828210172	...

Methodology for numerical tests

- All the tests are done by fixing CFL $\in (0, 1]$ and setting

$$\tau := \text{CFL} \times s \times \tau^*$$

$\Longrightarrow$ all the methods perform the same number of flux evaluations and limiting operations independently of s
$\Longrightarrow$ each method is IDP at least up to CFL $\leq c_{\text{eff}}$

Methodology for numerical tests

- All the tests are done by fixing CFL $\in (0, 1]$ and setting

$$\tau := \text{CFL} \times s \times \tau^*$$

  $\Longrightarrow$ all the methods perform the same number of flux evaluations and limiting operations independently of s
  $\Longrightarrow$ each method is IDP at least up to CFL $\leq c_{\text{eff}}$

- Local maximum/minimum principle enforced at every dof
  (relaxation performed as in [Guermond, Popov, Tomas, 19])

- Global maximum/minimum principle strictly enforced

Methodology for numerical tests

- All the tests are done by fixing CFL $\in (0,1]$ and setting

$$\tau := \text{CFL} \times s \times \tau^*$$

$\Longrightarrow$ all the methods perform the same number of flux evaluations and limiting operations independently of s
$\Longrightarrow$ each method is IDP at least up to CFL $\leq c_{\text{eff}}$

- Local maximum/minimum principle enforced at every dof
(relaxation performed as in [Guermond, Popov, Tomas, 19])

- Global maximum/minimum principle strictly enforced

- Affine constraints defining $\mathcal{A}$: Flux-Corrected Transport (FCT)
[Boris & Book 73; Zalesak 79; Kuzmin, Löhner, Turek 12]

- Non-affine constraints: some nonlinear technique
[Sanders 88; Coquel & LeFloch 91; Liu & Osher 96; Zhang & Shu 11; Lohman & Kuzmin 16; Guermond, Nazarov, Popov, Tomas 18]

1D linear transport, 4th-order FD (1/3)

- Linear transport, $D := (0,1)$, periodic BCs

$$\partial_t u + \partial_x u = 0, \qquad u_0(x) := \begin{cases} \left(4\frac{(x-x_0)(x_1-x)}{(x_1-x_0)^2}\right)^6 & x \in (x_0, x_1) := (0.1, 0.4) \\ 0 & \text{otherwise} \end{cases}$$

- 4th order Finite Differences in space

1D linear transport, 4th-order FD (1/3)

- Linear transport, $D := (0,1)$, periodic BCs

$$\partial_t u + \partial_x u = 0, \qquad u_0(x) := \begin{cases} \left(4\frac{(x-x_0)(x_1-x)}{(x_1-x_0)^2}\right)^6 & x \in (x_0,x_1) := (0.1, 0.4) \\ 0 & \text{otherwise} \end{cases}$$

- 4th order Finite Differences in space
- In the L^1-norm, all the methods achieve the expected convergence order with CFL of the order of 0.5
- Let us look at the more challenging L^∞-error measure

1D linear transport, 4th-order FD (2/3)

- Second-order methods: RK(2,2;1) outperforms SSPRK(2,2;$\frac{1}{2}$)

	CFL = 0.2				CFL = 0.25			
I	RK$(2,2;1)$	rate	SSPRK$(2,2;\frac{1}{2})$	rate	RK$(2,2;1)$	rate	SSPRK$(2,2;\frac{1}{2})$	rate
50	4.72E-02	–	1.23E-01	–	4.91E-02	–	1.30E-01	–
100	2.81E-03	4.07	1.50E-02	3.03	4.51E-03	3.44	4.32E-02	1.60
200	1.16E-03	1.28	1.24E-03	3.60	2.01E-03	1.17	2.14E-03	4.34
400	3.38E-04	1.78	3.47E-04	1.84	5.41E-04	1.89	5.67E-04	1.91
800	8.79E-05	1.94	9.28E-05	1.90	1.38E-04	1.97	1.48E-04	1.94
1600	2.22E-05	1.98	2.33E-05	1.99	3.47E-05	1.99	3.78E-05	1.97
3200	5.58E-06	1.99	5.92E-06	1.98	8.73E-06	1.99	5.36E-05	-.50

1D linear transport, 4th-order FD (2/3)

- Second-order methods: RK(2,2;1) outperforms SSPRK(2,2;$\frac{1}{2}$)

	CFL = 0.2				CFL = 0.25			
I	RK(2, 2; 1)	rate	SSPRK(2, 2; $\frac{1}{2}$)	rate	RK(2, 2; 1)	rate	SSPRK(2, 2; $\frac{1}{2}$)	rate
50	4.72E-02	–	1.23E-01	–	4.91E-02	–	1.30E-01	–
100	2.81E-03	4.07	1.50E-02	3.03	4.51E-03	3.44	4.32E-02	1.60
200	1.16E-03	1.28	1.24E-03	3.60	2.01E-03	1.17	2.14E-03	4.34
400	3.38E-04	1.78	3.47E-04	1.84	5.41E-04	1.89	5.67E-04	1.91
800	8.79E-05	1.94	9.28E-05	1.90	1.38E-04	1.97	1.48E-04	1.94
1600	2.22E-05	1.98	2.33E-05	1.99	3.47E-05	1.99	3.78E-05	1.97
3200	5.58E-06	1.99	5.92E-06	1.98	8.73E-06	1.99	5.36E-05	-.50

- Third-order methods: SSPRK(3,3;$\frac{1}{3}$) behaves poorly, RK(4,3;1) performs best

	CFL = 0.05						CFL = 0.25					
I	RK(3,3;1)	rate	SSPRK(3,3;$\frac{1}{3}$)	rate	RK(4,3;1)	rate	RK(3,3;1)	rate	SSPRK(3,3;$\frac{1}{3}$)	rate	RK(4,3;1)	rate
50	5.15E-02	–	4.76E-02	–	5.15E-02	–	5.48E-02	–	1.55E-01	–	6.08E-02	–
100	5.41E-03	3.25	5.41E-03	3.14	5.41E-03	3.25	5.15E-03	3.41	6.12E-02	1.35	6.15E-03	3.31
200	3.79E-04	3.83	3.79E-04	3.83	3.79E-04	3.83	3.92E-04	3.72	1.07E-03	5.84	3.83E-04	4.01
400	2.27E-05	4.06	2.27E-05	4.06	2.27E-05	4.06	2.89E-05	3.76	2.18E-04	2.29	2.30E-05	4.06
800	1.58E-06	3.85	1.58E-06	3.85	1.58E-06	3.85	3.20E-06	3.18	6.41E-05	1.77	1.59E-06	3.85
1600	9.12E-08	4.12	1.22E-07	3.69	8.13E-08	4.28	8.23E-07	1.96	1.83E-05	1.81	8.25E-08	4.27
3200	1.52E-08	2.58	6.84E-08	0.84	5.31E-09	3.94	2.40E-07	1.78	5.39E-06	1.76	5.39E-09	3.94

1D linear transport, 4th-order FD (3/3)

- Fourth-order methods: RK(5,4;1) outperforms SSPRK(5,4;$\frac{1}{2}$)

	CFL = 0.05						CFL = 0.2					
I	RK(4,4;$\frac{1}{2}$)	rate	SSPRK(5,4;$\frac{1}{2}$)	rate	RK(5,4;1)	rate	RK(4,4;$\frac{1}{2}$)	rate	SSPRK(5,4;$\frac{1}{2}$)	rate	RK(5,4;1)	rate
50	4.32E-02	–	5.37E-02	–	5.95E-02	–	1.26E-01	–	5.63E-02	–	5.55E-02	–
100	5.41E-03	3.00	5.09E-03	3.40	5.09E-03	3.54	1.65E-02	2.93	7.82E-03	2.85	5.72E-03	3.28
200	3.79E-04	3.84	3.04E-04	4.07	3.04E-04	4.07	4.10E-04	5.33	3.80E-04	4.36	3.82E-04	3.90
400	2.27E-05	4.06	1.91E-05	3.99	1.91E-05	3.99	5.02E-05	3.03	2.27E-05	4.06	2.29E-05	4.06
800	1.58E-06	3.85	1.19E-06	4.00	1.19E-06	4.00	1.10E-05	2.19	1.79E-06	3.67	1.60E-06	3.84
1600	8.13E-08	4.28	7.45E-08	4.00	7.45E-08	4.00	2.70E-06	2.03	3.66E-07	2.29	8.26E-08	4.28
3200	5.36E-09	3.92	4.65E-09	4.00	4.65E-09	4.00	7.69E-07	1.81	9.29E-08	1.98	5.38E-09	3.94

1D linear transport, 4th-order FD (3/3)

- Fourth-order methods: RK(5,4;1) outperforms SSPRK(5,4;$\frac{1}{2}$)

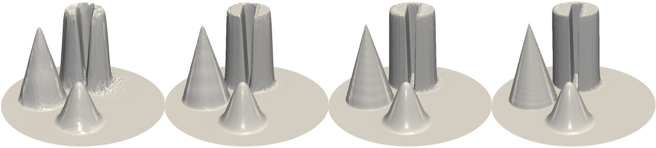
	CFL = 0.05						CFL = 0.2					
I	RK(4,4;$\frac{1}{2}$)	rate	SSPRK(5,4;$\frac{1}{2}$)	rate	RK(5,4;1)	rate	RK(4,4;$\frac{1}{2}$)	rate	SSPRK(5,4;$\frac{1}{2}$)	rate	RK(5,4;1)	rate
50	4.32E-02	–	5.37E-02	–	5.95E-02	–	1.26E-01	–	5.63E-02	–	5.55E-02	–
100	5.41E-03	3.00	5.09E-03	3.40	5.09E-03	3.54	1.65E-02	2.93	7.82E-03	2.85	5.72E-03	3.28
200	3.79E-04	3.84	3.04E-04	4.07	3.04E-04	4.07	4.10E-04	5.33	3.80E-04	4.36	3.82E-04	3.90
400	2.27E-05	4.06	1.91E-05	3.99	1.91E-05	3.99	5.02E-05	3.03	2.27E-05	4.06	2.29E-05	4.06
800	1.58E-06	3.85	1.19E-06	4.00	1.19E-06	4.00	1.10E-05	2.19	1.79E-06	3.67	1.60E-06	3.84
1600	8.13E-08	4.28	7.45E-08	4.00	7.45E-08	4.00	2.70E-06	2.03	3.66E-07	2.29	8.26E-08	4.28
3200	5.36E-09	3.92	4.65E-09	4.00	4.65E-09	4.00	7.69E-07	1.81	9.29E-08	1.98	5.38E-09	3.94

- Fifth-order methods: no SSP competitor!

	CFL = 0.02				CFL = 0.025			
I	RK(6,5;$\frac{1}{3}$)	rate	RK(7,5;1)	rate	RK(6,5;$\frac{2}{3}$)	rate	RK(7,5;1)	rate
50	5.19E-02	–	5.19E-02	–	5.19E-02	–	5.19E-02	–
100	5.41E-03	3.26	5.41E-03	3.26	5.41E-03	3.26	5.41E-03	3.26
200	3.79E-04	3.83	3.79E-04	3.83	3.79E-04	3.84	3.79E-04	3.83
400	2.27E-05	4.06	2.27E-05	4.06	2.27E-05	4.06	2.27E-05	4.06
800	1.58E-06	3.85	1.58E-06	3.85	1.58E-06	3.85	1.58E-06	3.85
1600	8.48E-08	4.22	8.13E-08	4.28	8.71E-08	4.18	8.13E-08	4.28
3200	7.10E-09	3.58	5.92E-09	3.78	1.16E-08	2.91	5.56E-09	3.87

Linear transport with non-smooth solution

- Three-solid problem with rotating advection field [Zalesak 79]
- Continuous $\mathbb{P}^1$-FEM on unstructured non-nested Delaunay meshes
- Solutions at $T = 1$ using RK(2,2;1) (midpoint rule) at CFL = 0.25
  [From left to right: $I = 6561$; $I = 24917$; $I = 98648$; $I = 389860$ dofs]



- Relative error in L^1-norm for RK(2,2;1) and RK(4,3;1)

I	RK(2,2;1)	rate	RK(4,3;1)	rate
1605	2.45E-01	–	2.49E-01	–
6561	1.28E-01	0.93	1.31E-01	0.92
24917	7.34E-02	0.81	7.49E-02	0.84
98648	4.26E-02	0.78	4.44E-02	0.76
389860	2.44E-02	0.81	2.56E-02	0.80

2D Burgers' equation (1/3)

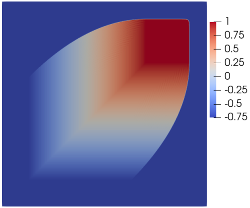
- 2D Burgers' equation in $D := (-.25, 1.75)^2$

$$\partial_t u + \nabla \cdot \boldsymbol{f}(u) = 0, \qquad \boldsymbol{f}(u) := \tfrac{1}{2}(u^2, u^2)^\mathsf{T}$$

with initial data

$$u_0(\boldsymbol{x}) := \begin{cases} 1 & \text{if } |x_1 - \frac{1}{2}| \le 1 \text{ and } |x_2 - \frac{1}{2}| \le 1 \\ -a & \text{otherwise} \end{cases}$$

- This problem exhibits many sonic points, which makes methods with too little low/high-order viscosity to fail [Guermond, Popov 17]
- Solution at $T = 0.65$ computed with RK(4,3;1) at CFL = 0.25 using 801^2 grid points



2D Burgers' equation (2/3)

- $T = 0.65$, CFL = 0.25, relative L^1-error for all the methods

I	RK(2,1;1)	rate	SSPRK(2,2; $\frac{1}{2}$)	rate
50	6.61E-02	–	6.70E-02	–
100	3.31E-02	1.00	3.34E-02	1.00
200	2.12E-02	0.65	2.12E-02	0.66
400	1.20E-02	0.82	1.16E-02	0.87
800	6.04E-03	0.99	5.73E-03	1.02

I	RK(3,3;1)	rate	SSPRK(3,3; $\frac{1}{3}$)	rate	RK(4,3;1)	rate
50	6.61E-02	–	6.74E-02	–	6.62E-02	–
100	3.31E-02	1.00	3.35E-02	1.01	3.31E-02	1.00
200	2.12E-02	0.65	2.13E-02	0.66	2.12E-02	0.65
400	1.20E-02	0.82	1.15E-02	0.89	1.20E-02	0.82
800	6.04E-03	0.99	5.72E-03	1.01	6.04E-03	0.99

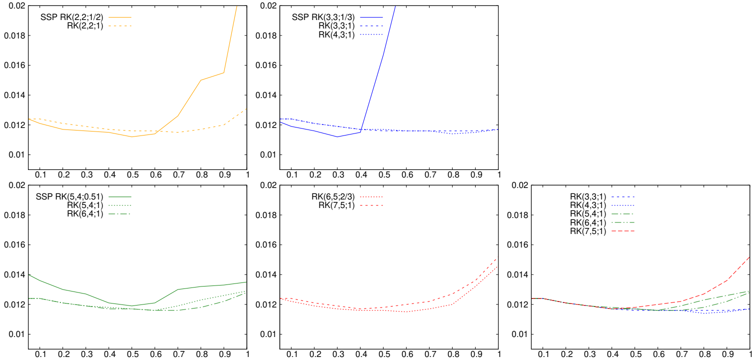
I	RK(4,4; $\frac{1}{2}$)	rate	RK(4,4; $\frac{3}{4}$)	rate	SSPRK(5,4; $\frac{1}{2}$)	rate	RK(5,4;1)	rate	RK(6,4;1)	rate
50	6.74E-02	–	6.63E-02	–	6.72E-02	–	6.63E-02	–	6.60E-02	–
100	3.35E-02	1.01	3.31E-02	1.00	3.43E-02	0.97	3.32E-02	1.00	3.30E-02	1.00
200	2.13E-02	0.66	2.11E-02	0.65	2.26E-02	0.60	2.12E-02	0.64	2.11E-02	0.64
400	1.17E-02	0.87	1.18E-02	0.84	1.28E-02	0.82	1.20E-02	0.82	1.20E-02	0.82
800	5.75E-03	1.02	5.84E-03	1.02	6.20E-03	1.05	6.06E-03	0.99	6.03E-03	0.99

I	RK(6,5; $\frac{2}{3}$)	rate	RK(7,5;1)	rate
50	6.65E-02	–	6.62E-02	–
100	3.32E-02	1.00	3.31E-02	1.00
200	2.11E-02	0.65	2.12E-02	0.65
400	1.18E-02	0.84	1.20E-02	0.82
800	5.79E-03	1.02	6.06E-03	0.99

- $\Longrightarrow$ at moderate CFL, all the methods converge equally well (all at order one)

2D Burgers' equation (3/3)

- Challenge methods by increasing CFL
- Results for second- and third-order methods (top), fourth-order, fifth-order methods plus a recap for all optimal methods



- $\Longrightarrow$ SSPRK (2,2) and SSPRK(3,3) start loosing accuracy at CFL ≈ 0.5, whereas IDP-ERK methods behave well over whole CFL range

Conclusions from numerical tests

- All IDP-ERK methods perform as well, and often better, than SSPRK methods of the same order

Conclusions from numerical tests

- All IDP-ERK methods perform as well, and often better, than SSPRK methods of the same order
- RK(2,2;1) (midpoint rule) outperforms popular SSPRK(2,2;$\frac{1}{2}$)
- RK(4,3;1) (vastly) outperforms popular SSPRK(3,3;$\frac{1}{3}$)

Conclusions from numerical tests

- All IDP-ERK methods perform as well, and often better, than SSPRK methods of the same order
- RK(2,2;1) (midpoint rule) outperforms popular SSPRK(2,2;$\frac{1}{2}$)
- RK(4,3;1) (vastly) outperforms popular SSPRK(3,3;$\frac{1}{3}$)
- The considered fourth-order methods provide comparable results

Conclusions from numerical tests

- All IDP-ERK methods perform as well, and often better, than SSPRK methods of the same order
- RK(2,2;1) (midpoint rule) outperforms popular SSPRK(2,2;$\frac{1}{2}$)
- RK(4,3;1) (vastly) outperforms popular SSPRK(3,3;$\frac{1}{3}$)
- The considered fourth-order methods provide comparable results
- Novel fifth-order IDP-ERK method with no SSP competitor

IDP IMEX schemes

Main ideas

- Consider low-order and high-order fluxes for
  - hyperbolic terms
  - parabolic (diffusion/relaxation) terms

Main ideas

- Consider low-order and high-order fluxes for
  - hyperbolic terms
  - parabolic (diffusion/relaxation) terms
- Quasi-linearization of parabolic fluxes (both low- and high-order)

Main ideas

- Consider low-order and high-order fluxes for
  - hyperbolic terms
  - parabolic (diffusion/relaxation) terms
- Quasi-linearization of parabolic fluxes (both low- and high-order)
- Key assumption: Under CFL condition, we have two IDP steps
  - forward Euler with low-order hyperbolic flux
  - backward Euler with low-order quasi-linear parabolic flux

Main ideas

- Consider low-order and high-order fluxes for
  - hyperbolic terms
  - parabolic (diffusion/relaxation) terms
- Quasi-linearization of parabolic fluxes (both low- and high-order)
- Key assumption: Under CFL condition, we have two IDP steps
  - forward Euler with low-order hyperbolic flux
  - backward Euler with low-order quasi-linear parabolic flux
- Rewrite IMEX scheme in incremental form

Main ideas

- Consider low-order and high-order fluxes for
  - hyperbolic terms
  - parabolic (diffusion/relaxation) terms
- Quasi-linearization of parabolic fluxes (both low- and high-order)
- Key assumption: Under CFL condition, we have two IDP steps
  - forward Euler with low-order hyperbolic flux
  - backward Euler with low-order quasi-linear parabolic flux
- Rewrite IMEX scheme in incremental form
- Apply (possibly distinct) limiters to hyperbolic and parabolic substeps

Butcher tableaux

- Explicit Butcher tableau

$$
\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathsf{e}}_{2,1} & 0 & & & \\
c_3 & a^{\mathsf{e}}_{3,1} & a^{\mathsf{e}}_{3,2} & 0 & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathsf{e}}_{s,1} & a^{\mathsf{e}}_{s,2} & \cdots & a^{\mathsf{e}}_{s,s-1} & 0 \\
\hline
1 & a^{\mathsf{e}}_{s+1,1} & a^{\mathsf{e}}_{s+1,2} & \cdots & a^{\mathsf{e}}_{s+1,s-1} & a^{\mathsf{e}}_{s+1,s}
\end{array}
$$

Butcher tableaux

- Explicit Butcher tableau

$$
\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathsf{e}}_{2,1} & 0 & & & \\
c_3 & a^{\mathsf{e}}_{3,1} & a^{\mathsf{e}}_{3,2} & 0 & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathsf{e}}_{s,1} & a^{\mathsf{e}}_{s,2} & \cdots & a^{\mathsf{e}}_{s,s-1} & 0 \\
\hline
1 & a^{\mathsf{e}}_{s+1,1} & a^{\mathsf{e}}_{s+1,2} & \cdots & a^{\mathsf{e}}_{s+1,s-1} & a^{\mathsf{e}}_{s+1,s}
\end{array}
$$

- Implicit Butcher tableau

$$
\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathsf{i}}_{2,1} & a^{\mathsf{i}}_{2,2} & & & \\
c_3 & a^{\mathsf{i}}_{3,1} & a^{\mathsf{i}}_{3,2} & a^{\mathsf{i}}_{3,3} & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathsf{i}}_{s,1} & a^{\mathsf{i}}_{s,2} & \cdots & a^{\mathsf{i}}_{s,s-1} & a^{\mathsf{i}}_{s,s} \\
\hline
1 & a^{\mathsf{i}}_{s+1,1} & a^{\mathsf{i}}_{s+1,2} & \cdots & a^{\mathsf{i}}_{s+1,s-1} & a^{\mathsf{i}}_{s+1,s}
\end{array}
$$

Butcher tableaux

- Explicit Butcher tableau

$$\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathrm{e}}_{2,1} & 0 & & & \\
c_3 & a^{\mathrm{e}}_{3,1} & a^{\mathrm{e}}_{3,2} & 0 & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathrm{e}}_{s,1} & a^{\mathrm{e}}_{s,2} & \cdots & a^{\mathrm{e}}_{s,s-1} & 0 \\
\hline
1 & a^{\mathrm{e}}_{s+1,1} & a^{\mathrm{e}}_{s+1,2} & \cdots & a^{\mathrm{e}}_{s+1,s-1} & a^{\mathrm{e}}_{s+1,s}
\end{array}$$

- Implicit Butcher tableau

$$\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathrm{i}}_{2,1} & a^{\mathrm{i}}_{2,2} & & & \\
c_3 & a^{\mathrm{i}}_{3,1} & a^{\mathrm{i}}_{3,2} & a^{\mathrm{i}}_{3,3} & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathrm{i}}_{s,1} & a^{\mathrm{i}}_{s,2} & \cdots & a^{\mathrm{i}}_{s,s-1} & a^{\mathrm{i}}_{s,s} \\
\hline
1 & a^{\mathrm{i}}_{s+1,1} & a^{\mathrm{i}}_{s+1,2} & \cdots & a^{\mathrm{i}}_{s+1,s-1} & a^{\mathrm{i}}_{s+1,s}
\end{array}$$

- Both tableaux share the same coefficients $(c_l)_{l\in\{1:s+1\}}$

Examples: second-order IMEX

- Heun + Crank–Nicolson: efficiency ratio is $\frac{1}{2}$

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

Examples: second-order IMEX

- Heun + Crank–Nicolson: efficiency ratio is $\frac{1}{2}$

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

- Explicit + implicit midpoint rules: efficiency ratio is 1

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

Examples: third-order IMEX (1/2)

- Three-stage, third-order method [Nørsett 74, Crouzeix 75] ($\gamma := \frac{1}{2} + \frac{1}{2\sqrt{3}} \approx 0.78867$)

$$\begin{array}{c|ccc} 0 & 0 & & \\ \gamma & \gamma & 0 & \\ 1-\gamma & \gamma-1 & 2-2\gamma & 0 \\ \hline 1 & 0 & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \begin{array}{c|ccc} 0 & 0 & & \\ \gamma & 0 & \gamma & \\ 1-\gamma & 0 & 1-2\gamma & \gamma \\ \hline 1 & 0 & \frac{1}{2} & \frac{1}{2} \end{array}$$

- Implicit method is A-stable, but efficiency ratio is only $\frac{1}{3}\gamma \approx 0.26$

Examples: third-order IMEX (1/2)

- Three-stage, third-order method [Nørsett 74, Crouzeix 75]
  ($\gamma := \frac{1}{2} + \frac{1}{2\sqrt{3}} \approx 0.78867$)

$$
\begin{array}{c|ccc}
0 & 0 & & \\
\gamma & \gamma & 0 & \\
1-\gamma & \gamma-1 & 2-2\gamma & 0 \\
\hline
1 & 0 & \frac{1}{2} & \frac{1}{2}
\end{array}
\qquad
\begin{array}{c|ccc}
0 & 0 & & \\
\gamma & 0 & \gamma & \\
1-\gamma & 0 & 1-2\gamma & \gamma \\
\hline
1 & 0 & \frac{1}{2} & \frac{1}{2}
\end{array}
$$

- Implicit method is A-stable, but efficiency ratio is only $\frac{1}{3}\gamma \approx 0.26$
- New scheme with optimal efficiency 1 [AE & JLG 22]

$$
\begin{array}{c|ccc}
0 & 0 & & \\
\frac{1}{3} & \frac{1}{3} & 0 & \\
\frac{2}{3} & 0 & \frac{2}{3} & 0 \\
\hline
1 & \frac{1}{4} & 0 & \frac{3}{4}
\end{array}
\qquad
\begin{array}{c|ccc}
0 & 0 & & \\
\frac{1}{3} & \frac{1}{3}-\gamma & \gamma & \\
\frac{2}{3} & \gamma & \frac{2}{3}-2\gamma & \gamma \\
\hline
1 & \frac{1}{4} & 0 & \frac{3}{4}
\end{array}
$$

- Implicit method has the same amplification function as above (and hence is A-stable)

Examples: third-order IMEX (2/2)

- Novel four-stage, third-order IMEX scheme with optimal efficiency 1 and implicit method is L-stable
- Explicit scheme is ERK(4,3;1) (already considered!)

$$
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{4} & \frac{1}{4} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
\frac{3}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 \\
\hline
1 & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}
\end{array}
$$

- Implicit scheme as follows:

$$
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{4} & -0.1858665215084591 & 0.4358665215084591 & & \\
\frac{1}{2} & -0.4367256409878701 & 0.5008591194794110 & 0.4358665215084591 & \\
\frac{3}{4} & -0.0423391342724147 & 0.7701152303135821 & -0.4136426175496265 & 0.4358665215084591 \\
\hline
1 & 0 & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}
\end{array}
$$

Examples: fourth-order IMEX

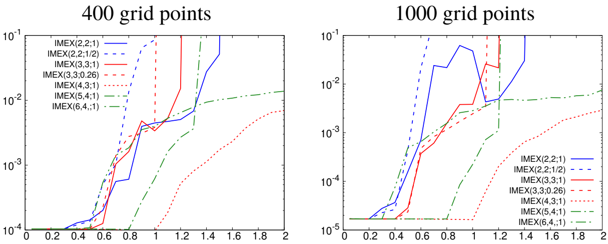
- Five- and six-stage schemes reviewed in [Carpenter & Kennedy 19]
- Novel five-stage scheme devised in [AE & JLG 22]
  - optimal efficiency 1
  - implicit scheme is singly diagonal and L-stable
- Novel six-stage scheme devised in [AE & JLG 22] with similar properties
  - the scheme is of linear order 5

Compressible Navier–Stokes equations, 1D

- Travelling viscous wave [Becker, 1922; Johnson, 13], $\Omega := [-0.5, 1]$, $T = 3$
- Ideal gas law, constant properties ($\mu = 0.01$, $Pr = 0.75$)
- Cumulated relative L^1-error on density, momentum and total energy
- Challenge all IMEX methods by increasing CFL

Compressible Navier–Stokes equations, 1D

- Travelling viscous wave [Becker, 1922; Johnson, 13], $\Omega := [-0.5, 1]$, $T = 3$
- Ideal gas law, constant properties ($\mu = 0.01$, $Pr = 0.75$)
- Cumulated relative L^1-error on density, momentum and total energy
- Challenge all IMEX methods by increasing CFL



- Main conclusions:
  - IMEX$(2,2;1)$ always outperforms IMEX$(2,2;\frac{1}{2})$
  - IMEX$(4,3;1)$ outperforms the other two third-order methods
  - IMEX$(6,4;1)$ slightly more robust than IMEX$(5,4;1)$

Compressible Navier–Stokes equations, 2D

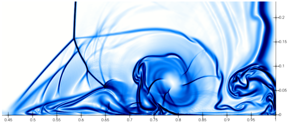
- Viscous shock tube problem [Daru & Tenaud, 01, 09]
- $\Omega := [0,1]\times[0,\frac{1}{2}]$, $T = 1$
- Ideal gas law, constant properties ($\mu = 0.001$, $Pr = 0.73$)
- $\mathbb{P}_1$ Lagrange FEM, IMEX(4, 3; 1) at CFL = 1.5

Compressible Navier–Stokes equations, 2D

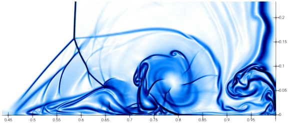
- Viscous shock tube problem [Daru & Tenaud, 01, 09]
- $\Omega := [0, 1] \times [0, \frac{1}{2}]$, $T = 1$
- Ideal gas law, constant properties ($\mu = 0.001$, $Pr = 0.73$)
- $\mathbb{P}_1$ Lagrange FEM, IMEX(4, 3; 1) at CFL $= 1.5$

- Density isocontours

572k grid points



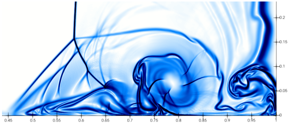
762k grid points



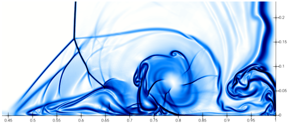
Compressible Navier–Stokes equations, 2D

- Viscous shock tube problem [Daru & Tenaud, 01, 09]
- $\Omega := [0, 1] \times [0, \frac{1}{2}]$, $T = 1$
- Ideal gas law, constant properties ($\mu = 0.001$, $Pr = 0.73$)
- $\mathbb{P}_1$ Lagrange FEM, IMEX$(4, 3; 1)$ at CFL $= 1.5$
- Density isocontours

572k grid points



762k grid points



- Numerical tests using non-ideal gas laws in progress

Thank you for your attention!

Euler IDP-IMEX scheme

- Gentle introduce ideas on Euler IDP-IMEX scheme

Euler IDP-IMEX scheme

- Gentle introduce ideas on Euler IDP-IMEX scheme
- F^{L}: Low-order approx. of hyperbolic flux $-\nabla\cdot\boldsymbol{f}(\boldsymbol{u})$
- $\mathsf{G}^{\mathrm{L,lin}}(\mathsf{W}^n;\cdot)$: Low-order quasi-linear approx. of parabolic flux $-\nabla\cdot\boldsymbol{g}(\boldsymbol{u},\nabla\boldsymbol{u})+\boldsymbol{S}(\boldsymbol{u})$

Euler IDP-IMEX scheme

- Gentle introduce ideas on Euler IDP-IMEX scheme
- $\mathbf{F}^{\mathrm{L}}$: Low-order approx. of hyperbolic flux $-\nabla\cdot\boldsymbol{f}(\boldsymbol{u})$
- $\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^n;\cdot)$: Low-order quasi-linear approx. of parabolic flux $-\nabla\cdot\boldsymbol{g}(\boldsymbol{u},\nabla\boldsymbol{u}) + \boldsymbol{S}(\boldsymbol{u})$
- Consider low-order quasi-linear update

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},n+1} = \underbrace{\mathbb{M}^{\mathrm{L}}\mathbf{U}^n + \tau\mathbf{F}^{\mathrm{L}}(\mathbf{U}^n)}_{=:\mathbb{M}^{\mathrm{L}}\mathbf{W}^{\mathrm{L},n}} + \tau\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{\mathrm{L},n};\mathbf{U}^{\mathrm{L},n+1})$$

Euler IDP-IMEX scheme

- Gentle introduce ideas on Euler IDP-IMEX scheme
- $\mathbf{F}^{\mathrm{L}}$: Low-order approx. of hyperbolic flux $-\nabla\cdot \boldsymbol{f}(\boldsymbol{u})$
- $\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^n;\cdot)$: Low-order quasi-linear approx. of parabolic flux $-\nabla\cdot \boldsymbol{g}(\boldsymbol{u},\nabla \boldsymbol{u}) + \boldsymbol{S}(\boldsymbol{u})$
- Consider low-order quasi-linear update

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},n+1} = \underbrace{\mathbb{M}^{\mathrm{L}}\mathbf{U}^n + \tau \mathbf{F}^{\mathrm{L}}(\mathbf{U}^n)}_{=:\mathbb{M}^{\mathrm{L}}\mathbf{W}^{\mathrm{L},n}} + \tau \mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{\mathrm{L},n};\mathbf{U}^{\mathrm{L},n+1})$$

- This can be decomposed as
  - hyperbolic sub-step (explicit update):

$$\mathbf{W}^{\mathrm{L},n} := \mathbf{U}^n + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{U}^n)$$

  - parabolic sub-step (quasi-linear solve):

$$\mathbf{U}^{\mathrm{L},n+1} := \big(\mathbb{I} - \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{\mathrm{L},n};\cdot)\big)^{-1}(\mathbf{W}^{\mathrm{L},n})$$

Key assumption on low-order fluxes

- Assumption 1. There exists $\tau^* > 0$ s.t. for all $\tau \in (0, \tau^*]$,
  - forward Euler with low-order hyperbolic flux is IDP:

$$\{\mathsf{V} \in \mathcal{A}^I\} \implies \{\mathsf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathsf{F}^{\mathrm{L}}(\mathsf{V}) \in \mathcal{A}^I\}$$

Key assumption on low-order fluxes

- Assumption 1. There exists $\tau^* > 0$ s.t. for all $\tau \in (0, \tau^*]$,
  - forward Euler with low-order hyperbolic flux is IDP:

$$\left\{\mathbf{V} \in \mathcal{A}^I\right\} \implies \left\{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\right\}$$

  - backward Euler with low-order quasi-linear parabolic flux is IDP: For all $\mathbf{W} \in \mathcal{A}^I$, $\mathbb{I} - \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}; \cdot) : (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$ is bijective and

$$\left\{\mathbf{V} \in \mathcal{A}^I\right\} \implies \left\{\left(\mathbb{I} - \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{G}^{\mathrm{L,lin}}(\mathbf{V}; \cdot)\right)^{-1}(\mathbf{V}) \in \mathcal{A}^I\right\}$$

Notice that quasi-linearization is performed at $\mathbf{V}$

Key assumption on low-order fluxes

- Assumption 1. There exists $\tau^* > 0$ s.t. for all $\tau \in (0, \tau^*]$,
  - forward Euler with low-order hyperbolic flux is IDP:

$$\{\mathbf{V} \in \mathcal{A}^I\} \implies \{\mathbf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{F}^{\mathrm{L}}(\mathbf{V}) \in \mathcal{A}^I\}$$

  - backward Euler with low-order quasi-linear parabolic flux is IDP: For all $\mathbf{W} \in \mathcal{A}^I$, $\mathbb{I} - \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W};\cdot) : (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$ is bijective and

$$\{\mathbf{V} \in \mathcal{A}^I\} \implies \{\left(\mathbb{I} - \tau(\mathbb{M}^{\mathrm{L}})^{-1}\mathbf{G}^{\mathrm{L,lin}}(\mathbf{V};\cdot)\right)^{-1}(\mathbf{V}) \in \mathcal{A}^I\}$$

    Notice that quasi-linearization is performed at $\mathbf{V}$

- **(Well-known) Proposition.** [Low-order Euler IDP-IMEX]
  Let Assumption 1 hold. Assume that $\mathbf{U}^n \in \mathcal{A}^I$ and $\tau \in (0, \tau^*]$. Then, $\mathbf{U}^{\mathrm{L},n+1} \in \mathcal{A}^I$

High-order Euler IDP-IMEX (1/2)

- We want to use high-order fluxes in space!

High-order Euler IDP-IMEX (1/2)

- We want to use high-order fluxes in space!
- Assumption 2. There exist two nonlinear limiting operators

$$\ell^{\text{hyp}}, \ell^{\text{par}} : \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$$

such that

  - for all $(\mathsf{V}, \boldsymbol{\Phi}^{\mathrm{L}}, \boldsymbol{\Phi}^{\mathrm{H}}) \in \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I$,

$$\{\mathsf{V} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\boldsymbol{\Phi}^{\mathrm{L}} \in \mathcal{A}^I\} \implies \{\ell^{\text{hyp}}(\mathsf{V}, \boldsymbol{\Phi}^{\mathrm{L}}, \boldsymbol{\Phi}^{\mathrm{H}}) \in \mathcal{A}^I\}$$

  - for all $(\mathsf{W}, \boldsymbol{\Psi}^{\mathrm{L}}, \boldsymbol{\Psi}^{\mathrm{H}}) \in \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I$,

$$\{\mathsf{W} + \tau(\mathbb{M}^{\mathrm{L}})^{-1}\boldsymbol{\Psi}^{\mathrm{L}} \in \mathcal{A}^I\} \implies \{\ell^{\text{par}}(\mathsf{W}, \boldsymbol{\Psi}^{\mathrm{L}}, \boldsymbol{\Psi}^{\mathrm{H}}) \in \mathcal{A}^I\}$$

High-order Euler IDP-IMEX (1/2)

- We want to use high-order fluxes in space!
- Assumption 2. There exist two nonlinear limiting operators

$$\ell^{\text{hyp}}, \ell^{\text{par}} : \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I \to (\mathbb{R}^m)^I$$

  such that
  - for all $(\mathsf{V}, \boldsymbol{\Phi}^{\text{L}}, \boldsymbol{\Phi}^{\text{H}}) \in \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I$,

$$\left\{\mathsf{V} + \tau(\mathbb{M}^{\text{L}})^{-1}\boldsymbol{\Phi}^{\text{L}} \in \mathcal{A}^I\right\} \implies \left\{\ell^{\text{hyp}}(\mathsf{V}, \boldsymbol{\Phi}^{\text{L}}, \boldsymbol{\Phi}^{\text{H}}) \in \mathcal{A}^I\right\}$$

  - for all $(\mathsf{W}, \boldsymbol{\Psi}^{\text{L}}, \boldsymbol{\Psi}^{\text{H}}) \in \mathcal{A}^I \times (\mathbb{R}^m)^I \times (\mathbb{R}^m)^I$,

$$\left\{\mathsf{W} + \tau(\mathbb{M}^{\text{L}})^{-1}\boldsymbol{\Psi}^{\text{L}} \in \mathcal{A}^I\right\} \implies \left\{\ell^{\text{par}}(\mathsf{W}, \boldsymbol{\Psi}^{\text{L}}, \boldsymbol{\Psi}^{\text{H}}) \in \mathcal{A}^I\right\}$$

- Important remarks
  - the invariant domains enforced by the two limiters can be different
  - bounds for limiting are deduced from the low-order updates

High-order Euler IDP-IMEX (2/2)

- Given $\mathbf{U}^n \in \mathcal{A}^I$, high-order Euler IDP-IMEX proceeds as follows:

$$\mathbf{U}^n \underbrace{\xrightarrow{(1)} (\mathbf{W}^{\mathrm{L},n+1}, \mathbf{W}^{\mathrm{H},n+1}) \xrightarrow{(2)}}_{\text{hyperbolic step}} \mathbf{W}^{n+1} \underbrace{\xrightarrow{(3)} (\mathbf{U}^{\mathrm{L},n+1}, \mathbf{U}^{\mathrm{H},n+1}) \xrightarrow{(4)}}_{\text{parabolic step}} \mathbf{U}^{n+1}$$

High-order Euler IDP-IMEX (2/2)

- Given $\mathbf{U}^n \in \mathcal{A}^I$, high-order Euler IDP-IMEX proceeds as follows:

$$\underbrace{\mathbf{U}^n \xrightarrow{(1)} (\mathbf{W}^{\mathrm{L},n+1}, \mathbf{W}^{\mathrm{H},n+1}) \xrightarrow{(2)}}_{\text{hyperbolic step}} \mathbf{W}^{n+1} \underbrace{\xrightarrow{(3)} (\mathbf{U}^{\mathrm{L},n+1}, \mathbf{U}^{\mathrm{H},n+1}) \xrightarrow{(4)}}_{\text{parabolic step}} \mathbf{U}^{n+1}$$

- Hyperbolic steps (1) and (2): compute low/high-order updates and limit

$$\mathbb{M}^{\mathrm{L}}\mathbf{W}^{\mathrm{L},n+1} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^n + \tau \mathbf{F}^{\mathrm{L}}(\mathbf{U}^n),$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{W}^{\mathrm{H},n+1} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^n + \tau \mathbf{F}^{\mathrm{H}}(\mathbf{U}^n),$$

$$\mathbf{W}^{n+1} := \ell^{\mathrm{hyp}}(\mathbf{U}^n, \mathbf{\Phi}^{\mathrm{L}}, \mathbf{\Phi}^{\mathrm{H}})$$

High-order Euler IDP-IMEX (2/2)

- Given $\mathbf{U}^n \in \mathcal{A}^I$, high-order Euler IDP-IMEX proceeds as follows:

$$\underbrace{\mathbf{U}^n \xrightarrow{(1)} (\mathbf{W}^{\mathrm{L},n+1}, \mathbf{W}^{\mathrm{H},n+1}) \xrightarrow{(2)}}_{\text{hyperbolic step}} \mathbf{W}^{n+1} \underbrace{\xrightarrow{(3)} (\mathbf{U}^{\mathrm{L},n+1}, \mathbf{U}^{\mathrm{H},n+1}) \xrightarrow{(4)}}_{\text{parabolic step}} \mathbf{U}^{n+1}$$

- Hyperbolic steps (1) and (2): compute low/high-order updates and limit

$$\mathbb{M}^{\mathrm{L}}\mathbf{W}^{\mathrm{L},n+1} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^n + \tau\mathbf{F}^{\mathrm{L}}(\mathbf{U}^n), \qquad \mathbf{W}^{n+1} := \ell^{\mathrm{hyp}}(\mathbf{U}^n, \mathbf{\Phi}^{\mathrm{L}}, \mathbf{\Phi}^{\mathrm{H}})$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{W}^{\mathrm{H},n+1} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^n + \tau\mathbf{F}^{\mathrm{H}}(\mathbf{U}^n),$$

- Parabolic steps (3) and (4): compute low/high-order updates (quasi-linear solves) and limit

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},n+1} - \tau\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{n+1};\mathbf{U}^{\mathrm{L},n+1}) := \mathbb{M}^{\mathrm{L}}\mathbf{W}^{n+1}, \qquad \mathbf{U}^{n+1} := \ell^{\mathrm{par}}(\mathbf{W}^{n+1}, \mathbf{\Psi}^{\mathrm{L}}, \mathbf{\Psi}^{\mathrm{H}})$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},n+1} - \tau\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{\mathrm{H},n+1}) := \mathbb{M}^{\mathrm{H}}\mathbf{W}^{n+1},$$

High-order Euler IDP-IMEX (2/2)

- Given $\mathbf{U}^n \in \mathcal{A}^I$, high-order Euler IDP-IMEX proceeds as follows:

$$\underbrace{\mathbf{U}^n \xrightarrow{(1)} (\mathbf{W}^{\mathrm{L},n+1}, \mathbf{W}^{\mathrm{H},n+1}) \xrightarrow{(2)}}_{\text{hyperbolic step}} \mathbf{W}^{n+1} \underbrace{\xrightarrow{(3)} (\mathbf{U}^{\mathrm{L},n+1}, \mathbf{U}^{\mathrm{H},n+1}) \xrightarrow{(4)}}_{\text{parabolic step}} \mathbf{U}^{n+1}$$

- Hyperbolic steps (1) and (2): compute low/high-order updates and limit

$$\mathbb{M}^{\mathrm{L}}\mathbf{W}^{\mathrm{L},n+1} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^n + \tau\mathbf{F}^{\mathrm{L}}(\mathbf{U}^n), \qquad \mathbb{M}^{\mathrm{H}}\mathbf{W}^{\mathrm{H},n+1} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^n + \tau\mathbf{F}^{\mathrm{H}}(\mathbf{U}^n),$$

$$\mathbf{W}^{n+1} := \ell^{\mathrm{hyp}}(\mathbf{U}^n, \boldsymbol{\Phi}^{\mathrm{L}}, \boldsymbol{\Phi}^{\mathrm{H}})$$

- Parabolic steps (3) and (4): compute low/high-order updates (quasi-linear solves) and limit

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},n+1} - \tau\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{n+1};\mathbf{U}^{\mathrm{L},n+1}) := \mathbb{M}^{\mathrm{L}}\mathbf{W}^{n+1}, \qquad \mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},n+1} - \tau\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{\mathrm{H},n+1}) := \mathbb{M}^{\mathrm{H}}\mathbf{W}^{n+1},$$

$$\mathbf{U}^{n+1} := \ell^{\mathrm{par}}(\mathbf{W}^{n+1}, \boldsymbol{\Psi}^{\mathrm{L}}, \boldsymbol{\Psi}^{\mathrm{H}})$$

- **(Well-known) Proposition.** [High-order Euler IDP-IMEX]
Let Assumptions 1 and 2 hold. Assume that $\mathbf{U}^n \in \mathcal{A}^I$ and $\tau \in (0, \tau^*]$.
Then, $\mathbf{U}^{n+1} \in \mathcal{A}^I$

High-order IDP-IMEX

- We are now ready to go high-order in time!
- Key idea. Consider the following two ODE systems on (t^n, t^{n+1}):

$$\mathbb{M}^{\mathrm{L}}\partial_t\mathbf{U} = \underbrace{\mathbf{F}^{\mathrm{L}}(\mathbf{U})}_{\text{explicit}} + \underbrace{\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{n,l};\mathbf{U})}_{\text{implicit}} \qquad \text{(at each stage } l\text{)}$$

$$\mathbb{M}^{\mathrm{H}}\partial_t\mathbf{U} = \underbrace{\mathbf{F}^{\mathrm{H}}(\mathbf{U}) + \mathbf{G}^{\mathrm{H}}(\mathbf{U}) - \mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^n;\mathbf{U})}_{\text{explicit}} + \underbrace{\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^n;\mathbf{U})}_{\text{implicit}}$$

Butcher tableaux

- Explicit Butcher tableau

$$
\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathsf{e}}_{2,1} & 0 & & & \\
c_3 & a^{\mathsf{e}}_{3,1} & a^{\mathsf{e}}_{3,2} & 0 & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathsf{e}}_{s,1} & a^{\mathsf{e}}_{s,2} & \cdots & a^{\mathsf{e}}_{s,s-1} & 0 \\
\hline
1 & a^{\mathsf{e}}_{s+1,1} & a^{\mathsf{e}}_{s+1,2} & \cdots & a^{\mathsf{e}}_{s+1,s-1} & a^{\mathsf{e}}_{s+1,s}
\end{array}
$$

Butcher tableaux

- Explicit Butcher tableau

$$\begin{array}{c|cccccc} 0 & 0 & & & & \\ c_2 & a^{\mathsf{e}}_{2,1} & 0 & & & \\ c_3 & a^{\mathsf{e}}_{3,1} & a^{\mathsf{e}}_{3,2} & 0 & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ c_s & a^{\mathsf{e}}_{s,1} & a^{\mathsf{e}}_{s,2} & \cdots & a^{\mathsf{e}}_{s,s-1} & 0 \\ \hline 1 & a^{\mathsf{e}}_{s+1,1} & a^{\mathsf{e}}_{s+1,2} & \cdots & a^{\mathsf{e}}_{s+1,s-1} & a^{\mathsf{e}}_{s+1,s} \end{array}$$

- Implicit Butcher tableau

$$\begin{array}{c|cccccc} 0 & 0 & & & & \\ c_2 & a^{\mathsf{i}}_{2,1} & a^{\mathsf{i}}_{2,2} & & & \\ c_3 & a^{\mathsf{i}}_{3,1} & a^{\mathsf{i}}_{3,2} & a^{\mathsf{i}}_{3,3} & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ c_s & a^{\mathsf{i}}_{s,1} & a^{\mathsf{i}}_{s,2} & \cdots & a^{\mathsf{i}}_{s,s-1} & a^{\mathsf{i}}_{s,s} \\ \hline 1 & a^{\mathsf{i}}_{s+1,1} & a^{\mathsf{i}}_{s+1,2} & \cdots & a^{\mathsf{i}}_{s+1,s-1} & a^{\mathsf{i}}_{s+1,s} \end{array}$$

Butcher tableaux

- Explicit Butcher tableau

$$\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathsf{e}}_{2,1} & 0 & & & \\
c_3 & a^{\mathsf{e}}_{3,1} & a^{\mathsf{e}}_{3,2} & 0 & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathsf{e}}_{s,1} & a^{\mathsf{e}}_{s,2} & \cdots & a^{\mathsf{e}}_{s,s-1} & 0 \\
\hline
1 & a^{\mathsf{e}}_{s+1,1} & a^{\mathsf{e}}_{s+1,2} & \cdots & a^{\mathsf{e}}_{s+1,s-1} & a^{\mathsf{e}}_{s+1,s}
\end{array}$$

- Implicit Butcher tableau

$$\begin{array}{c|cccccc}
0 & 0 & & & & \\
c_2 & a^{\mathsf{i}}_{2,1} & a^{\mathsf{i}}_{2,2} & & & \\
c_3 & a^{\mathsf{i}}_{3,1} & a^{\mathsf{i}}_{3,2} & a^{\mathsf{i}}_{3,3} & & \\
\vdots & \vdots & \ddots & \ddots & \ddots & \\
c_s & a^{\mathsf{i}}_{s,1} & a^{\mathsf{i}}_{s,2} & \cdots & a^{\mathsf{i}}_{s,s-1} & a^{\mathsf{i}}_{s,s} \\
\hline
1 & a^{\mathsf{i}}_{s+1,1} & a^{\mathsf{i}}_{s+1,2} & \cdots & a^{\mathsf{i}}_{s+1,s-1} & a^{\mathsf{i}}_{s+1,s}
\end{array}$$

- Both tableaux share the same coefficients $(c_l)_{l\in\{1:s+1\}}$; recall the notation $l'(l) := \max\{k < l \mid c_k \le c_l\}$ (think of $l'(l) = l - 1$)

Details (1/2)

- Given $\mathbf{U}^n \in \mathcal{A}^I$, set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- At each stage $l \in \{2{:}s+1\}$, one performs the following steps:

$$\color{red}{\mathbf{U}^{n,l'}} \underbrace{\xrightarrow{(1)} (\mathbf{W}^{\mathrm{L},l}, \mathbf{W}^{\mathrm{H},l}) \xrightarrow{(2)}}_{\color{blue}{\text{hyperbolic step}}} \mathbf{W}^{n,l} \underbrace{\xrightarrow{(3)} (\mathbf{U}^{\mathrm{L},l}, \mathbf{U}^{\mathrm{H},l}) \xrightarrow{(4)}}_{\color{teal}{\text{parabolic step}}} \mathbf{U}^{n,l}$$

Details (1/2)

- Given $\mathbf{U}^n \in \mathcal{A}^I$, set $\mathbf{U}^{n,1} := \mathbf{U}^n$
- At each stage $l \in \{2{:}s+1\}$, one performs the following steps:

$$\mathbf{U}^{n,l'} \underbrace{\xrightarrow{(1)} (\mathbf{W}^{\mathrm{L},l}, \mathbf{W}^{\mathrm{H},l}) \xrightarrow{(2)}}_{\text{hyperbolic step}} \mathbf{W}^{n,l} \underbrace{\xrightarrow{(3)} (\mathbf{U}^{\mathrm{L},l}, \mathbf{U}^{\mathrm{H},l}) \xrightarrow{(4)}}_{\text{parabolic step}} \mathbf{U}^{n,l}$$

- Hyperbolic steps (1) and (2): compute low/high-order updates

$$\mathbb{M}^{\mathrm{L}}\mathbf{W}^{\mathrm{L},l} := \mathbb{M}^{\mathrm{L}}\mathbf{U}^{n,l'} + \tau(c_l - c_{l'})\mathbf{F}^{\mathrm{L}}(\mathbf{U}^{n,l'})$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{W}^{\mathrm{H},l} := \mathbb{M}^{\mathrm{H}}\mathbf{U}^{n,l'} + \tau \sum_{k\in\{1:l-1\}} (a^{\mathrm{e}}_{l,k} - a^{\mathrm{e}}_{l',k})\mathbf{F}^{\mathrm{H}}(\mathbf{U}^{n,k})$$

and limit

$$\mathbf{W}^{n,l} := \ell^{\mathrm{hyp}}(\mathbf{U}^{n,l'}, \mathbf{\Phi}^{\mathrm{L}}, \mathbf{\Phi}^{\mathrm{H}})$$

Details (2/2)

- Recall $\mathbf{W}^{n,l}$ just computed from hyperbolic steps (1) and (2)

Details (2/2)

- Recall $\mathbf{W}^{n,l}$ just computed from hyperbolic steps (1) and (2)
- Parabolic steps (3) and (4): compute low/high-order updates

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},l} - \tau(c_l - c_{l'})\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{n,l};\mathbf{U}^{\mathrm{L},l}) := \mathbb{M}^{\mathrm{L}}\mathbf{W}^{n,l}$$
$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},l} - \tau a^{\mathsf{i}}_{l,l}\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{\mathrm{H},l}) := \mathbb{M}^{\mathrm{H}}\mathbf{W}^{n,l} + \tau\Delta_l$$

$$\Big(\Delta_l := \sum_{k\in\{1:l-1\}}(a^{\mathsf{i}}_{l,k} - a^{\mathsf{i}}_{l',k})\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{n,k}) + \sum_{k\in\{1:l-1\}}(a^{\mathsf{e}}_{l,k} - a^{\mathsf{e}}_{l',k})\big(\mathbf{G}^{\mathrm{H}}(\mathbf{U}^{n,k}) - \mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{n,k})\big)\Big)$$

- Notice that $a^{\mathsf{i}}_{l,l} = 0$ for $l = s + 1$ (final high-order stage is explicit)
- Limit: $\mathbf{U}^{n+1} := \ell^{\mathrm{par}}(\mathbf{W}^{n,l}, \boldsymbol{\Psi}^{\mathrm{L}}, \boldsymbol{\Psi}^{\mathrm{H}})$

Details (2/2)

- Recall $\mathbf{W}^{n,l}$ just computed from hyperbolic steps (1) and (2)
- Parabolic steps (3) and (4): compute low/high-order updates

$$\mathbb{M}^{\mathrm{L}}\mathbf{U}^{\mathrm{L},l} - \tau(c_l - c_{l'})\mathbf{G}^{\mathrm{L,lin}}(\mathbf{W}^{n,l};\mathbf{U}^{\mathrm{L},l}) := \mathbb{M}^{\mathrm{L}}\mathbf{W}^{n,l}$$

$$\mathbb{M}^{\mathrm{H}}\mathbf{U}^{\mathrm{H},l} - \tau a^{\mathsf{i}}_{l,l}\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{\mathrm{H},l}) := \mathbb{M}^{\mathrm{H}}\mathbf{W}^{n,l} + \tau\Delta_l$$

$$\Big(\Delta_l := \sum_{k\in\{1:l-1\}}(a^{\mathsf{i}}_{l,k} - a^{\mathsf{i}}_{l',k})\mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{n,k}) + \sum_{k\in\{1:l-1\}}(a^{\mathsf{e}}_{l,k} - a^{\mathsf{e}}_{l',k})\big(\mathbf{G}^{\mathrm{H}}(\mathbf{U}^{n,k}) - \mathbf{G}^{\mathrm{H,lin}}(\mathbf{U}^{n};\mathbf{U}^{n,k})\big)\Big)$$

- Notice that $a^{\mathsf{i}}_{l,l} = 0$ for $l = s+1$ (final high-order stage is explicit)
- Limit: $\mathbf{U}^{n+1} := \ell^{\mathrm{par}}(\mathbf{W}^{n,l}, \mathbf{\Psi}^{\mathrm{L}}, \mathbf{\Psi}^{\mathrm{H}})$
- **Theorem.** [High-order IDP-IMEX]
  Let Assumptions 1 and 2 hold. Assume that $\mathbf{U}^{n} \in \mathcal{A}^{I}$. Then, $\mathbf{U}^{n+1} \in \mathcal{A}^{I}$ (as well as all intermediate stages) $\forall\tau \in (0, \tau^{*}/\max_{l\in\{2:s+1\}}(c_l - c_{l'})]$

Important omitted details

- The design of low-order linearized parabolic flux $\mathbf{G}^{\mathrm{L,lin}}$ is problem-dependent

Important omitted details

- The design of low-order linearized parabolic flux $\mathbf{G}^{\mathrm{L,lin}}$ is problem-dependent
- The whole scheme can be rewritten using conservative limiters

Important omitted details

- The design of low-order linearized parabolic flux $\mathbf{G}^{\mathrm{L,lin}}$ is problem-dependent
- The whole scheme can be rewritten using conservative limiters
- The setting allows for the hyperbolic and parabolic problems to be solved each with its own natural set of variables
  - conservative for Euler, primitive for Navier–Stokes