Discontinuous Galerkin methods for first-order PDEs

Alexandre Ern

Université Paris-Est, CERMICS (ENPC)

CNAM, Paris, 19.06.2015
Outline

- **Friedrichs’ systems** (steady linear PDEs)
  - design of dG methods
  - convergence analysis for smooth solutions
  - unified view on linear stabilization
  - cf. [AE & Guermond, 06-..], [Di Pietro & AE, 12]

- **dG in time** (time-dependent linear PDEs)
  - convergence analysis for smooth solutions
  - cf. [AE & Schieweck, 15]

- **Weighting linear stabilization** (conservation laws)
  - linear stabilization for rough solutions/nonlinear PDEs
  - cf. [AE & Guermond, 13]
Friedrichs’ systems

- Open, bounded, connected, strongly Lipschitz subset $\Omega \subset \mathbb{R}^d$
- $\mathbb{K}^m$-valued functions, $m \geq 1$ and $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$
- $(d + 1)$ functions $\mathcal{K}, \{A^k\}_{1 \leq k \leq d} : \Omega \to \mathbb{K}^{m \times m}$
  - $\mathcal{K}, \{A^k\}_{1 \leq k \leq d}$ and $\mathcal{X} := \sum_{k=1}^{d} \partial_k A^k$ are bounded
  - $A^k$ is symmetric (Hermitian)
  - $\mathcal{K} + \mathcal{K}^H - \mathcal{X}$ is uniformly positive ($\geq 2\mu_0 I$)

- Given $f : \Omega \to \mathbb{K}^m$, find $u : \Omega \to \mathbb{K}^m$ s.t. $Au = f$ in $\Omega$ with

$$Au = \mathcal{K}u + \sum_{k=1}^{d} A^k \partial_k u$$

- cf. [Friedrichs, 58]
Examples

▶ Advection-reaction $m = 1$, $K = \mathbb{R}$
   - $\mu u + \beta \cdot \nabla u = f$
   - $\mu \in L^\infty$, $\beta \in L^\infty$, $\nabla \cdot \beta \in L^\infty$, $\mu - \frac{1}{2} \nabla \cdot \beta \geq \mu_0 > 0$

▶ Darcy (grad-div) $m = d + 1$, $K = \mathbb{R}$
   - $u = (\sigma, p)$, $\mathfrak{d}^{-1} \sigma + \nabla p = f_1$, $\mu p + \nabla \cdot \sigma = f_2$
   - $\mu \in L^\infty$ and uniformly positive, $\mathfrak{d}$ bounded, symmetric, uniformly positive definite

▶ Maxwell (eddy currents, curl-curl) $m = 6$, $K = \mathbb{C}$
   - $u = (E, H)$, $\sigma E - \nabla \times H = f_1$, $i\omega \mu H + \nabla \times E = 0$
   - $\sigma, \mu \in L^\infty$, uniformly positive (for simplicity)
Boundary conditions

- Symmetric boundary field $\mathcal{N} : \partial \Omega \rightarrow \mathbb{K}^m$ (unit outward normal $n$)
  \[ \mathcal{N} = \sum_{k=1}^{d} n_k A^k \]

- Assume there is an additional boundary field $\mathcal{M} : \partial \Omega \rightarrow \mathbb{K}^m$
  \[ (\text{real part of}) \quad \mathcal{M} \quad \text{is non-negative} \]
  \[ \ker(\mathcal{M} - \mathcal{N}) + \ker(\mathcal{M} + \mathcal{N}) = \mathbb{K}^m \]

- The boundary condition is $\mathcal{M} - \mathcal{N} u = 0$ on $\partial \Omega$

- Examples
  - advection-reaction $\mathcal{N} u = (\beta \cdot n) u$, $\mathcal{M} u = |\beta \cdot n| u$
  - Darcy $\mathcal{N}(\sigma, p) = (pn, \sigma \cdot n)$, $\mathcal{M}(\sigma, p) = (\pm pn, \mp \sigma \cdot n)$
  - Maxwell $\mathcal{N}(E, H) = (H \times n, E \times n)$, $\mathcal{M}(E, H) = (\pm H \times n, \mp E \times n)$
  - note that $\mathcal{M}$ is skew-symmetric for Darcy and Maxwell
Mathematical theory

- \( L^2 \)-based theory: pivot space \( L = L^2(\Omega; \mathbb{K}^m) \)

- Graph space \( V = \{ v \in L \mid Av \in L \} \)
  - Friedrichs’ operator \( Av = \mathcal{K}v + \sum_{k=1}^{d} A^k \partial_k v \)
  - formal adjoint \( \tilde{A}v = (\mathcal{K}^H - \mathcal{X})v - \sum_{k=1}^{d} A^k \partial_k v \)
  - \( A, \tilde{A} \in \mathcal{L}(V; L) \)

- Boundary operators \( N, M \in \mathcal{L}(V; V') \)
  - \( \langle Nv, w \rangle_{V', V} = (Av, w)_L - (v, \tilde{A}w)_L \)
  - \( \langle Mv, v \rangle_{V', V} \geq 0 \) and \( \ker(M - N) + \ker(M + N) = V \)
  - \( L \)-dissipativity on \( \ker(M - N) \): \( (Av, v)_L \geq \mu_0 \| v \|_L^2 + \frac{1}{2} \langle Mv, v \rangle_{V', V} \)

- Given \( f \in L \), there is a unique \( u \in V \) s.t.

\[
Au = f \quad (M - N)u = 0
\]

(and there is a unique \( \tilde{u} \in V \) s.t. \( \tilde{A}\tilde{u} = f \) and \( (M^* + N)\tilde{u} = 0 \)
dG setting

- Admissible mesh sequence \( \{ \mathcal{T}_h \}_{h>0} \)
  - matching simplicial meshes: Ciarlet's *shape-regularity*
  - general meshes (non-matching, polyhedral): *shape- and contact-regularity*, essentially one length scale for mesh faces and cells [Di Pietro & AE, 12]
  - usual FE tools: inverse & discrete trace ineq., polynomial approx.

- Broken polynomial space (of order \( r \geq 0 \))

\[
P_r(\mathcal{T}_h; \mathbb{R}) = \{ v_h \in L^1(\Omega; \mathbb{R}) \mid v_h|_T \in P_r(T; \mathbb{R}) \forall T \in \mathcal{T}_h \}
\]

- Jumps and averages at mesh interfaces

\[
\begin{align*}
F &= \partial T_l \cap \partial T_r \\
n_F \text{ points from } T_l \text{ to } T_r \\
\{ v \} &= \frac{1}{2} (v|_{T_l} + v|_{T_r}) \\
[v] &= v|_{T_l} - v|_{T_r}
\end{align*}
\]
dG approximation: centered fluxes

- Standard Galerkin setting with $V_h = \mathbb{P}_r(\mathcal{T}_h; \mathbb{K}_m)$

  Find $u_h \in V_h$ s.t. $a^\text{cf}_h(u_h, w_h) = (f, w_h)_L$ for all $w_h \in V_h$

  with discrete bilinear form $a^\text{cf}_h$ satisfying two key properties
  - exact consistency $a^\text{cf}_h(u, w_h) = (f, w_h)_L$, $\forall w_h \in V_h$
  - $L$-dissipativity $a^\text{cf}_h(v_h, v_h) \geq \mu_0 \|v_h\|_L^2 + \frac{1}{2} (M v_h, v_h)_L(\partial \Omega)$, $\forall v_h \in V_h$

  - Centered fluxes (interfaces $F \in \mathcal{F}^i_h$) and boundary penalty ($F \in \mathcal{F}^b_h$)

    $a^\text{cf}_h(v_h, w_h) = \sum_{T \in \mathcal{T}_h} (v_h, \tilde{A}w_h)_L(T) + \sum_{F \in \mathcal{F}^i_h} (\phi^i_F(v_h), [w_h])_L(F) + \sum_{F \in \mathcal{F}^b_h} (\phi^b_F(v_h), w_h)_L(F)$

    - $\phi^i_F(v_h) = \mathcal{N}_F \{v_h\}$ (for AR, $\phi^i(v_h) = (\beta \cdot n_F) \{v_h\}$)
    - $\phi^b_F(v_h) = \frac{1}{2} (M_+ - \mathcal{N}) v_h$, $M_+ = M = |\beta \cdot n|$ for AR, $M_+$ adds least-squares penalty on BC for Darcy and Maxwell

  - For smooth solution $u \in H^{r+1}(\Omega; \mathbb{K}_m)$, $\|u - u_h\|_L \lesssim h^r$
Linear stabilization (upwinding)

- Upwinding amounts to adding a least-squares penalty on interface jumps [Brezzi et al., 04]

\[ a_h(v_h, w_h) = a_h^{cf}(v_h, w_h) + \sum_{F \in F_h} (S_F[v_h], [w_h])_{L(F)} \]

with \( S_F \sim |N_F| \), so that

- \( a_h \) is still exactly consistent
- \( a_h \) enjoys strengthened \( L \)-dissipativity

\[ a_h(v_h, v_h) \geq \|v_h\|^2 = \mu_0 \|v_h\|^2_L + \frac{1}{2} (M v_h, v_h)_{L(\partial\Omega)} + \sum_{F \in F_h} \|S_F^{1/2}[v_h]\|^2_{L(F)} \]

- Incidence on the flux: \( \phi^i_F(v_h) = N_F \{v_h\} + S_F[v_h] \)
  - for AR, \( S_F = \frac{1}{2} |\beta \cdot n_F| \) leads to \( \phi^i_F(v_h) = (\beta \cdot n_F) u_h^{\uparrow} \)
  - for Darcy, jumps of both \( \sigma_h \cdot n_F \) and \( p_h \) are penalized
  - for Maxwell, jumps of both \( \mathbf{H}_h \times n_F \) and \( \mathbf{E}_h \times n_F \) are penalized
Error analysis with upwinding

- Assume smooth solution $u \in H^{r+1}(\Omega; \mathbb{K}^m)$

- Strengthened $L$-dissipativity leads to $\|u - u_h\| \lesssim h^{r+1/2} \to$ quasi-optimal $L$-norm estimate

- Full stability norm and discrete inf-sup stability

$$\|v_h\|_\# \lesssim \sup_{w_h \in V_h} \frac{a_h(v_h, w_h)}{\|w_h\|_\#} \quad \forall v_h \in V_h$$

$$\|v_h\|_\#^2 = \|v_h\|^2 + \sum_{T \in T_h} h_T \|Av_h\|_{L(T)}^2$$

- We obtain $\|u - u_h\|_\# \lesssim h^{r+1/2} \to$ optimal graph-norm estimate

- For mixed elliptic PDEs, it is possible to modify the penalty strategy so as to eliminate locally the auxiliary variable
Unified view on linear stabilization

- Many recent $H^1$-conforming stabilized FEM are analyzed with the same tools and lead to similar error estimates

- Example: Continuous interior penalty

$$a_h^{cip}(v_h, w_h) = (v_h, \tilde{A}w_h)_L + \sum_{F \in \mathcal{F}_h^i} (S_F[\nabla v_h], [\nabla w_h])_{L(F)} + \sum_{F \in \mathcal{F}_h^b} (\phi_F^b(v_h), w_h)_{L(F)}$$
  - penalizes gradient jumps with $S_F \sim h_F^2$
  - cf. [Burman & Hansbo, 04; Burman, 05; Burman & AE, 07]

- Other examples
  - Subgrid Viscosity penalizes gradient of subscale fluctuation, cf. [Guermond, 99]
  - Local Projection Stabilization penalizes subscale fluctuation of gradient, cf. [Braack & Burman, 06; Matthies et al., 07]

- Stabilization bilinear form is symmetric (contrast with GaLS/SUPG)
Time-dependent linear PDEs

- Overview
- Main results
- Some analysis tools
- Error estimates for smooth solutions
dG in time

▶ **Time semi-discretization of evolution problem by dG method**
- piecewise polynomials in time of order \( k \geq 0 \)
- time interval \( I = (0, T] \) decomposed as \( I = \bigcup_{n=1}^{N} I_n \)
- subintervals \( I_n = (t_{n-1}, t_n] \) *(open at left, closed at right endpoint)*
- discrete times \( 0 = t_0 < t_1 < \cdots < t_N = T \), time steps \( \tau_n = t_n - t_{n-1} \)

▶ For Banach space \( B \) (functions in space), let

\[
\mathbb{P}_k(I_n, B) = \{ w : I_n \to B : w(t) = \sum_{j=0}^{k} W^j t^j, \forall t \in I_n, W^j \in B, \forall j \} 
\]

\[
X^k_{\tau}(B) = \{ w_\tau : \bar{I} \to B : w_\tau|_{I_n} \in \mathbb{P}_k(I_n, B) \quad \forall n \}
\]

▶ a function \( w_\tau \in X^k_{\tau}(B) \) can be **discontinuous** at discrete times \( t_n \)
and is **continuous from the left** at all \( t_n \)

▶ jump of \( w_\tau \) at \( t_n \) is \( [w_\tau]_n = w_\tau(t_n^+) - w_\tau(t_n^-) \)
Evolution problems with coercivity

- **Parabolic problems**: dG in time (order $k$), dG in space (order $r$) [Thomée, 07]
  - $\ell^\infty(L^2)$ (at discrete time nodes) and $L^2(L^2)$ error estimates of order $\tau^{2k+1} + h^{r+1}$: **super-convergence in time**

- **Nonlinear advection-diffusion**, dG in space
  - $\ell^\infty(L^2)$ and $L^2(L^2)$ estimates of order $(\tau^{k+1} + h^r)$ on time-varying meshes (under condition $h^2 \lesssim \tau$) [Feistauer et al., 11-..]

- **Linear advection-diffusion**, $H^1$-conforming FEM with LPS
  - $\ell^\infty(L^2)$ and $L^2(L^2)$ estimates of order $(\tau^{k+1} + h^{r+1/2} + \varepsilon^{1/2}h^r)$ on static meshes [Ahmed, Matthies, Tobiska & Xie, 11]
Evolution problems without coercivity

- Linear first-order operator $A \nu = \mu \nu + \beta \cdot \nabla \nu$ in space
  - $\mu : \Omega \to \mathbb{R}$ is a bounded reaction function
  - $\beta : \Omega \to \mathbb{R}^d$ is a given Lipschitz advection field
  - both are time-independent

- Mathematical setting of Friedrichs’ systems (spaces $V$ and $L$)

- Linear evolution problem
  - data $f \in C^0([0, T], L)$ and $u_0 \in V$
  - find $u \in C^0([0, T], V) \cap C^1([0, T], L)$ s.t.

  \[
  (\partial_t u(t), \nu)_L + (Au(t), \nu)_L = (f(t), \nu)_L \quad \forall \nu \in L \quad \forall t \in (0, T)
  \]

  and $u(0) = u_0$

- well-posedness results from Hille–Yosida Theorem
dG-in-time semi-discretization

- Time semi-discrete solution $u_{\tau}$ belongs to $X^k_{\tau}(V)$
- For all $n = 1 \ldots N$, $u_{\tau}|_{I_n} \in P_k(I_n, V)$ s.t. for all $v_{\tau} \in P_k(I_n, L)$,
  \[
  \int_{I_n} \left( \partial_t u_{\tau} + Au_{\tau}, v_{\tau} \right)_L \, dt + \left( [u_{\tau}]_{n-1}, v_{\tau}(t^+_{n-1}) \right)_L = \int_{I_n} (f, v_{\tau})_L \, dt
  \]
  $(k + 1)$ coupled first-order PDEs in space within each time step
- RHS evaluated using the $(k + 1)$-point right-sided GR quadrature on each subinterval $I_n$
  \[
  Q_n(g) = \frac{\tau_n}{2} \sum_{\mu=1}^{k+1} \hat{w}_\mu g(t_{n,\mu}) \approx \int_{I_n} g(t) \, dt
  \]
  - weights $\hat{w}_\mu > 0$, $t_{n,k+1} = t_n$, $Q_n(g)$ exact for all $g \in P_{2k}(I_n, \mathbb{R})$
- Time semi-discrete problem with quadrature becomes
  \[
  \int_{I_n} \left( \partial_t u_{\tau} + Au_{\tau}, v_{\tau} \right)_L \, dt + \left( [u_{\tau}]_{n-1}, v_{\tau}(t^+_{n-1}) \right)_L = Q_n((f, v_{\tau})_L)
Full space-time discretization

- Discrete space $V^n_h \subset L$ built from a mesh $\mathcal{T}^n_h$ which can change from one time interval to the next

- FEM with linear stabilization (dG, CIP, ...)
  - $A^n_h : V^n_h \to V^n_h$ s.t. $(A^n_h v_h, w_h)_L = a^n_h(v_h, w_h)$ ($a^n_h$ depends on $\mathcal{T}^n_h$ ...)

- **Fully discrete problem**: $u_{\tau h}|_{l_n} \in P_k(l_n, V^n_h)$ s.t. for all $v_{\tau h} \in P_k(l_n, V^n_h)$ and all $n = 1 \ldots N$,

\[
\int_{l_n} (\partial_t u_{\tau h} + A^n_h u_{\tau h}, v_{\tau h})_L \, dt + ([u_{\tau h}]_{n-1}, v_{\tau h} (t_{n-1}^+) )_L = Q_n((f, v_{\tau h})_L)
\]
Example: dG(1) in time

- On each time interval $I_n$, we can solve for the two unknowns

$$U_h^n = u_{\tau h}(t_{n,j}) \in V_h^n \quad j = 1, 2$$

- The coupled $(2\times2)$-block system reads

$$\frac{3}{4} U_{hn}^1 + \frac{\tau_n}{2} A_h^n U_{hn}^1 + \frac{1}{4} U_{hn}^2 = u_{\tau h}(t_{n-1}) + \frac{\tau_n}{2} P^n_h f(t_{n,1})$$

$$- \frac{9}{4} U_{hn}^1 + \frac{5}{4} U_{hn}^2 + \frac{\tau_n}{2} A_h^n U_{hn}^2 = -u_{\tau h}(t_{n-1}) + \frac{\tau_n}{2} P^n_h f(t_{n,2})$$

where $P^n_h$ is the $L$-orthogonal projector onto $V_h^n$
Main results

- **Improved and new error estimates** for smooth solutions
  - polynomial order $k \geq 1$ in time
  - unified analysis for FEM with linear stabilization in space

- **Two main analysis tools** in time
  - post-processed, time-continuous discrete solution $\mathcal{L}_\tau u_{\tau h}$
  - special time-interpolate $R_{\tau}^{k+1} u$ of order $(k + 1)$

- $\ell^\infty(L^2)$ and $L^2(L^2)$ estimates for $(u - \mathcal{L}_\tau u_{\tau h})$
  - super-convergent bound of order $(\tau^{k+2} + h^{r+1/2})$ on static meshes
  - novel estimate on projection error for time-varying meshes

- Estimates on **error derivatives** (on static meshes)
  - bound on $(\partial_t u - \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h})$ of order $(\tau^{k+1} + h^{r+1/2})$ in $\ell^\infty(L^2)$ and in $L^2(L^2)$
  - optimal bound on the discrete graph norm of $(u - \mathcal{L}_\tau u_{\tau h})$
Comparison with RK methods (1)

- Explicit RK methods in time combined with \textbf{dG in space} (and suitable limiters) [Cockburn, Shu et al., 89-..]

- Explicit time-marching schemes are conditionally stable
  - error bounds require \textit{Gronwall’s argument}
  - error constant blows up \textit{exponentially} in $T$

- Analysis of explicit RK2 and RK3 schemes: $\ell^\infty(L^2)$ estimates
  - nonlinear conservation laws and dG in space [Zhang & Shu, 04, 10]
  - Friedrichs’ systems, stabilized FEM [Burman, AE & Fernández, 10]
  - $O(\tau^2 + h^{r+1/2})$ for RK2 under tightened CFL condition $\tau = O(h^{4/3})$
  - for RK2 with $r = 1$, usual CFL suffices ($\tau = O(h)$)
  - $O(\tau^3 + h^{r+1/2})$ for RK3 under usual CFL
  - no unified analysis available for arbitrary order in time
Comparison with RK methods (2)

- Advantages of time-dG schemes are
  - unconditional stability
  - super-convergent error estimates
  - error constants behave as $T^{1/2}$
  - unified analysis for all polynomial orders $k \geq 1$ (implicit Euler corresponding to $k = 0$ being slightly different)

- The prize to pay is increased computational cost
  - can be tamed by efficient multigrid solvers
  - heat, Stokes and NS equations [Hussain, Schieweck & Turek, 11, 12]

- Implicit RK schemes share various advantages with dG in time
  - recent analysis for linear Maxwell equations [Hochbruck & Pažur, 15]
Analysis tools

- Recall \( X^k_\tau(B) = \{ w_\tau : \bar{I} \to B : w_\tau|_{l_n} \in \mathbb{P}_k(l_n, B), \forall n = 1 \ldots N \} \)

- **Lifting operator**

\[
\mathcal{L}_\tau : X^k_\tau(B) \to X^{k+1}_\tau(B) \cap C^0(\bar{I}, B)
\]

such that \( \mathcal{L}_\tau w_\tau(0) = w_\tau(0) \) and, for all \( n = 1 \ldots N \),

\[
\mathcal{L}_\tau w_\tau(t) = w_\tau(t) - [w_\tau]_{n-1} \vartheta_n(t) \quad \forall t \in l_n = (t_{n-1}, t_n)
\]

where \( \vartheta_n \in \mathbb{P}_{k+1}(l_n, \mathbb{R}) \), \( \vartheta_n(t_{n-1}) = 1 \) and vanishes at the \((k+1)\) RS GR points, so that \( \mathcal{L}_\tau w_\tau(t_n, \mu) = w_\tau(t_n, \mu) \) for all \( \mu = 1 \ldots (k+1) \)

- The fully discrete problem can be rewritten as

\[
\int_{l_n} (\partial_t \mathcal{L}_\tau u_{\tau h} + A^n_h u_{\tau h}, v_{\tau h})_L \, dt = Q_n((f, v_{\tau h})_L)
\]
A higher-order time interpolate (1)

- Let \( u \in C^1(\bar{I}, B) \)

- **Step 1.** Choose a Lagrange/Hermite interpolate \( I_{\bar{\tau}}^{k+2}u \in C^1(\bar{I}, B) \) such that, for all \( n = 1 \ldots N \), \( I_{\bar{\tau}}^{k+2}u|_{I_n} \in \mathbb{P}_{k+2}(I_n, B) \) and

\[
I_{\bar{\tau}}^{k+2}u(t_n) = u(t_n) \quad \text{and} \quad \partial_t I_{\bar{\tau}}^{k+2}u(t_n) = \partial_t u(t_n)
\]

- for \( k = 1 \), these conditions fully determine \( I_{\bar{\tau}}^{k+2}u \) in \( I_n \)
- for \( k \geq 2 \), values at additional Lagrange nodes in \( I_n \) are prescribed
- for \( k = 0 \), this construction is not possible

- **Step 2.** Define \( R_{\bar{\tau}}^{k+1}u|_{I_n} \in \mathbb{P}_{k+1}(I_n, B) \) by the \((k + 2)\) conditions

\[
\partial_t R_{\bar{\tau}}^{k+1}u(t_n, \mu) = \partial_t I_{\bar{\tau}}^{k+2}u(t_n, \mu) \quad \forall \mu = 1 \ldots (k + 1)
\]

\[
R_{\bar{\tau}}^{k+1}u(t_{n-1}^+) = I_{\bar{\tau}}^{k+2}u(t_{n-1})
\]

and set \( R_{\bar{\tau}}^{k+1}u(0) = u(0) \)
A higher-order time interpolate (2)

- **Continuity**: \( R_{\tau}^{k+1} u \in C^0(\bar{I}, B) \) and \( R_{\tau}^{k+1} u(t_n) = u(t_n) \) for all \( n = 0 \ldots N \)

- **Approximation of smooth functions**

\[
\| u - R_{\tau}^{k+1} u \|_{C^0(\bar{I}_n, B)} \lesssim \tau_n^{k+2} \| u \|_{C^{k+2}(\bar{I}_n, B)}
\]
\[
\| \partial_t u - \partial_t R_{\tau}^{k+1} u \|_{C^0(\bar{I}_n, B)} \lesssim \tau_n^{k+1} \| u \|_{C^{k+2}(\bar{I}_n, B)}
\]

- **Stability**: \( \| R_{\tau}^{k+1} u \|_{C^1(\bar{I}_n, B)} \lesssim \| u \|_{C^1(\bar{I}_n, B)} \) for all \( u \in C^1(\bar{I}_n, B) \)
\( \ell^\infty(L^2) \) and \( L^2(L^2) \) error estimates

- **Static meshes**

- **Post-processed error** \( \tilde{e} = u - \mathcal{L}_\tau u_{\tau h} \): For all \( m = 1 \ldots N \),

\[
\| \tilde{e}(t_m) \|_L^2 \lesssim (E_0)^2 + t_m \max_{1 \leq n \leq m} \left\{ C^T_n(u) \tau_n^{2(k+2)} + C^S_n(u) h^{2r+1} \right\} + \text{hot}
\]

and under the mild assumption \( \tau_n \lesssim \tau_{n-1} \),

\[
\| \tilde{e} \|_{L^2(I,L)}^2 \lesssim (E_0)^2 + T \max_{1 \leq n \leq N} \left\{ C^T_n(u) \tau_n^{2(k+2)} + C^S_n(u) h^{2r+1} \right\}
\]

- For the error \( u - u_{\tau h} \), same super-convergent bound in \( \ell^\infty(L^2) \), but only optimal \( (\tau^{k+1} + h^{r+1/2}) \) bound in \( L^2(L^2) \)

- Super-convergence does not hold for implicit Euler \( (k = 0) \)
Time-varying meshes

- Time-varying meshes lead to an additional projection error

- Assume that $\mathcal{T}_h^n$ is created from $\mathcal{T}_h^{n-1}$ by local refinements and coarsenings (using a common finest mesh)

- The local (in time) projection error is defined as

$$E_n^P(u) = \sup_{v_h \in V_h^n} \frac{(u(t_{n-1}) - P_h^{n-1}u(t_{n-1}), v_h - \Pi_h^{n-1}v_h)_L}{\|v_h - \Pi_h^{n-1}v_h\|_L}$$

- $\Pi_h^{n-1}: V_h^{n-1} + V_h^n \rightarrow V_h^{n-1}$ denotes an $L^2$-stable, linear quasi-interpolation operator

- Lagrange interpolate for $H^1$-conf. FEM, $L^2$-projection for dG

- local projection error vanishes if there is only mesh coarsening

- The global projection error entering the $\ell^\infty(L^2)$ and $L^2(L^2)$ error estimates is $$(E_{P,m}(u))^2 = \sum_{n=1}^m (E_n^P(u))^2$$
Bound on projection error

- Decompose mesh as $\mathcal{T}_h^n = \mathcal{T}_h^{n,\text{ref}} \cup \mathcal{T}_h^{n,\text{coa}}$ where $\mathcal{T}_h^{n,\text{coa}}$ collects mesh cells in $\mathcal{T}_h^n$ that can be decomposed into one or more cells of $\mathcal{T}_h^{n-1}$

- Quasi-interpolation operator satisfies $(v_h - \Pi_{h}^{n-1} v_h)|_K = 0$, $\forall v_h \in V_h^n$, $\forall K \in \mathcal{T}_h^{n,\text{coa}}$

- On dG spaces, the local projection error is bounded as

$$E_n^P(u) \lesssim |\Omega_n^{\text{ref}}|^{1/2} (h_n^{\text{ref}})^{1/2} \left\{ (h_n^{\text{ref}})^{r+1/2} |u(t_{n-1})|_{W^{r+1,\infty}(\Omega_n^{\text{ref}})} \right\}$$

and on $H^1$-conforming spaces, it is bounded as

$$E_n^P(u) \lesssim (h_n^{\text{ref}})^{1/2} \left\{ (h_n)^{r+1/2} |u(t_{n-1})|_{H^{r+1}(\Omega)} \right\}$$

- The bound on dG spaces can exploit that, often in practice, $|\Omega_n^{\text{ref}}| \lesssim h_n^{\text{ref}}$ (up to a slightly stronger regularity on $u$)
Estimates on error derivatives

- Bounds on error derivatives are rarely explored in the literature
- Assume static meshes
- General methodology
  - derive super-convergent (in time) $\ell^\infty(L^2)$ and $L^2(L^2)$ error bounds on
    time-derivative
  - infer optimal (in time) discrete graph-norm error estimate using
    discrete inf-sup stability
Estimate on time derivative

- Key idea: error on time-derivative is defined as

\[ \hat{e} = \partial_t u - L_{\tau} \partial_t L_{\tau} u_{\tau} h \]

- For all \( m = 1 \ldots N, \)

\[ \| \hat{e}(t_m) \|^2_L \lesssim (\hat{E}_0)^2 + t_m \max_{1 \leq n \leq m} \left\{ \hat{C}_n^T(u, f)_{\tau_n}^{2(k+1)} + C_n^S(u)h^{2r+1} \right\} + \text{hot} \]

and under the mild assumption \( \tau_n \lesssim \tau_{n-1}, \)

\[ \| \hat{e} \|^2_{L^2(I, L)} \lesssim (\hat{E}_0)^2 + T \max_{1 \leq n \leq N} \left\{ \hat{C}_n^T(u, f)_{\tau_n}^{2(k+1)} + C_n^S(u)h^{2r+1} \right\} \]
Discrete graph norm error estimate

- Recall discrete inf-sup stability with stability norm

\[ \| v_h \|_h^2 = \| v_h \|_h^2 + \sum_{T \in T_h} h_T \| \beta \cdot \nabla v_h \|_{L^2(T)} \]

- $\ell^2(V)$-estimate on $\tilde{e} = u - \mathcal{L}_\tau u_{\tau} h$: For all $m = 1 \ldots N$,

\[ \sum_{n=1}^{m} Q_n(\tilde{\| e \|}_h^2) \lesssim (\hat{E}_0)^2 + t_m \max_{1 \leq n \leq m} \left\{ \tilde{C}_n^T(u, f) \tau_n^{2(k+1)} + C_n^S(u) h^{2r+1} \right\} \]

- This bound is optimal in time and exhibits the usual (quasi-)optimal behavior in space for steady problems.
Weighting linear stabilization

- Motivations
- Weighting LS: theory
- Weighting LS: numerics
- We focus on Continuous Interior Penalty, but conjecture most conclusions extend to other LS
Motivations

- LS adds least-squares penalty to standard Galerkin FEM
  - acts as a high-order dissipation (in contrast to first-order viscosity)
  - LS is very effective for linear first-order PDEs with smooth data

- The situation is not so bright when it comes to solving
  - linear problems with non-smooth data
  - nonlinear problems with non-unique weak solutions

- LS promotes the Gibbs phenomenon, leading to
  - spurious oscillations in the vicinity of shocks
  - failure to satisfy a maximum principle
  - convergence to non-entropic weak solutions
Nonlinear viscosity

- LS is often supplemented by some nonlinear viscosity technique
  - shock-capturing [Hughes & Mallet, 86; Johnson & Szepessy, 87]
  - crosswind diffusion [Codina, 93; Burman & AE, 02; Burman, 07]
  - entropy viscosity [Guermond, 08; G. & Pasquetti, 08; G., Pasquetti & Popov, 11]

- It is not clear that LS and nonlinear viscosity work hand in hand

- Numerical tests indicate they can antagonize each other
Some illustrations

- **Nonlinear conservation law**

\[
\begin{aligned}
\left\{ \begin{array}{l}
\partial_t u + \nabla \cdot f(u) = 0 \\
\quad (x, t) \in \Omega \times (0, T) \\
\quad u|_{t=0} = u_0 \\
\quad x \in \Omega
\end{array} \right.
\]

- \(\Omega\) open polyhedral domain in \(\mathbb{R}^d\); \(f \in C^1(\mathbb{R}; \mathbb{R}^d)\)
- *no issues with BCs* (either periodic or compactly supported \(u_0\))
- we assume that (1) admits a **unique weak entropic solution**
- we consider space semi-discretization

- **Galerkin solution** \(u_h \in C^1([0, T]; V_h)\) s.t. \(u_h|_{t=0} = u_{0,h}\) and

\[
\int_{\Omega} w_h \partial_t u_h \, d\Omega + \int_{\Omega} w_h \nabla \cdot f(u_h) \, d\Omega = 0 \quad \forall w_h \in V_h \quad \forall t \in (0, T)
\]

with \(H^1\)-conforming FE space \(V_h\) (of order \(r\))

... globally polluted by spurious oscillations
Viscous solution

\[ \int_{\Omega} w_h \partial_t u_h \, d\Omega + \int_{\Omega} w_h \nabla \cdot f(u_h) \, d\Omega + n_{\text{visc}}(u_h; w_h) = 0 \]

with

\[ n_{\text{visc}}(v_h; w_h) = c_{\text{max}} \sum_{T \in \mathcal{T}_h} h_T \| f'(v_h) \|_{L^\infty(T)} \int_T \nabla v_h \cdot \nabla w_h \, dT \]

- typically \( c_{\text{max}} = \frac{1}{2r} \) in 1D and \( c_{\text{max}} = \frac{1}{4r} \) in 2D
- for linear transport, \( f(v_h) = \beta v_h \) so that \( \| f'(v_h) \|_{L^\infty(T)} = \| \beta \|_{L^\infty(T)} \)

... only first-order accurate
CIP stabilized solution

\[
\int_{\Omega} w_h \partial_t u_h \, d\Omega + \int_{\Omega} w_h \nabla \cdot f(u_h) \, d\Omega + n_{\text{CIP}}(u_h; w_h) = 0
\]

with

\[
n_{\text{CIP}}(v_h; w_h) = c_{\text{CIP}} \sum_{F \in \mathcal{F}_h^i} h_F^2 \| f'(v_h) \|_{L^\infty(F)} \int_F [\nabla v_h] : [\nabla w_h] \, dF
\]

- typically, \( c_{\text{CIP}} = 0.05 \)

... \( O(h^{r+1/2}) \) \( L^2 \)-estimates for linear transport and smooth solutions
Entropy-viscosity solution

- Entropy-viscosity solution (nonlinear stabilization)

\[ \int_{\Omega} w_h \partial_t u_h \, d\Omega + \int_{\Omega} w_h \nabla \cdot f(u_h) \, d\Omega + n_{\text{entr}}(u_h; u_h, w_h) = 0 \]

with

\[ n_{\text{entr}}(z_h; v_h, w_h) = \sum_{T \in T_h} \nu_T(z_h) \int_T \nabla v_h \cdot \nabla w_h \, dT \]

and \( \nu_T(z_h) \) is designed s.t.

\[ \nu_T(z_h) = \min(c_{\text{max}} \beta_T(z_h) h_T, c_{\text{ev}} D_T(z_h) h_T^2) \]

and \( \beta_T(z_h) = \|f'(z_h)\|_{L^\infty(T)} \), \( D_T(z_h) \) is the local residual for a chosen entropy (e.g., the quadratic one)

... weak maximum principle (proof in 1D)

\[ \|u_h(t)\|_{L^\infty(\Omega)} \leq \|u_0\|_{L^\infty(\Omega)} + ch^\alpha \]
Illustration of difficulties

- Numerical tests in 1D
  - linear transport with non-smooth data
  - nonlinear transport with composite wave (non-convex flux)
  - CIP stabilization and first-order viscosity

- Time discretization is performed using SSP RK3
  - (very) small time steps to avoid time discretization errors

- The mass matrix is never lumped
Linear transport with non-smooth data I

- $\partial_t u + \partial_x u = 0$, $u(x, 0) = 1_{(0.4, 0.7)}$, periodic BCs, and $T = 1$

- Stabilizing capability of CIP stabilization, but inability to counter Gibbs phenomenon

- Maximum principle indicators at final time

$$e_{\text{Max}} = \max_{x \in \Omega} u_h(x, T) - 1 \quad e_{\text{Min}} = -\min_{x \in \Omega} u_h(x, T)$$

remain bounded away from zero for CIP
Linear transport with non-smooth data II

<table>
<thead>
<tr>
<th>$h$</th>
<th>entropy $e_{\text{Max}}$</th>
<th>rate</th>
<th>entropy + CIP $e_{\text{Max}}$</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.500E-03</td>
<td>6.715E-03</td>
<td>–</td>
<td>1.597E-02</td>
<td>–</td>
</tr>
<tr>
<td>1.250E-03</td>
<td>5.434E-03</td>
<td>0.305</td>
<td>1.600E-02</td>
<td>-0.003</td>
</tr>
<tr>
<td>6.250E-04</td>
<td>2.854E-03</td>
<td>0.929</td>
<td>1.633E-02</td>
<td>-0.030</td>
</tr>
<tr>
<td>3.125E-04</td>
<td>2.235E-03</td>
<td>0.353</td>
<td>1.626E-02</td>
<td>0.006</td>
</tr>
<tr>
<td>1.563E-04</td>
<td>1.785E-03</td>
<td>0.324</td>
<td>1.646E-02</td>
<td>-0.017</td>
</tr>
</tbody>
</table>

- Entropy-viscosity solution satisfies a **weak maximum principle**
- Adding CIP to entropy-viscosity, the **WMP is lost!**
Nonlinear transport with composite wave I

- Riemann problem with non-convex flux (S-shaped)

\[ f(u) = \begin{cases} 
\frac{1}{4} u(1 - u) & \text{if } u < \frac{1}{2} \\
\frac{1}{2} u(u - 1) + \frac{3}{16} & \text{if } \frac{1}{2} \leq u 
\end{cases} \]

\[ u_0(x) = \begin{cases} 
0 & x \in [0, 0.35] \\
1 & x \in (0.35, 1] 
\end{cases} \]

- Entropy solution at \( T = 1 \) is composed of a shock wave followed by a rarefaction wave

- Many second-order central schemes with limiters converge to a non-entropic (weak) solution
  - e.g., central upwind with second-order reconstruction and either superbee or minmod2 limiters [Kurganov, Petrova & Popov, 07]
Nonlinear transport with composite wave II

- Uniform mesh with 1000 cells, SSP RK3 with CFL = 0.01

Galerkin

CIP

- The CIP-stabilized solution converges to a non-entropic solution
Nonlinear transport with composite wave III

- Entropy-viscosity solution **converges to (correct) entropic solution**
- Adding CIP **destroys this property**!
CIP stabilization and first-order viscosity

- CIP can have adverse effects even on first-order viscosity
- **(Inviscid) Burgers equation** with $u(x, 0) = \sin(2\pi x)$, 200 mesh cells, $r = 1$, CFL = 0.025
  - adding CIP to 1st-order visc. leads to over/under-shoots
  - $c_{\text{max}} = 2$ makes 1st-order visc. overcome Gibbs phenomenon triggered by CIP
- **Riemann problem with non-convex flux**, 4,000 and 10,000 cells
  - viscous solution converges to entropic solution (as expected!)
  - adding CIP stabilization destroys this property
Conclusions from numerical tests

- CIP does a great job at suppressing oscillations in smooth regions
- It promotes the Gibbs phenomenon
  - failure to satisfy a (weak) maximum principle
  - convergence to non-entropic weak solutions
- These effects can even overcome convergent viscosity methods (both nonlinear and first-order)
Key idea

- Temper the amount of LS in the vicinity of shocks
  - nonlinear weights depending on the local gradient of discrete solution
  - may seem counter-intuitive at first glance since LS is often motivated to counter spurious oscillations near large gradients ...

- We show that CIP stabilization can be tempered in such a way that
  - $O(h^{r+1/2})$ $L^2$-error estimates are preserved for smooth solutions in linear problems [proof]
  - LS no longer antagonizes nonlinear viscosity methods [numerics]

- This is a win-win situation
  - nonlinear viscosity alone does not deliver full-order accuracy in smooth regions
Theoretical insight

- Weighted CIP-stabilized solution

\[
\int_{\Omega} w_h \partial_t u_h \, d\Omega + \int_{\Omega} w_h \nabla \cdot f(u_h) \, d\Omega + n_{\text{wei,ed}}(u_h; u_h, w_h) = 0
\]

with

\[
n_{\text{wei,ed}}(z_h; v_h, w_h) = c_{\text{CIP}} \sum_{F \in F_h^i} \alpha(g_F(z_h)) h_F^2 \|f'(v_h)\|_{L^\infty} \int_F [\nabla v_h] \cdot [\nabla w_h] \, dF
\]

where \( g_F(z_h) = |\langle \nabla z_h \rangle_{\Delta_F}|/\ell(u_0) \) is a local measure of \( \nabla z_h \) around \( F \)

- The weighting function \( \alpha \) is non-increasing and

\[
\exists \lambda > 0, \quad (r \geq r_0) \Rightarrow (\alpha(r) \geq r^{-\lambda})
\]

\( \alpha \) cannot decrease too fast (typically \( \alpha(0) = 1 \) and \( \alpha(\infty) = 0 \))
Convergence analysis

- **Linear transport, smooth solutions**
- For all $t \in [0, T]$, with $e = u - u_h$,

\[
\|e(t)\|_{L^2(\Omega)}^2 + \int_0^t n_{\text{wei,ed}}(u_h; e, e) \, d\tau \lesssim h^{2r+1}
\]

with
- for all $\lambda > 0$, if $d = 2$ or if $d = 3$ and $r \geq 3$
- for $d = 3$ and $r \in \{1, 2\}$, upper bound is $h^{r+\epsilon \lambda}$ with $\epsilon \lambda \in (0, \frac{1}{2})$ and $\lambda \in (0, 2)$ for $r = 2$ and $\lambda \in (0, \frac{2}{3})$ if $r = 1$

- Proof on quasi-uniform meshes
Principle of proof I

▶ Classical techniques lead to

\[
\frac{d}{dt} \|e\|^2_{L^2(\Omega)} + n_{\text{wei,ed}}(u_h; e, e) \leq \text{RHS}(\Omega) \lesssim h^r \|e\|_{L^2(\Omega)}
\]

where control on \( n_{\text{wei,ed}}(u_h; e, e) \) is not yet used

▶ Let \( \epsilon \geq 0 \) and consider the sets collecting “bad” and “good” cells

\[
\Omega^\# = \{g_F(u_h) \geq h^{-\epsilon}\}
\]

\[
\Omega^b = \Omega \setminus \Omega^\#
\]

\( \Omega^\# \) collects mesh cells where the gradient of \( u_h \) is locally high
Principle of proof II

- On $\Omega^b$, owing to the behavior of weighting function $\alpha$, there is enough CIP stabilization to infer that

$$\text{RHS}(\Omega^b) \lesssim h^{r+\frac{1}{2}-\frac{1}{2}\lambda\epsilon} n_{\text{wei,ed}}(u_h; e, e)^{\frac{1}{2}}$$

- On $\Omega^\#$, the following holds:

$$\text{RHS}(\Omega^\#) \lesssim h^{2r} |\Omega^\#|^{\frac{1}{2}} \quad \text{and} \quad |\Omega^\#| \lesssim h^{2(r-1+\epsilon)}$$

since $\|\nabla u\|_{L^2(\Omega^\#)}$ and $\|\nabla e\|_{L^2(\Omega^\#)}$ are bounded

- This yields

$$\frac{d}{dt} \|e\|_{L^2(\Omega)}^2 + n_{\text{wei,ed}}(u_h; e, e) \lesssim h^{3r-1+\epsilon} + h^{2r+1-\lambda\epsilon}$$

Choose $\epsilon$ to equilibrate both terms and derive an improved error estimate $O(h^{r+\rho})$, and then use a bootstrap argument.
Numerical examples

- We study the effectiveness of the **weighted CIP-stabilization** on
  - linear transport with smooth data
  - linear transport with non-smooth data
  - nonlinear transport with composite wave

- 1D and 2D tests are considered
1D tests I

- $\Omega = (0, 1)$ with periodic BCs, $r = 1$, SSP RK3 with CFL = 0.2
  - stab. parameters $c_{\text{CIP}} = 0.05$, $c_{\text{max}} = 0.5$, and $c_{\text{ev}} = 0.5$

- Linear transport with smooth data, CIP stabilization with and without weighting
  \[ \| e \|_{L^1(\Omega)} \sim h^2 \quad \| e \|_{L^2(\Omega)} \sim h^2 \]

- Linear transport with non-smooth data, entropy viscosity plus CIP stabilization, uniform and non-uniform meshes
  \[ \| e \|_{L^1(\Omega)} \sim h^{0.75} \quad \| e \|_{L^2(\Omega)} \sim h^{0.37} \]

and weak maximum principle is satisfied (with rate $h^{0.5}$)
1D tests II

- **Riemann problem with non-convex flux**
  - five uniform meshes from 100 up to 1,600 cells
  - entropy viscosity plus CIP stabilization

- Convergence to the (correct) entropic solution
2D tests I

- **Linear transport** (rotating velocity field in unit disk)
  - $r \in \{1, 2\}$, RK4, CFL $= 0.25$
  - stab. parameters $c_{\text{CIP}} = 0.025$, $c_{\text{max}} = \frac{1}{4r}$, and $c_{\text{ev}} = 0.1$

- CIP stabilization with and without weighting leads to optimal convergence on smooth solutions

- Entropy viscosity plus weighted CIP stabilization
  - $r = 1$: entropy viscosity alone and with CIP is second-order
  - $r = 2$: entropy viscosity alone is $h^{2+\epsilon}$, while adding CIP improves CV at least to $h^{2.5} \rightarrow \text{win-win situation}$
2D tests II

- Linear transport, **non-smooth data**, entropy visc. + CIP, $r \in \{1, 2\}$

- CV rates (in $L^1$-norm, rates are $h^{0.75}$ for $r = 1$ and $h^{0.8}$ for $r = 2$)

<table>
<thead>
<tr>
<th>$h$</th>
<th>$r = 1$</th>
<th></th>
<th>$r = 2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$L^2$-norm</td>
<td>rate</td>
<td>$L^2$-norm</td>
</tr>
<tr>
<td>5.00E-02</td>
<td>4.172E-01</td>
<td>–</td>
<td>2.794E-01</td>
<td>–</td>
</tr>
<tr>
<td>2.50E-02</td>
<td>3.158E-01</td>
<td>0.402</td>
<td>2.114E-01</td>
<td>0.402</td>
</tr>
<tr>
<td>1.25E-02</td>
<td>2.411E-01</td>
<td>0.389</td>
<td>1.601E-01</td>
<td>0.401</td>
</tr>
<tr>
<td>1.00E-02</td>
<td>2.214E-01</td>
<td>0.383</td>
<td>1.466E-01</td>
<td>0.394</td>
</tr>
</tbody>
</table>

- Weak maximum principle for $e_{\text{Max}}$ (similar results for $e_{\text{Min}}$)

<table>
<thead>
<tr>
<th>$h$</th>
<th>$r = 1$</th>
<th></th>
<th>$r = 2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$e_{\text{Max}}$</td>
<td>rate</td>
<td>$e_{\text{Max}}$</td>
</tr>
<tr>
<td>5.00E-02</td>
<td>3.546E-02</td>
<td>–</td>
<td>7.904E-03</td>
<td>–</td>
</tr>
<tr>
<td>2.50E-02</td>
<td>1.283E-02</td>
<td>1.467</td>
<td>6.943E-03</td>
<td>0.187</td>
</tr>
<tr>
<td>1.25E-02</td>
<td>7.776E-02</td>
<td>0.722</td>
<td>5.953E-03</td>
<td>0.222</td>
</tr>
<tr>
<td>1.00E-02</td>
<td>6.798E-02</td>
<td>0.603</td>
<td>5.211E-03</td>
<td>0.596</td>
</tr>
</tbody>
</table>
2D tests III

- **Cauchy problem in** $\mathbb{R}^2$ **with non-convex flux**

$$f(u) = (\sin u, \cos u) \quad u(x, y, 0) = \begin{cases} 
3.5\pi & x^2 + y^2 < 1 \\
0.25\pi & \text{otherwise}
\end{cases}$$

- entropy viscosity ($c_{\text{max}} = \frac{1}{2}, c_{\text{ev}} = 1$) predicts correct rotating composite wave structure
- adding CIP ($c_{\text{CIP}} = 1$) leads to non-physical layers
- weighting CIP pushes spurious layer back to the shock
Conclusions

- In the literature, much efforts are devoted to constructing LS techniques in various flavors.
- It is often believed that LS is the workhorse, whereas shock-capturing is only meant to remove remaining oscillations.
- We believe that
  - nonlinear viscosities should be the workhorses killing the Gibbs phenomenon and ensuring convergence to the entropic solution.
  - LS plays the role of an auxiliary tool whose job is to improve convergence in smooth regions.