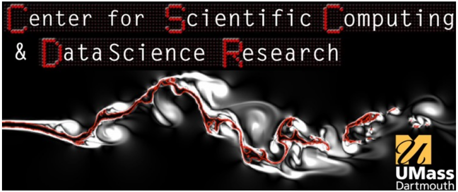
Strong Stability Preserving Two Derivative Methods

Zachary J. Grant
Mathematics Department
University of Massachusetts Dartmouth



Thanks to my colleagues: Andrew Christlieb, Sining Gong, Sigal Gottlieb, Jingwei Hu, Elisha Rogatch, David Seal, Ruiwen Shu & to the AFOSR and DOE for funding much of this work

Motivation

In the numerical solution of hyperbolic partial differential equations

$$u_t + f(u)_x = 0$$

with discontinuous solutions, we need the spatial discretization to satisfy certain nonlinear stability properties usually when coupled with forward Euler.

Strong stability preserving (SSP) high order time discretizations were developed to ensure that these stability properties of the spatial discretization are preserved when coupled with these higher order methods.

The forward Euler condition

Given a system of ODEs, generally resulting from a spatial discretization of a PDE, of the form

$$u_t = \mathbf{F}(u)$$

Forward Euler condition:

$$\|u^n + \Delta t \mathbf{F}(u^n)\| \leq \|u^n\| \quad \forall \Delta t \leq \Delta t_{\text{FE}}$$

where $\|\cdot\|$ is some convex functional.

SSP methods extend this strong stability to higher order methods. How?

High order SSP Runge–Kutta methods

Consider the two-stage second order Runge–Kutta method

$$\begin{aligned} u^{(1)} &= u^n + \Delta t \mathbf{F}(u^n) \\ u^{n+1} &= \frac{1}{2} u^n + \frac{1}{2}\left(u^{(1)} + \Delta t F(u^{(1)})\right), \end{aligned}$$

we can write this method as a convex combination of the forward Euler method

$$\begin{aligned} \| \, u^{(1)} \, \| &= \|u^n + \Delta t \mathbf{F}(u^n)\| \leq \|\mathbf{u}^n\| \;\; \text{for} \;\; \Delta t \leq \Delta t_{\text{FE}} \\ \|u^{n+1}\| &= \|\frac{1}{2} u^n + \frac{1}{2}\left(u^{(1)} + \Delta t \mathbf{F}(u^{(1)})\right) \| \\ &\leq \frac{1}{2}\|u^n\| + \frac{1}{2}\left\| u^{(1)} + \Delta t \mathbf{F}(u^{(1)})\right\| \leq \|u^n\| \;\; \text{for} \;\; \Delta t \leq \Delta t_{\text{FE}} \end{aligned}$$

This method is SSP with $\mathcal{C} = 1$ and $\mathcal{C}_{\text{eff}} = \frac{1}{2}$.

UMass Dartmouth

High order SSP Runge–Kutta methods

Consider the two-stage second order Runge–Kutta method

$$\begin{aligned} u^{(1)} &= u^n + \Delta t \mathbf{F}(u^n) \\ u^{n+1} &= \frac{1}{2} u^n + \frac{1}{2}\left(u^{(1)} + \Delta t F(u^{(1)}) \right), \end{aligned}$$

we can write this method as a convex combination of the forward Euler method

$$\begin{aligned} \| \, u^{(1)} \, \| &= \|u^n + \Delta t \mathbf{F}(u^n)\| \leq \|\mathbf{u}^n\| \;\; \text{for} \;\; \Delta t \leq \Delta t_{\mathsf{FE}} \\ \|u^{n+1}\| &= \| \frac{1}{2} u^n + \frac{1}{2}\left(u^{(1)} + \Delta t \mathbf{F}(u^{(1)}) \right) \| \\ &\leq \frac{1}{2}\|u^n\| + \frac{1}{2}\left\| u^{(1)} + \Delta t \mathbf{F}(u^{(1)}) \right\| \leq \|u^n\| \;\; \text{for} \;\; \Delta t \leq \Delta t_{\mathsf{FE}} \end{aligned}$$

This method is SSP with $\mathcal{C} = 1$ and $\mathcal{C}_{\text{eff}} = \frac{1}{2}$.

UMass Dartmouth

Can I always do this? No!

IF

we can decompose a higher order method into convex combinations of forward Euler

THEN

the Runge–Kutta or multistep method will preserve the strong stability property that $\mathbf{F}$ satisfies when coupled with FE.

SSP methods were introduced by Chi-Wang Shu and Stanley Osher in 1988 for time-evolution of hyperbolic PDEs with shocks. These were initially called TVD time stepping methods.

SSP Runge–Kutta Methods

- Explicit SSP Runge–Kutta methods have order $p \leq 4$ and $\mathcal{C}_{\text{eff}} \leq 1$.
- Explicit SSP Runge–Kutta methods for linear problems have no order barrier but $\mathcal{C}_{\text{eff}} \leq 1$.
- Implicit SSP Runge–Kutta methods have order $p \leq 6$
- Implicit Runge–Kutta cannot be unconditionally SSP.
- All optimal implicit SSP Runge–Kutta methods found have $\mathcal{C}_{\text{eff}} \leq 2$.

Going back to basics: Downwinding

We try to find ways to break the order barriers and in particular the SSP coefficient bounds.

Idea: The key to the SSP property was to decompose the higher order methods into a convex combination of forward Euler steps, i.e.

$$\alpha \mathbf{u}^{(j)} + \Delta t \beta \mathbf{F}(\mathbf{u}^{(j)})$$

where α and β are non-negative.

But what if some of the β coefficients are negative?

Downwinding

If we have a spatial discretization $\tilde{\mathbf{F}}$ that approximates the same derivative as the spatial approximation $\mathbf{F}$, but has the property that

$$||\mathbf{u}^n - \Delta t \tilde{\mathbf{F}}(\mathbf{u}^n)|| \leq ||\mathbf{u}^n|| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{FE}},$$

while $\mathbf{F}$ has the property

$$||\mathbf{u}^n + \Delta t \mathbf{F}(\mathbf{u}^n)|| \leq ||\mathbf{u}^n|| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{FE}}.$$

If a method can be decomposed into convex combination of $\mathbf{u}^n + \Delta t \mathbf{F}(\mathbf{u}^n)$ and $\mathbf{u}^n - \Delta t \tilde{\mathbf{F}}(\mathbf{u}^n)$ then the Runge–Kutta method will preserve the strong stability property that $\mathbf{F}$ and $\tilde{\mathbf{F}}$ satisfy when coupled with forward Euler or backward Euler (respectively).

UMass Dartmouth

Downwinding: Breaking the order barrier

Explicit Runge–Kutta methods with downwinding can **break the order barrier**

If you have a PDE that admits an upwind stable discretization $\mathbf{F}$ and a downwind stable discretization $\tilde{\mathbf{F}}$ such that we can obtain

- A 4-stage fourth order method with $\mathcal{C} = 0.936$ (Gottlieb & Shu 1998)
- A fifth order explicit Runge–Kutta method with $\mathcal{C} = 1.17$ (Ruuth & Spiteri 2004)
- This may be more efficient than it first appear as we can compute the downwind and upwind operators together at less than double the cost (Gottlieb & Ruuth 2006).

UMass | Dartmouth

Unconditional SSP methods with downwinding

Ketcheson (2011): Implicit Runge–Kutta methods with downwinding can break the SSP coefficient bound

$$
\begin{aligned}
y_1 &= \frac{2}{r(r-2)} u^n + \frac{2}{r}\left(y_1 + \frac{\Delta t}{r} F(y_1)\right) + \frac{r^2 - 4r + 2}{r(r-2)}\left(y_2 - \frac{\Delta t}{r}\tilde{F}(y_2)\right) \\
y_2 &= y_1 + \frac{\Delta t}{r} F(y_1) \\
u^{n+1} &= y_2 + \frac{\Delta t}{r} F(y_2)
\end{aligned}
$$

Similarly, a family of third order fully implicit downwind methods with arbitrarily large $\mathcal{C}$ was found by Ketcheson's student Yiannis Hadjimichael and appears in his thesis (2017).

UMass | Dartmouth

Another approach: using additional derivatives

To break the order barriers and in particular the SSP coefficient bounds we can also use an additional derivative.
This approach is based on a Taylor series idea, where $u_{tt} = \dot{\mathbf{F}}$ is used. In our case we add this to the Runge–Kutta method.

- Enhancing Runge–Kutta methods with additional derivatives as far back as 1950 (Turan) and 1963 (Stroud & Stroud).
- Multistage multiderivative time integrators for ordinary differential equations were studied by Shintani (1971,1972) and Kastlunger and G. Wanner (1972) and Mitsui (1982), Ono (2004) and Tsai (2010).
- More recently, for the the time evolution of partial differential equations, these were studied by Seal & Christlieb, Tsai, Li & Du. This makes sense for hyperbolic conservation laws, where it is natural (i.e. cheap) to include the second derivative.

UMass Dartmouth

Two derivatives Runge–Kutta methods

In 2014 Seal and Christlieb studied high-order multiderivative time integrators for hyperbolic conservation laws. In particular they used the 4th order method:

$$
\begin{aligned}
y^{(1)} &= u^n + \frac{\Delta t}{2}\mathbf{F}(u^n) + \frac{\Delta t^2}{8}\dot{\mathbf{F}}(u^n) \\
u^{n+1} &= u^n + \Delta t \mathbf{F}(u^n) + \frac{\Delta t^2}{6}(\dot{\mathbf{F}}(u^n) + 2\dot{\mathbf{F}}(y^{(1)})).
\end{aligned}
$$

Is this method SSP?

UMass Dartmouth

Two derivative Runge–Kutta methods

We proved that the fourth order method above gives

$$\|u^{n+1}\| \leq \|u^n\|$$

if $\mathbf{F}$ satisfies the forward Euler condition

Forward Euler

$$||\mathbf{u}^n + \Delta t \mathbf{F}(\mathbf{u}^n)|| \leq ||\mathbf{u}^n|| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{FE}}.$$

and $\dot{\mathbf{F}}$ satisfies the condition

Second derivative

$$\|u^n + \Delta t^2 \dot{\mathbf{F}}(u^n)\| \leq \|u^n\| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{SD}} = K \Delta t_{\text{FE}}$$

where $K > 0$.

These building blocks make sense

Given the simple linear one-way wave equation $U_t = U_x$, let

$$u_t \approx u_x \approx \mathbf{F}(u^n)_j := \frac{1}{\Delta x}\left(u^n_{j+1} - u^n_j\right).$$

and

$$u_{tt} = u_{xx} \approx \dot{\mathbf{F}}(u^n)_j := \frac{1}{\Delta x^2}\left(u^n_{j+1} - 2u^n_j + u^n_{j-1}\right).$$

Then F and $\dot{F}$ satisfy the total variation diminishing (TVD) property:

$$\|u^n + \Delta t \mathbf{F}(u^n)\|_{TV} \leq \|u^n\|_{TV} \quad \text{for} \quad \Delta t \leq \Delta x,$$

$$\left\|u^n + \Delta t^2 \dot{\mathbf{F}}(u^n)\right\|_{TV} \leq \|u^n\|_{TV} \quad \text{for} \quad \Delta t \leq \frac{\sqrt{2}}{2}\Delta x.$$

Optimal one stage second order methods

The Taylor series method is the unique explicit one-stage two-derivative second order method:

$$u^{n+1} \;\; = \;\; u^n + \Delta t F(u^n) + \frac{1}{2}\Delta t^2 \dot{F}(u^n).$$

It is one method, but the optimal Shu-Osher decomposition and the corresponding SSP coefficient, depend on the value of K in the second derivative condition.

Optimal one stage second order methods:

This method can be written as a convex combination of the building blocks:

$$u^{n+1} = (1-\alpha)\left(u^n + \frac{1}{1-\alpha}\Delta t F(u^n)\right) + \alpha\left(u^n + \frac{1}{2\alpha}\Delta t^2 \dot{F}(u^n)\right).$$

This is SSP for $\Delta t \leq \max\{(1-\alpha)\Delta t_{\text{FE}}, \sqrt{2\alpha} K \Delta t_{\text{FE}}\}$, so we set

$$(1-\alpha)^2 = 2\alpha K^2 \quad \Longrightarrow \quad \alpha = 1 + K^2 \pm K\sqrt{2+K^2}$$

to obtain

$$\mathcal{C} = K\sqrt{2+K^2} - K^2.$$

UMass Dartmouth

Two-stage third order methods

Optimal SSP explicit two-stage two-derivative third order methods take the form

$$\begin{aligned} u^* &= u^n + a\Delta t F(u^n) + \hat{a}\Delta t^2 \dot{F}(u^n), \\ u^{n+1} &= u^n + b_1 \Delta t F(u^n) + b_2 \Delta t F(u^*) \\ &\quad + \hat{b_1}\Delta t^2 \dot{F}(u^n) + \hat{b_2}\Delta t^2 \dot{F}(u^*). \end{aligned}$$

The optimal method has SSP coefficient $\mathcal{C} = r$ that depends on K. This is a little messy!

Two-stage third order methods

The optimal method has SSP coefficient $\mathcal{C} = r$ given by the smallest positive root of the polynomial

$$2K(a_0 - 2K) + 4K^3 a_0 - a_0 r + \frac{1 - a_0}{2K^2} r^2 - \frac{\frac{a_0}{2K} + K}{6K^3} r^3,$$

for $a_0 = \sqrt{K^2 + 2} - K$. The coefficients of the optimal method for each K depend on K and r and are given by

$$a = \frac{1}{r}\left(K a_0\right), \qquad b_1 = 1 - b_2, \qquad b_2 = \frac{2K^2(1 - \frac{1}{r}) + r}{K a_0 + 2K^2} - \frac{r^2}{3K^2},$$

$$\hat{a} = \frac{1}{2}a^2, \quad \hat{b}_1 = \frac{1}{2} - \frac{1}{2}ab_2 - \frac{1}{6a}, \qquad \hat{b}_2 = \frac{1}{6a} - \frac{1}{2}ab_2.$$

Fourth order methods

The two-stage two-derivative fourth order method is

$$y^{(1)} = u^n + \frac{\Delta t}{2}\mathbf{F}(u^n) + \frac{\Delta t^2}{8}\dot{\mathbf{F}}(u^n)$$
$$u^{n+1} = u^n + \Delta t\mathbf{F}(u^n) + \frac{\Delta t^2}{6}(\dot{\mathbf{F}}(u^n) + 2\dot{\mathbf{F}}(y^{(1)})).$$

Although the method is unique, the optimal decomposition, and therefore the SSP coefficient, depends on K. The SSP coefficient $\mathcal{C} = r$ is given by the smallest positive root of

$$r^4 + 4K^2r^3 - 12K^2r^2 - 24K^4r + 24K^4.$$

UMass Dartmouth

Three-stage fifth order methods

Increasing the number of stages to three breaks the order barrier and allows fifth order method:

$$
\begin{aligned}
y^{(1)} &= u^n + a_{21} \Delta t F(u^n) + \dot{a}_{21} \Delta t^2 \dot{F}(u^n) \\
y^{(2)} &= u^n + a_{31} \Delta t F(u^n) + \dot{a}_{31} \Delta t^2 \dot{F}(u^n) + \dot{a}_{32} \Delta t^2 \dot{F}(y^{(1)}) \\
u^{n+1} &= u^n + \Delta t F(u^n) + \Delta t^2 \left(\dot{b}_1 \dot{F}(u^n) + \dot{b}_2 \dot{F}(y^{(1)}) + \dot{b}_3 \dot{F}(y^{(2)}) \right).
\end{aligned}
$$

The SSP coefficient $\mathcal{C} = r$ depends on K and is messy.

UMass Dartmouth

Three-stage fifth order methods

The SSP coefficient $\mathcal{C} = r$ is the largest positive root of

$$10r^2 a_{21}^4 - (100K^2 + 10r^2)a_{21}^3 + (130K^2 + 3r^2)a_{21}^2 - 50K^2 a_{21} + 6K^2,$$

where

$$a_{21} = \frac{K^6}{r^6}\left(-\frac{2}{K^4}r^5 + \frac{10}{K^4}r^4 + \frac{40}{K^2}r^3 - \frac{120}{K^2}r^2 - 240r + 240\right).$$

Given K, we can find the corresponding r, and the coefficients are then given as a one-parameter system

$$\begin{aligned}
\dot{a}_{21} &= \frac{1}{2}a_{21}^2, \qquad a_{31} = \frac{3/5 - a_{21}}{1 - 2a_{21}}, \\
\dot{a}_{32} &= \frac{1}{10}\left(\frac{(\frac{3}{5} - a_{21})^2}{a_{21}(1 - 2a_{21})^3} - \frac{\frac{3}{5} - a_{21}}{(1 - 2a_{21})^2}\right), \qquad \dot{a}_{31} = \frac{1}{2}\frac{(\frac{3}{5} - a_{21})^2}{(1 - 2a_{21})^2} - \dot{a}_{32}, \\
\dot{b}_2 &= \frac{2a_{31} - 1}{12a_{21}(a_{31} - a_{21})}, \qquad \dot{b}_3 = \frac{1 - 2a_{21}}{12a_{31}(a_{31} - a_{21})}, \qquad \dot{b}_1 = \frac{1}{2} - \dot{b}_2 - \dot{b}_3.
\end{aligned}$$

UMass Dartmouth

Two derivative Runge–Kutta methods

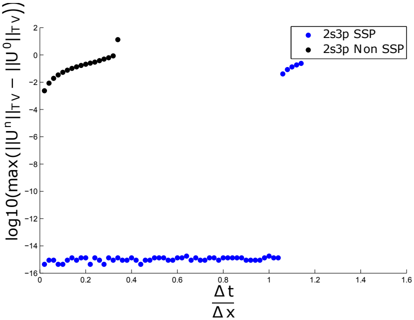
- Decomposing methods into convex combinations of the building blocks $u^n + \Delta t \mathbf{F}(u^n)$ and $u^n + \Delta t^2 \dot{\mathbf{F}}(u^n)$ allows us to build SSP two derivative Runge–Kutta methods.
- The SSP coefficient depends not only on the number of stages but also on K
- Using these building blocks we found explicit optimal SSP methods up to sixth order
- Two derivative methods allow us to break the order barrier of explicit Runge–Kutta methods
- Does the SSP property really matter in practice?

Yes!

Using an SSP and non-SSP two-stage two-derivative third order method on the motivating example above with a step function initial condition

$$u_0(x) = \begin{cases} 1 & \text{if } \frac{1}{4} \leq x \leq \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

and periodic boundary conditions $u(0,t) = u(1,t)$.



Maximal TV rise above the initial TV.

Different building blocks

Forward Euler

$$||\mathbf{u}^n + \Delta t \mathbf{F}(\mathbf{u}^n)|| \leq ||\mathbf{u}^n|| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{FE}}.$$

&

Taylor Series condition

$$\|u^n + \Delta t \mathbf{F}(\mathbf{u}^n) + \frac{1}{2}\Delta t^2 \dot{\mathbf{F}}(u^n)\| \leq \|u^n\| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{TS}}$$

$\uparrow$ allows for **more** spatial discretizations
$\downarrow$ fewer time-discretization methods are SSP (barrier of $p \leq 6$).

UMass Dartmouth

More flexibility in spatial discretization

For $U_t = U_x$ we define (as above)

$$\mathbf{F}(u^n)_j := \frac{1}{\Delta x}\left(u^n_{j+1} - u^n_j\right) \approx U_x(x_j).$$

To get $\dot{\mathbf{F}}$ apply this differentiation operator twice

$$\dot{\mathbf{F}}(u^n)_j := \frac{1}{\Delta x^2}\left(u^n_{j+2} - 2u^n_{j+1} + u^n_j\right).$$

$\mathbf{F}$ satisfies the total variation diminishing (TVD) property for $\Delta t \leq \Delta x$, while the Taylor series term using $\mathbf{F}$ and $\dot{\mathbf{F}}$ satisfies

$$\|u^n + \Delta t \mathbf{F}(u^n) + \frac{1}{2}\Delta t^2 \dot{\mathbf{F}}(u^n)\|_{TV} \leq \|u^n\|_{TV} \quad \text{for} \quad \Delta t \leq \Delta x.$$

Less flexibility in time discretization

For example, the fourth order method

$$y^{(1)} = u^n + \frac{\Delta t}{2} F(u^n) + \frac{\Delta t^2}{8} \dot{F}(u^n)$$
$$u^{n+1} = u^n + \Delta t F(u^n) + \frac{\Delta t^2}{6} (\dot{F}(u^n) + 2\dot{F}(y^{(1)})).$$

cannot be written as convex combinations of these building blocks.
But this fourth order method **can**, with $\mathcal{C} = 1$

$$\begin{aligned}
y^{(1)} &= u^n \\
y^{(2)} &= u^n + \Delta t F(y^{(1)}) + \frac{1}{2} \Delta t^2 \dot{F}(y^{(1)}) \\
y^{(3)} &= u^n + \frac{1}{27} \Delta t \left(14 F(y^{(1)}) + 4 F(y^{(2)}) \right) + \frac{2}{27} \Delta t^2 \dot{F}(y^{(1)}) \\
u^{n+1} &= u^n + \frac{1}{48} \Delta t \left(17 F(y^{(1)}) + 4 F(y^{(2)}) + 27 F(y^{(3)}) \right) + \frac{1}{24} \Delta t^2 \dot{F}(y^{(1)}).
\end{aligned}$$

UMass Dartmouth

Implicit methods with two derivatives

Whichever of these we use

Second derivative

$$u^n + \Delta t^2 \dot{\mathbf{F}}(u^n)$$

Taylor Series

$$u^n + \Delta t F(u^n) + \frac{1}{2}\Delta t^2 \dot{F}(u^n)$$

when we move to implicit methods our time-step is very limited. Why? Look at implicit Taylor series method

$$u^{n+1} = u^n + \Delta t F(u^{n+1}) - \frac{1}{2}\Delta t^2 \dot{F}(u^{n+1}).$$

Negative derivative condition

This suggests* the negative derivative condition

Negative derivative condition

$$\|u^n - \Delta t^2 \dot{F}(u^n)\| \leq \|u^n\| \quad \text{for} \quad \Delta t \leq \Delta t_{\text{ND}}.$$

Or even

Implicit negative derivative condition

$$u^{n+1} = u^n - \Delta t^2 \dot{F}(u^{n+1}) \quad \Longrightarrow \quad \|u^{n+1}\| \leq \|u^n\| \quad \forall \Delta t > 0.$$

↑ **Better SSP properties (unconditional)**

↓ **Fewer discretizations satisfy these conditions**

UMass Dartmouth

Unconditionally SSP implicit multiderivative schemes

For $u_t = \mathbf{G}(u)$, with $u_{tt} = \dot{\mathbf{G}}(u)$ we write the s-stage method in a special Shu-Osher form with only implicit computations

$$u^{(i)} = r_i u^n + \sum_{j=1}^{i-1} p_{ij} u^{(j)} + \Delta t d_{ii} G(u^{(i)}) + \Delta t^2 \dot{d}_{ii} \dot{G}(u^{(i)}),$$
$$u^{n+1} = u^{(s)}.$$

In matrix form, this becomes

$$U = \mathbf{R}\mathbf{e}u^n + \mathbf{P}U + \Delta t \mathbf{D} G(U) + \Delta t^2 \dot{\mathbf{D}} \dot{G}(U),$$

where $\mathbf{P}$, and $\mathbf{R} = I - \mathbf{P}$ are $s \times s$ matrices, r_i are the ith row sum of $\mathbf{R}$, $\mathbf{D}$ and $\dot{\mathbf{D}}$ are $s \times s$ diagonal matrices, and $\mathbf{e}$ is a vector of ones.

UMass | Dartmouth

Theorem: Unconditionally SSP implicit schemes

Let the operators G and $\dot{G}$ satisfy

Implicit Euler

$$u^{n+1} = u^n + \Delta t \mathbf{G}(u^{n+1}) \quad \Longrightarrow \quad \|u^{n+1}\| \leq \|u^n\| \quad \forall \Delta t > 0.$$

and

Implicit negative derivative condition

$$u^{n+1} = u^n - \Delta t^2 \dot{\mathbf{G}}(u^{n+1}) \quad \Longrightarrow \quad \|u^{n+1}\| \leq \|u^n\| \quad \forall \Delta t > 0.$$

for some convex functional $\| \cdot \|$.

UMass Dartmouth

A method given by

$$U = \mathbf{Re} u^n + \mathbf{P} U + \Delta t \mathbf{D} G(U) + \Delta t^2 \dot{\mathbf{D}} \dot{G}(U),$$

that satisfies the conditions

$$\mathbf{Re} \geq 0, \quad \mathbf{P} \geq 0, \quad \mathbf{D} \geq 0, \quad \dot{\mathbf{D}} \leq 0, \quad \text{(componentwise)}$$

will preserve the strong stability property $\|u^{n+1}\| \leq \|u^n\|$ for any positive time-step $\Delta t > 0$.

These building blocks give diagonally implicit Runge–Kutta methods of order $p = 2, 3, 4$ that are unconditionally SSP.
S. Gottlieb, Z. Grant, J. Hu, R. Shu (2022).

UMass | Dartmouth

Unconditionally SSP implicit multiderivative schemes

Second order The one-stage, second order method is simply the implicit Taylor series method

$$u^{n+1} = u^n + \Delta t G(u^{n+1}) - \frac{1}{2}\Delta t^2 \dot{G}(u^{n+1}).$$

Third order

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \dot{\mathbf{D}} = \begin{bmatrix} -\frac{1}{6} & 0 \\ 0 & -\frac{1}{3} \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{Re} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

UMass Dartmouth

Unconditionally SSP implicit multiderivative schemes

Fourth order A five-stage, fourth order unconditionally SSP implicit two-derivative Runge–Kutta method is given by the Shu-Osher coefficients

$$diag(\mathbf{D}) = \begin{bmatrix} 0.660949255604937 \\ 0.242201390400848 \\ 1.137542996287740 \\ 0.191388711018110 \\ 0.625266691721946 \end{bmatrix}, \quad diag(\dot{\mathbf{D}}) = \begin{bmatrix} -0.177750705279127 \\ -0.354733903778084 \\ -0.403963513682271 \\ -0.161628266349058 \\ -0.218859021269943 \end{bmatrix},$$

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0.084036809261019 & 0.915963190738981 & 0 & 0 & 0 \\ 0.001511648458457 & 0 & 0.090254853867587 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{Re} = [1, 0, 0, 0.908233497673956, 0]^T .$$

UMass Dartmouth

IMEX schemes

We don't often deal with problems $u_t = \mathbf{G}(u)$ that have these properties, but we do if we consider equations of the form $u_t = \mathbf{F}(u) + \mathbf{G}(u)$,

- $\mathbf{F}$ satisfies a forward Euler condition where the time-step restriction is of a reasonable size (i.e. $\mathbf{F}$ is non-stiff),
- $\mathbf{G}$ satisfies an implicit Euler condition, with no time-step restriction.
- $\dot{\mathbf{G}}$ satisfies a negative derivative condition, with no time-step restriction.

Such equations are common in kinetic models; e.g. BGK, Broadwell, etc. These have a transport term $\mathbf{F}$ and stiff collision term $\mathbf{G}$ that satisfy these conditions.

SSP IMEX schemes

An s-stage multi-derivative IMEX method ($\mathcal{C} = r$)

$$
\begin{aligned}
u^{(i)} &= r_i u^n + \sum_{j=1}^{i-1} p_{ij} u^{(j)} + \sum_{j=1}^{i-1} w_{ij} \left(u^{(j)} + \frac{\Delta t}{r} F(u^{(j)}) \right) \\
&\quad + \Delta t d_{ii} G(u^{(i)}) + \Delta t^2 \dot{d}_{ii} \dot{G}(u^{(i)}), \qquad i = 1, ..., s, \\
u^{n+1} &= u^{(s)}.
\end{aligned}
$$

In matrix form,

$$
U = \mathbf{R}\mathbf{e}u^n + \mathbf{P}U + \mathbf{W}\left(U + \frac{\Delta t}{r} F(U) \right) + \Delta t \mathbf{D} G(U) + \Delta t^2 \dot{\mathbf{D}} \dot{G}(U),
$$

where $\mathbf{P}$, $\mathbf{W}$, and $\mathbf{R} = I - \mathbf{P} - \mathbf{W}$ are $s \times s$ matrices, $\mathbf{D}$ and $\dot{\mathbf{D}}$ are $s \times s$ diagonal matrices, and $\mathbf{e}$ is a vector of ones.

SSP IMEX schemes with second derivative

Given operators F and G that satisfy for some $\| \cdot \|$:

Forward Euler condition on $\mathbf{F}$

$$\|u + \Delta t F(u)\| \leq \|u\| \quad \text{for all} \quad \Delta t \leq \Delta t_{\text{FE}},$$

for some $\Delta t_{\text{FE}} > 0$

Implicit Euler condition on $\mathbf{G}$

$$u^{n+1} = u^n + \Delta t \mathbf{G}(u^{n+1}) \quad \Longrightarrow \quad \|u^{n+1}\| \leq \|u^n\| \quad \forall \Delta t > 0.$$

Implicit negative derivative condition on $\dot{\mathbf{G}} = \mathbf{G}'(u)\mathbf{G}(u)$

$$u^{n+1} = u^n - \Delta t^2 \dot{\mathbf{G}}(u^{n+1}) \quad \Longrightarrow \quad \|u^{n+1}\| \leq \|u^n\| \quad \forall \Delta t > 0.$$

UMass Dartmouth

SSP IMEX schemes with second derivative

Under the conditions above on $\mathbf{F}$, $\mathbf{G}$, and $\dot{\mathbf{G}}$, if the method

$$U = \mathbf{Re}u^n + \mathbf{P}U + \mathbf{W}\left(U + \frac{\Delta t}{r}\mathbf{F}(U)\right) + \Delta t\mathbf{D}\mathbf{G}(U) + \Delta t^2\dot{\mathbf{D}}\dot{\mathbf{G}}(U),$$

satisfies

$$\mathbf{Re} \geq 0, \quad \mathbf{P} \geq 0, \quad \mathbf{W} \geq 0, \quad \mathbf{D} \geq 0, \quad \dot{\mathbf{D}} \leq 0,$$

then it preserves the strong stability property $\|u^{n+1}\| \leq \|u^n\|$ under the time-step condition $\Delta t \leq r\Delta t_{\mathsf{FE}}$.

We found such methods of up to order $p = 3$.

There is no time step restriction coming from $\mathbf{G}$, which is unusual for IMEX methods!

UMass Dartmouth

We begin with a method that has Shu-Osher coefficients

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{bmatrix}, \quad \mathbf{Re} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

and

$$diag(\mathbf{D}) = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}, \quad diag(\dot{\mathbf{D}}) = -\begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix},$$

with $r = 1$.

Looser conditions

If we allow the final stage to be explicit and contain only $\mathbf{F}$ terms, we get a larger SSP coefficient $\mathcal{C} = (1+\sqrt{2})/2 > 1$:

$$
\begin{aligned}
y^{(1)} &= u^n + \frac{1}{2+\sqrt{2}}\Delta t \mathbf{G}(y^{(1)}) - \frac{1}{2+\sqrt{2}}\Delta t^2 \dot{\mathbf{G}}(y^{(1)}) \\
y^{(2)} &= \left(y^{(1)} + \frac{\Delta t}{r}\mathbf{F}(y^{(1)}) \right) \\
y^{(3)} &= \frac{6-\sqrt{2}}{8} y^{(1)} + \frac{2+\sqrt{2}}{8}\left(y^{(2)} + \frac{\Delta t}{r}\mathbf{F}(y^{(2)}) \right) + \frac{1}{\sqrt{2}}\Delta t \mathbf{G}(y^{(3)}) \\
u^{n+1} &= \left(1 - \frac{\sqrt{2}}{4}\right) y^{(3)} + \frac{2+\sqrt{2}}{4(1+\sqrt{2})}\left(y^{(3)} + \frac{\Delta t}{r}\mathbf{F}(y^{(3)}) \right).
\end{aligned}
$$

UMass Dartmouth

Third order SSP IMEX schemes

This method has $r = 0.904402174130635$ with:

$$diag(\mathbf{D}) = \begin{pmatrix} 0 \\ 2 \\ 0.388820513661584 \\ 0.083529464436389 \\ 1.793313488277995 \\ 0 \end{pmatrix},$$

$$diag(\dot{\mathbf{D}}) = - \begin{pmatrix} 0.871358934880525 \\ 0.856842702601821 \\ 0 \\ 0 \\ 2 \\ 0.205134529930013 \end{pmatrix}.$$

Note that $d_{ii} + \left|\dot{d}_{ii}\right| > 0$ for each stage i.

Third order SSP IMEX scheme

$$\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.058453072749259 & 0 & 0 & 0 & 0 & 0 \\ 0.764266518291495 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.292520982667463 & 0 & 0 & 0 \\ 0.173788618990251 & 0 & 0 & 0.281050180194829 & 0 & 0 \\ 0.016811671845949 & 0 & 0 & 0.448630511341543 & 0 & 0 \end{pmatrix},$$

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.253395246357353 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.235733481708505 & 0 & 0 & 0 & 0 \\ 0 & 0.123961833526104 & 0 & 0 & 0 & 0 \\ 0.409037644509411 & 0.136123556305509 & 0 & 0 & 0 & 0 \\ 0.203353399602184 & 0 & 0 & 0 & 0.331204417210324 & 0 \end{pmatrix},$$

$$\mathbf{Re} = \begin{pmatrix} 1 \\ 0.688151680893388 \\ 0 \\ 0.583517183806433 \\ 0 \\ 0 \end{pmatrix}.$$

UMass Dartmouth

Second order with $(k \geq 3)$-step two-derivative two-stages

More steps allow larger SSP coefficients, e.g. a family of methods with $r = \mathcal{C} = \frac{k-2}{k-1}$

$$
\begin{aligned}
y^{(1)} &= u^n - (k-1)\Delta t^2 \dot{\mathbf{G}}(y^{(1)}) \\
y^{(2)} &= \frac{1}{k-1} u^{n-k+1} + \frac{k-2}{k-1}\left(y^{(1)} + \frac{\Delta t}{r}\mathbf{F}(y^{(1)}) \right) \\
&\quad + k\Delta t \mathbf{G}(y^{(2)}) - k\Delta t^2 \dot{\mathbf{G}}(y^{(2)}) \\
u^{n+1} &= \frac{1}{k-1} y^{(2)} + \frac{k-2}{k-1}\left(y^{(1)} + \frac{\Delta t}{r}\mathbf{F}(y^{(1)}) \right)
\end{aligned}
$$

Families of problems that satisfy these conditions

The negative derivative is naturally found as part of kinetic equations of the form

$$u_t = T(u) + \frac{1}{\varepsilon} Q(u)$$

where T satisfies a forward Euler condition

$$\|u^n + \Delta t T(u^n)\| \leq \|u^n\| \quad \forall \Delta t \leq \Delta t_F$$

and Q satisfies a forward Euler condition

$$\|u^n + \Delta t Q(u^n)\| \leq \|u^n\| \quad \forall \Delta t \leq \Delta t_Q$$

and a negative derivative condition

$$\|u^n - \Delta t^2 \dot{Q}(u^n)\| \leq \|u^n\| \quad \text{for} \quad \Delta t \leq \Delta t_{\dot{Q}}$$

Numerical Results: Broadwell Model

The Broadwell model is a simple discrete velocity kinetic model:

$$\begin{cases} \partial_t f_+ + \partial_x f_+ = \dfrac{1}{\varepsilon}(f_0^2 - f_+ f_-), \\ \partial_t f_0 = -\dfrac{1}{\varepsilon}(f_0^2 - f_+ f_-), \\ \partial_t f_- - \partial_x f_- = \dfrac{1}{\varepsilon}(f_0^2 - f_+ f_-), \end{cases} \tag{1}$$

where $f_+ = f_+(t,x)$, $f_0 = f_0(t,x)$, and $f_- = f_-(t,x)$ denote the densities of particles with speed 1, 0, and -1, respectively.
Define $f = (f_+, f_0, f_-)^T$,
$T(f) = (-\partial_x f_+, 0, \partial_x f_-)^T$,
$Q(f) = (f_0^2 - f_+ f_-, -(f_0^2 - f_+ f_-), f_0^2 - f_+ f_-)^T$
(these should be defined for the system after spatial discretization).

UMass Dartmouth

Numerical Results: Broadwell Model

Consider the Broadwell model (1) on the domain $x \in [0, 2]$
with periodic boundary condition
and inconsistent initial data

$$
\begin{aligned}
f_+(0,\cdot) &= 1 + 0.2\exp(0.3\sin(\pi x)), \quad f_-(0,\cdot) = \exp(0.2\cos(2\pi x)), \\
f_0(0,\cdot) &= \frac{1}{1 + 0.3\sin(\pi x)}.
\end{aligned}
$$

We discretize in space by the fifth order finite volume WENO scheme, and the collision operator Q is evaluated pointwise on the Gauss quadrature points in each cell.

We fix the the CFL number as $\Delta t = \frac{1}{2}\Delta x$, and solve (1) by the new second and third order methods up to final time $T = 0.1$.

Numerical Results: Broadwell Model

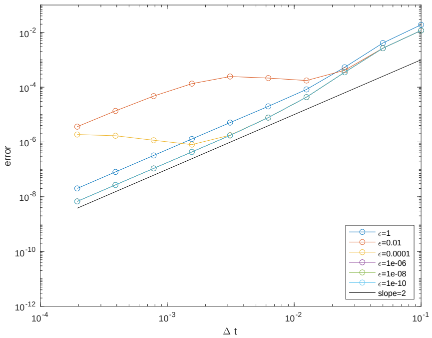


Figure: Accuracy test of the new second order IMEX scheme for the Broadwell model.

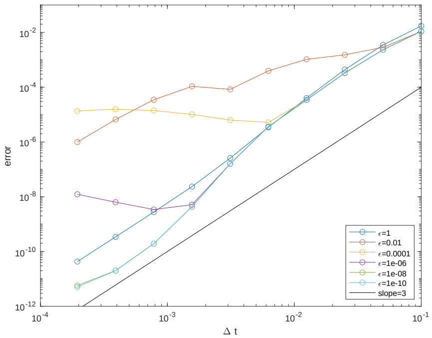


Figure: Accuracy test of the new third order IMEX scheme for the Broadwell model.

UMass Dartmouth

Example: The Bhatnagar-Gross-Krook (BGK) model

The negative derivative condition can be found in (part of) the BGK equation, a widely used kinetic model introduced to mimic the full Boltzmann equation:

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}(M - f), \quad x, v \in \mathbb{R}^d,$$

where $f = f(t, x, v)$ is the probability density function and M is the Maxwellian:

$$M(t, x, v) = \frac{\rho(t, x)}{(2\pi T(t, x))^{d/2}} \exp\left(-\frac{|v - u(t, x)|^2}{2T(t, x)}\right),$$

the density ρ, bulk velocity u and temperature T are given by the moments of f.

UMass Dartmouth

The Bhatnagar-Gross-Krook (BGK) model

We see that the BGK model has the property

$$\dot{Q}(f) = Q'Q = -Q(f).$$

So that if Q satisfies a forward Euler condition

$$\|u^n + \Delta t Q(u^n)\| \leq \|u^n\| \quad \forall \Delta t \leq \Delta t_Q$$

then $\dot{Q}$ will automatically satisfy the negative derivative condition

$$\|u^n - \Delta t^2 \dot{Q}(u^n)\| = \|u^n + \Delta t^2 Q(u^n)\| \leq \|u^n\| \quad \text{for} \quad \Delta t \leq \sqrt{\Delta t_Q}$$

UMass | Dartmouth

Numerical Results: The BGK model

The BGK model is a widely used kinetic model introduced to mimic the full Boltzmann equation:

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}(M - f), \quad x, v \in \mathbb{R}^d, \tag{2}$$

where $f = f(t, x, v)$ is the probability density function and M is the so-called Maxwellian given by

$$M(t, x, v) = \frac{\rho(t, x)}{(2\pi T(t, x))^{d/2}} \exp\left(-\frac{|v - u(t, x)|^2}{2T(t, x)}\right), \tag{3}$$

where the density ρ, bulk velocity u and temperature T are given by the moments of f:

$$\rho = \int_{\mathbb{R}^d} f \, \mathrm{d}v, \quad \rho u = \int_{\mathbb{R}^d} f v \, \mathrm{d}v, \quad \frac{1}{2}\rho d T = \frac{1}{2}\int_{\mathbb{R}^d} f |v - u|^2 \, \mathrm{d}v. \tag{4}$$

UMass Dartmouth

Numerical Results: The BGK model

Consider the 1D BGK model on the physical domain $x \in [0, 2]$ with periodic boundary condition, and inconsistent initial data given by

$$f(0, x, v) = 0.7M[\tilde{\rho}(x), \tilde{u}(x), \tilde{T}(x)](v) + 0.3M[\tilde{\rho}(x), -0.5\tilde{u}(x), \tilde{T}(x)](v),$$

with

$$\tilde{\rho}(x) = 1 + 0.2\sin(2\pi x), \quad \tilde{u}(x) = 1, \quad \tilde{T}(x) = \frac{1}{1 + 0.2\sin(\pi x)}.$$

The velocity domain is truncated into $[-v_{max}, v_{max}]$ with $v_{max} = 15$ and discretized with $N_v = 150$ grid points, and the physical space is discretized in the same way as for the Broadwell model.
We fix the the CFL number as $\Delta t = \frac{1}{2}\frac{\Delta x}{v_{max}}$.

UMass Dartmouth

Numerical Results: The BGK model

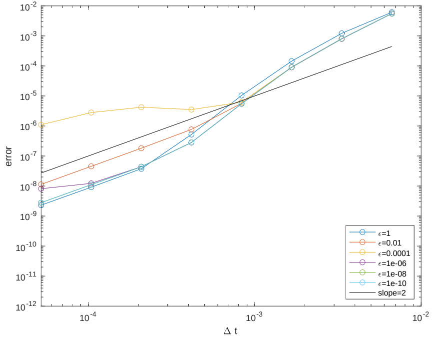


Figure: Accuracy test of the new second order IMEX scheme for the BGK model.

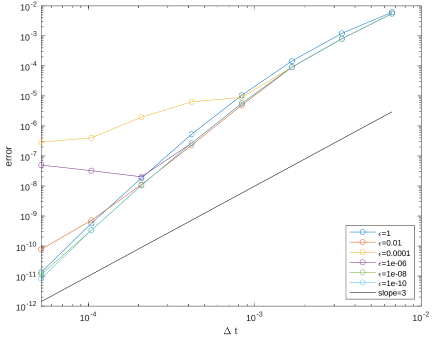


Figure: Accuracy test of the new third order IMEX scheme for the BGK model.

UMass Dartmouth

Additive SSP methods

- Additive methods handle T and Q with different coefficients, and ensure the SSP property will be preserved for the combination.
- We want methods that will handle T with downwinding because it is a transport term, so that T satisfies a forward Euler condition, and has a downwind version $\tilde{T}$.
- Because Q satisfies a forward Euler condition and a negative derivative condition, we want the additive method to handle Q with an implicit term only and an implicit negative derivative term only.

Additive SSP methods

If T and Q satisfy the conditions above, then an additive method of the form

$$
\begin{aligned}
u^{(i)} &= r_i u^n + \sum_{j=1}^{i-1} p_{ij} u^{(j)} + \sum_{j=1}^{i-1} w_{ij} \left(u^{(j)} + \frac{\Delta t}{r} T(u^{(j)}) \right) \\
&\quad + \sum_{j=1}^{i-1} \tilde{w}_{ij} \left(u^{(j)} - \frac{\Delta t}{r} T(u^{(j)}) \right) \\
&\quad + \frac{1}{\varepsilon} \Delta t d_{ii} Q(u^{(i)}) + \frac{1}{\varepsilon^2} \Delta t^2 \dot{d}_{ii} \dot{Q}(u^{(i)}), \qquad i = 1, ..., s, \\
u^{n+1} &= u^{(s)}
\end{aligned}
$$

is SSP with $\mathcal{C} = r$ if $r_i, p_{ij}, w_{ij}, \tilde{w}_{ij}$ and d_{ii} are all non-negative and $\dot{d}_{ii}$ is non-positive.

UMass Dartmouth

In matrix form

$$U = \mathbf{R}\mathbf{e}u^n + \mathbf{P}U + \mathbf{W}\left(U + \frac{\Delta t}{r}T(U)\right) + \tilde{\mathbf{W}}\left(U - \frac{\Delta t}{r}\tilde{T}(U)\right)$$

$$+\frac{\Delta t}{\epsilon}\mathbf{D}Q(U) + \frac{\Delta t^2}{\epsilon}\dot{\mathbf{D}}\dot{Q}(U),$$

where $\mathbf{P}$, $\mathbf{W}$, $\tilde{\mathbf{W}}$, and $\mathbf{R} = I - \mathbf{P} - \mathbf{W} - \tilde{\mathbf{W}}$ are $s \times s$ matrices,
r_i is the $i-$th row sum of $\mathbf{R}$,
$\mathbf{D}$ and $\dot{\mathbf{D}}$ are $s \times s$ diagonal matrices,
and $\mathbf{e}$ is a vector of ones.

If all elements of $\mathbf{R}\mathbf{e}, \mathbf{P}, \mathbf{W}, \tilde{\mathbf{W}}$ and $\mathbf{D}$ are non-negative, and the elements of $\dot{\mathbf{D}}$ are non-positive,
then this method is SSP with $\mathcal{C} = r$.

Can we find additive SSP methods with arbitrarily large r?

Ongoing work by Andrew Christlieb, **Sining Gong**, Sigal Gottlieb, and Zack Grant found an unconditional additive method using downwinding for T and the second derivative for Q:

$$
\begin{aligned}
u^{(i)} &= r_i u^n + \sum_{j=1}^{s} p_{ij} u^{(j)} + \sum_{j=1}^{s} w_{ij} \left(u^{(j)} + \frac{\Delta t}{r} T(u^{(j)}) \right) \\
&\quad + \sum_{j=1}^{s} \tilde{w}_{ij} \left(u^{(j)} - \frac{\Delta t}{r} \tilde{T}(u^{(j)}) \right) \\
&\quad + \frac{1}{\varepsilon} \Delta t d_{ii} Q(u^{(i)}) + \frac{1}{\varepsilon^2} \Delta t^2 \dot{d}_{ii} \dot{Q}(u^{(i)}), \;\; i = 1, ..., s \\
u^{n+1} &= u^{(s)}.
\end{aligned}
$$

UMass | Dartmouth

Unconditional additive second order SSP methods

The coefficients are given by:

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & \frac{2}{4+r} \\ \frac{1}{2r^2} & \frac{5}{13} & \frac{3}{8r} - \frac{3}{4r^2} \\ 0 & \frac{8}{13} & 0 \end{bmatrix}, \qquad \tilde{\mathbf{W}} = \begin{bmatrix} \frac{2+r}{4+r} & 0 & 0 \\ 0 & 0 & \frac{8}{13} - \frac{1}{4r^2} - \frac{3}{8r} \\ 0 & 0 & \frac{5}{13} \end{bmatrix}$$

$$diag(\mathbf{D}) = \begin{bmatrix} \frac{2}{4+r} \\ 0 \\ 0 \end{bmatrix}, \quad diag(\dot{\mathbf{D}}) = \begin{bmatrix} 0 \\ -\frac{3}{8r^2} \\ -\frac{3}{8r^2} \end{bmatrix}, \quad \mathbf{Re} = \begin{bmatrix} 0 \\ \frac{1}{2r^2} \\ 0 \end{bmatrix}, \quad \mathbf{P} = \mathbf{0},$$

Note that when $r \geq 2$, all the coefficients here are positive.

Unconditional Additive SSP methods

Given

$$u_t = T(u) + \frac{1}{\epsilon} Q(u),$$

where

- T satisfies a forward Euler condition for some $\Delta t_{\text{FE}} > 0$.
- $\tilde{T}$ satisfies a backward Euler condition for some $\Delta t_{\text{DW}} > 0$.
- $\frac{1}{\epsilon} Q$ satisfies an implicit Euler condition, with no time-step restriction.
- $\frac{1}{\epsilon^2} \dot{Q}$ satisfies a negative derivative condition, with no time-step restriction.

We found a family of second order SSP methods with arbitrarily large $\mathcal{C}$!

Conclusions

Downwinding for the transport term and negative derivative conditions for the collision term give good* additive SSP methods!

Conclusions

SSP properties depend on properties of the underlying problem and discretization, so that it is helpful to have different methods

- Optimal SSP Runge–Kutta methods with forward Euler and second derivative building blocks are in A.J. Christlieb, S. Gottlieb, Z. Grant, and D. C. Seal, JSC (2016)
- Optimal SSP Runge–Kutta methods with forward Euler and Taylor series building blocks are in Z. Grant, S. Gottlieb, D.C. Seal, CAMC (2019)
- Optimal SSP Runge–Kutta methods with forward Euler and negative derivative building blocks, S. Gottlieb, Z. Grant, J. Hu, R. Shu, SINUM (2022)
- A review as well as an extension to multi-step, multi-stage, two-derivative methods in Applied Mathematical Modelling Volume 149, January 2026.
- **Ongoing** Unconditional SSP Runge–Kutta methods using downwinding for the transport term and the implicit negative derivative building blocks for the collision term A. Christlieb, S. Gong, S. Gottlieb, Z. Grant

Conclusions

Thank You!