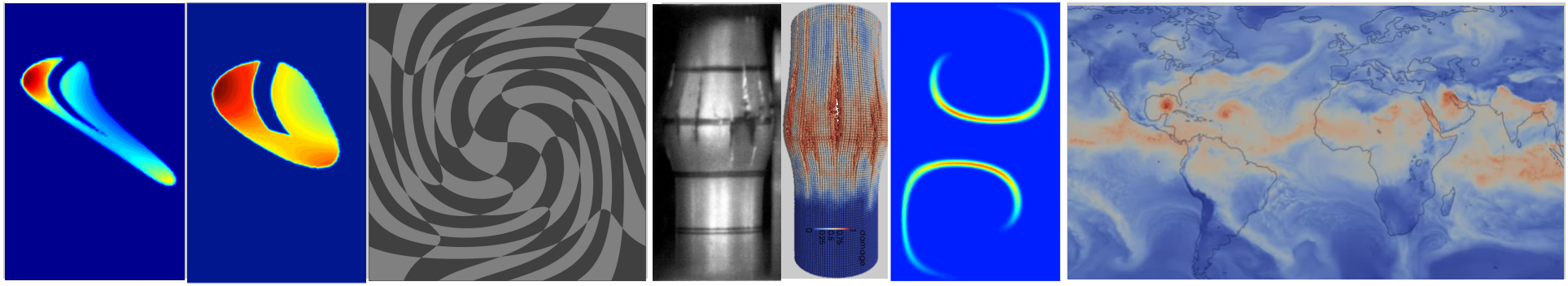
SAND2016-0022C

Exceptional service in the national interest

Sandia National Laboratories



Optimization-Based Property Preserving Methods, or

Going Boldly Beyond Compatible Discretizations

CCR Center for Computing Research

Pavel Bochev
Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY | Office of Science

Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4)

WONAPDE

Fifth Chilean Workshop on Numerical Analysis of PDEs

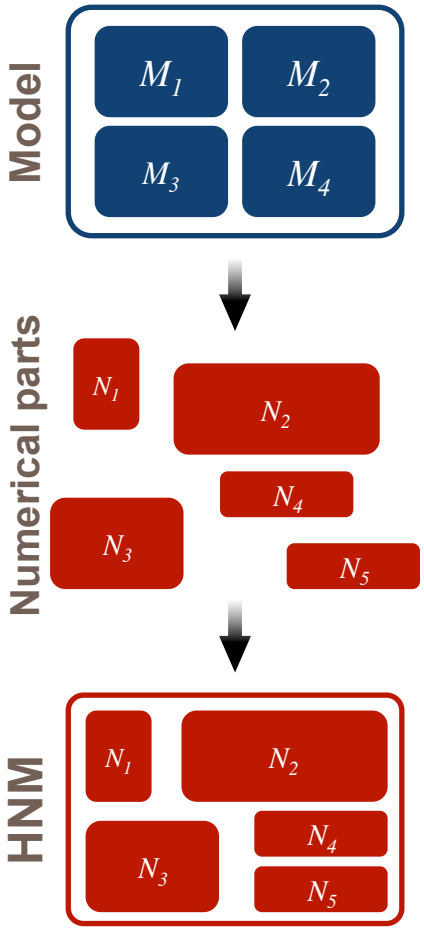
University of Concepcion, Chile, January 11-15.

U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration | ASC

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Research drivers & objectives

DOE uses **computer models** to understand, predict, and verify **complex systems** in **high consequences analyses** that would be difficult or even impossible by other means.



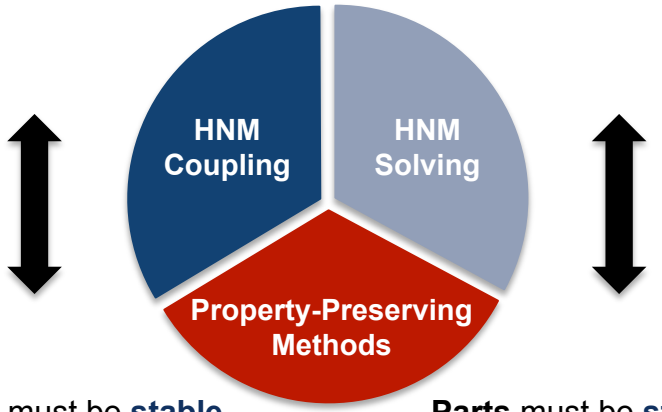
Complex systems require **diverse "mathematical parts"**: PDEs, integral equations, classical DFT, potential-based atomistic…

Diverse math models require **diverse "numerical parts":** mesh based (FE, FV, FD), meshless (SPH, MLS), implicit, explicit, Eulerian, Lagrangian…

HNM = Collection of **dissimilar numerical parts** from multiple disciplines functioning together as a **unified simulation tool**

Challenges: beyond compatible methods

The parts must function **together** as a **unified simulation tool (HNM).** ⟷ We must be able to **solve our HNMs efficiently**



HNMs must be **stable, accurate** and **preserve key physical properties** ⟷ **Parts** must be **stable, accurate** and **preserve key physical properties.**

We carry out a comprehensive research effort to address these challenges

Taxonomy of challenges

Sandia National Laboratories

1. Achieving Stability & Accuracy (Structural aspects)

- **Game changer**: **Homological techniques**: FE exterior calculus (DEC), mimetic FD,…
- Typically achieved by *topological means:*
  - Careful placement of the variables on the mesh.
  - Special grid structure, e.g., topologically dual grids.
- **Challenges**:
  1. **Models that don't fit EC structure**, e.g., heterogeneous methods: FEM+cDFT
  2. **Stable and accurate** does not imply **property preserving**…

2. Preserving Physical Properties (Qualitative aspects)

- Maximum principles, local bounds, symmetries, Geometric Conservation Laws,…
- Correlations between variables, e.g., between two passive tracers.
- **Challenges**: conventional ways to preserve these properties are either
  - **Restrictive**: Cartesian mesh, angle conditions, etc, and/or,
  - **Entangle accuracy** with the property preservation, e.g., limiters.
- **Game changer**?

Taxonomy of challenges

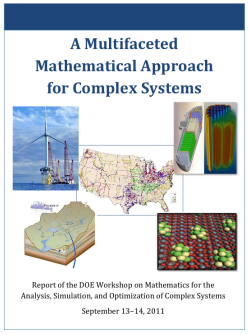
Sandia National Laboratories

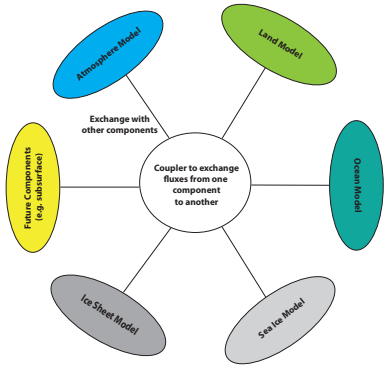
3. Assembling Diverse Numerical Parts into HNMs and solving them

☞"Exascale computing will enable consideration of new classes of multiscale problems in which different types of discretizations, appropriate to a particular scale in different portions of the domain, are employed and models which treat distinct phenomena in different parts of the domain, such as ocean-atmosphere coupling…"

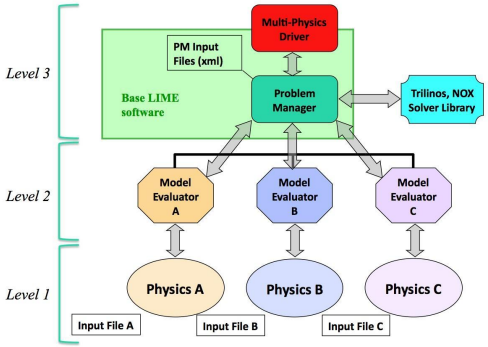
☞"Effective models must be hierarchical and include multiple sub-models that represent different phenomena with vastly differing scales."

"As this type of simulation expands, there is a critical need to develop systematic approaches for coupling across the range of scales and quantification of the properties of these types of coupling strategies"





Global Earth System Model



LIME: Lightweight Integrating Multiphysics Environment (CASL)

Traditional monolithic and operator-splitting modeling approaches fall short of meeting the crosscutting challenges; see *Multifaceted Mathematical Approach for Complex Systems.*

Game changer?

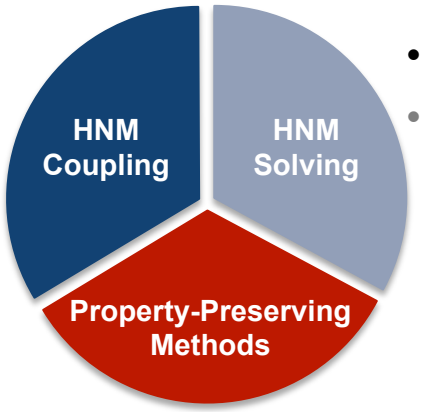
Game changer?

Sandia National Laboratories

The use of optimization ideas to couple heterogeneous numerical methods and to preserve the relevant physical properties could be a game changer.

Optimization-based operator coupling

- Local-to-Nonlocal couplings (D'Elia talk, Tuesday)
- Atomistic-to-Continuum
- Interface problems: Friday, 11:20, AUD1.



Optimization-based operator splitting

- Abstract decomposition theory
- Application to the Navier-Stokes eqs.
- Application to advection-diffusion equations

Optimization-based property-preserving methods

- Semi-Lagrangian transport of passive tracers
- Volume correction (Geometric Conservation Law)
- Property-preserving data transfer (remap)

Thanks to:

- M. D'Elia, P. Kuberry, D. Littlewood, M. Perego, K. Peterson, D. Ridzal (SNL), M. Shashkov (LANL)
- M. Gunzburger (FSU), A. Shapeev (SkolTech), S.Moe (U. WA), M. Luskin, D. Olson (U. MN)

The optimization approach in a nutshell

Sandia National Laboratories

Couch assembly of numerical parts and preservation of properties into an optimization problem:

Minimize

Coupling mismatch $\left\| u_1^h - u_2^h \right\|^2$ + $\sum_{i=1,2} \left\| u_i^h - u_i^T \right\|^2$ **Target mismatch**

Subject to

Component physics $L_1^h(u_1^T, \theta) = f_1$ $\underline{C}_i \le C u_i^h \le \overline{C}_i$ **Physical properties**

$L_2^h(u_2^T, \theta) = f_2$ $B u_i^h = b$

- **Reverses the roles** of the coupling conditions and the models.
- **Divide and conquer approach:**
  - **separates numerical parts: facilitates merging of heterogeneous methods.**
  - **separates *accuracy* from *physical properties*** (local bounds, conservation, etc..)

Sandia National Laboratories

Part 1

Optimization-based operator splitting

Sandia National Laboratories

In Part 1 we consider application of optimization to operator splitting.

Minimize

Coupling mismatch $\left\| u_1^h - u_2^h \right\|^2$ + $\sum_i \left\| u_i^h - u_i^T \right\|^2$ Target mismatch

Subject to

Component physics $L_1^h(u_1^h, \theta) = f_1$

$L_2^h(u_2^h, \theta) = f_2$

$\underline{C}_i \le C u_i^h \le \overline{C}_i$ Physical properties

$B u_i^h = b$

This case study highlights the use of optimization ideas for the design of robust and efficient solvers for multiphysics problems.

Related work: Lions (2001), Quarteroni (2000), Gunzburger (2000), Du (2001) – applications to PDEs, Oden (2011 – Atomistic to Continuum), Discacciati (2013 – heterogeneous domain decomposition), Karniadakis (2014-Stochastic PDE)

Abstract additive operator-splitting theory

Sandia National Laboratories

Model problem

U, V, H - Hilbert spaces, V^* - dual of V, s.t. $\{V, H, V^*\}$ is Gelfand triple

$Q(\cdot,\cdot): U \times V \to \mathbf{R}$ **weak form of a "multiphysics" operator**

Seek $u \in U$ such that $Q(u,v) = \langle f, v \rangle \quad \forall v \in V, \quad f \in V^*$

Assumptions

$$\begin{cases} \sup_{v \in V} \dfrac{Q(u,v)}{\|v\|_V} \geq \underline{\gamma} \|u\|_U \quad \text{and} \quad \sup_{u \in U} \dfrac{Q(u,v)}{\|u\|_U} > 0 \\ Q(u,v) \leq \overline{\gamma} \|u\|_U \qquad \forall u \in U, \forall v \in V \end{cases} \quad \Rightarrow \quad \|u\|_U \leq \frac{1}{\underline{\gamma}} \|f\|_{V^*}$$

Sufficient for a well-posed variational formulation

P. Bochev and D. Ridzal, *Optimization-based additive operator decomposition of weakly coercive problems with applications,* CAMWA, 2016.

Abstract additive operator-splitting theory

Sandia National Laboratories

Assumptions

- $Q(u,v) = Q_1(u,v) + Q_2(u,v)$ with weakly coercive component forms:

$$\begin{cases} \sup_{v \in V} \dfrac{Q_i(u,v)}{\|v\|_V} \geq \underline{\gamma}_i \|u\|_U \quad \text{and} \quad \sup_{u \in U} \dfrac{Q_i(u,v)}{\|u\|_U} > 0 \\ Q_i(u,v) \leq \overline{\gamma}_i \|u\|_U \qquad \forall u \in U, \forall v \in V \end{cases} \Rightarrow \quad Q_i(u,v) = \langle f, v \rangle \quad \text{is well-posed}$$

- **Component problems** are **easier to solve** that the **monolithic problem**

Reformulation of $Q(u,v) = \langle f, v \rangle$ as a constrained optimization problem

$$\begin{cases} \text{minimize} \quad J(u_1,u_2) = \dfrac{1}{2}\|u_1 - u_2\|_U^2 \\ \text{subject to} \begin{cases} Q_1(u_1,v_1) - (\theta,v_1)_V = \langle f, v_1 \rangle & \forall v_1 \in V \\ Q_2(u_2,v_2) + (\theta,v_2)_V = 0 & \forall v_2 \in V \end{cases} \end{cases}$$

- u_1, u_2 – the states
- θ – virtual (distributed) control

Optimization **exposes** the **constituent components** of the **multiphysics** operator

Abstract additive operator-splitting theory

Sandia National Laboratories

Lagrange multiplier solution: saddle-point optimality system

$$\begin{cases} \left(u_1-u_2,\hat{u}_1-\hat{u}_2\right)_U+Q_1\left(\hat{u}_1,\lambda_1\right)+Q_2\left(\hat{u}_2,\lambda_2\right)=0 & \forall \hat{u}_1,\hat{u}_2\in U \\ \left(\hat{\theta},\lambda_2-\lambda_1\right)_V=0 & \forall \hat{\theta}\in V \\ Q_1\left(u_1,\hat{\lambda}_1\right)+Q_2\left(u_2,\hat{\lambda}_2\right)+\left(\theta,\hat{\lambda}_2-\hat{\lambda}_1\right)_V=\left\langle f,\hat{\lambda}_1\right\rangle & \forall \hat{\lambda}_1,\hat{\lambda}_2\in V \end{cases}$$

Theorem *(Bochev & Ridzal, 2015)*

The optimality system is well-posed problem with a unique solution $(u_1,u_2,\theta,\lambda_1,\lambda_2)$.

Moreover, if u is a solution of the original variational equation, then $u=u_1=u_2$

Notable facts

- Control penalty is **not required** for well-posedness of the optimality system!
- As a result, **original** and **reformulated** problems are **completely equivalent**

There's no splitting error!

Abstract additive operator-splitting theory

Sandia National Laboratories

Discretization

Assume $U^h \subset U,\ V^h \subset V$ is a pair of LBB-stable spaces for the form Q:

$$\sup_{v^h \in V^h} \frac{Q\left(u^h, v^h\right)}{\left\|v^h\right\|_V} \geq \underline{\gamma} \left\|u^h\right\|_U \quad \text{and} \quad \sup_{u^h \in U^h} \frac{Q\left(u^h, v^h\right)}{\left\|u^h\right\|_U} > 0$$

This turns out to be sufficient for the well-posedness of the discrete reformulated problem

Theorem

(Bochev & Ridzal, 2015)

- The KKT optimality system is well-posed with a unique solution $(u_1^h, u_2^h, \theta^h, \lambda_1^h, \lambda_2^h)$
- Discrete reformulated and monolithic problems are equivalent: $u^h = u_1^h = u_2^h$
- The following quasi-optimal error estimate holds:

$$\sum_{i=1,2} \left\|u_i^h - u_i\right\|_U + \sum_{i=1,2} \left\|\lambda_i^h - \lambda_i\right\|_V \leq C \left(\inf_{U^h} \sum_{i=1,2} \left\|v_i^h - u_i\right\|_U + \inf_{V^h} \sum_{i=1,2} \left\|\mu_i^h - \lambda_i\right\|_U \right)$$

Abstract additive operator-splitting theory

Sandia National Laboratories

Solving the KKT system

Monolithic problem

$$(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{u} = \mathbf{f}$$

KKT

$$\begin{bmatrix} \mathbf{U} & -\mathbf{U} & 0 & \mathbf{Q}_1^T & 0 \\ -\mathbf{U} & \mathbf{U} & 0 & 0 & \mathbf{Q}_1^T \\ 0 & 0 & 0 & -\mathbf{V} & \mathbf{V} \\ \mathbf{Q}_2 & 0 & -\mathbf{V} & 0 & 0 \\ 0 & \mathbf{Q}_2 & \mathbf{V} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \theta \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{f} \\ 0 \end{bmatrix}$$

Can we really solve this 5X larger problem faster and more efficiently than the monolithic one?

Theorem

(Bochev & Ridzal, 2015)

Let $(\mathbf{u}_1, \mathbf{u}_2, \theta, \lambda_1, \lambda_2)$ be the solution of the full KKT system. Then

- $\lambda_1 = \lambda_2 = 0$ and $\mathbf{u} = \mathbf{u}_1 = \mathbf{u}_2$
- The triple $(\mathbf{u}_1, \mathbf{u}_2, \theta)$ solves the reduced KKT system

$$\begin{bmatrix} \mathbf{Q}_1 & 0 & -\mathbf{V} \\ 0 & \mathbf{Q}_2 & \mathbf{V} \\ 0 & 0 & (\mathbf{Q}_1^{-1} + \mathbf{Q}_2^{-1})\mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \theta \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \\ -\mathbf{Q}_1^{-1}\mathbf{f} \end{bmatrix}$$

This system provides a foundation for an efficient iterative procedure for the solution of the KKT system

Additive splitting ⟹ solution algorithm

Sandia National Laboratories

1. **Use GMRES to solve the reduced space equation** $\left(\mathbf{Q}_1^{-1}+\mathbf{Q}_2^{-1}\right)\mathbf{V}\theta=-\mathbf{Q}_1^{-1}\mathbf{f}$
   - Application of $\left(\mathbf{Q}_1^{-1}+\mathbf{Q}_2^{-1}\right)$ decouples trivially into linear system solves with $\mathbf{Q}_1$ and $\mathbf{Q}_2$.
   - By assumption sub-problems are **easier to solve** that the monolithic problem:

     $\Rightarrow$ $\mathbf{Q}_1$, $\mathbf{Q}_2$ are **easier to invert** than $\left(\mathbf{Q}_1+\mathbf{Q}_2\right)$

2. **Recover the state by solving either** $\mathbf{Q}_1\mathbf{u}_1=\mathbf{f}+\mathbf{V}\theta$ **or** $\mathbf{Q}_2\mathbf{u}_2=-\mathbf{V}\theta$
   - Since $\mathbf{u}=\mathbf{u}_1=\mathbf{u}_2$ both yield **the solution of the monolithic problem**!
   - Note that this also allows to further simplify the KKT system to

$$\begin{bmatrix}\mathbf{Q}_1 & -\mathbf{V}\\ \mathbf{Q}_2 & \mathbf{V}\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \theta\end{bmatrix}=\begin{bmatrix}\mathbf{f}\\ 0\end{bmatrix}$$

⟹ In principle one could bypass optimization and derive the split via auxiliary variables

- However, derivation of this system is not obvious at first!
- The variational setting and its discretization are left to a guesswork and serendipity.
- Optimization automates and formalizes the discovery of decompositions.

Application to the Navier-Stokes equations

Sandia National Laboratories

Focus on the Oseen equations

$$\begin{cases} -\nu\Delta\mathbf{u} + (\mathbf{b}\cdot\nabla)\mathbf{u} + \nabla p = \mathbf{f} & \text{in } \Omega \\ \nabla\cdot\mathbf{u} = 0 & \text{in } \Omega \\ \mathbf{u} = 0 & \text{on } \Gamma \end{cases}$$

Result from linearization of the Navier-Stokes equations (NSE) by fixed point or a Newton-type method

⇒ Availability of robust solvers with optimal complexity is prerequisite to solve the NSE

- Convergence should be at best independent of the mesh size and the viscosity
- Formulation of such solvers remains a challenge. Existing approaches include
  - physics-based splitting: vector Laplacian + convection term (Hamilton et al NLA, 2010)
  - dimension-based splitting: 1D scalar advection-diffusion (Benzi et al, ANM 2011)
  - Iterative algorithm design must be tailored to the splitting employed

- We apply optimization-based splitting to develop efficient solvers for Oseen equations
- Approach is agnostic to the type of splitting used, only requires well-posed subproblems
- Could in principle apply it with the same splitting as above

Specialization to the Oseen equations

Sandia National Laboratories

Variational setting for the monolithic problem

$$Q(\mathbf{u},p;\mathbf{v},q) = \nu(\nabla\mathbf{u},\nabla\mathbf{v}) + (\mathbf{b}\cdot\nabla\mathbf{u},\mathbf{v}) - (p,\nabla\cdot\mathbf{v}) + (q,\nabla\cdot\mathbf{u})$$

$$U = V = H_0^1(\Omega)\times L_0^2(\Omega); \quad H = L^2(\Omega)$$

$$(\mathbf{u},p;\mathbf{v},q)_U = (\nabla\mathbf{u},\nabla\mathbf{v}) + (p,q); \quad \|\mathbf{u},p\|_U^2 = \|\nabla\mathbf{u}\|_0^2 + \|p\|_0^2$$

Q is weakly coercive on $U\times V$: the monolithic problem satisfies our assumptions

Additive splitting: $Q(\mathbf{u},p;\mathbf{v},q) = Q_1(\mathbf{u},p;\mathbf{v},q) + Q_2(\mathbf{u},p;\mathbf{v},q)$

$$Q_1(\mathbf{u},p;\mathbf{v},q) = \sigma(\nabla\mathbf{u},\nabla\mathbf{v}) + (\mathbf{b}\cdot\nabla\mathbf{u},\mathbf{v}) - 2(p,\nabla\cdot\mathbf{v}) + 2(q,\nabla\cdot\mathbf{u})$$

$$Q_2(\mathbf{u},p;\mathbf{v},q) = (\nu-\sigma)(\nabla\mathbf{u},\nabla\mathbf{v}) + (p,\nabla\cdot\mathbf{v}) - (q,\nabla\cdot\mathbf{u})$$

Choosing a sufficiently large splitting parameter σ ensures that each subproblem is dominated by the Laplacian

Additive splitting: strong form

$$\begin{bmatrix} -\sigma\Delta\mathbf{u} + (\mathbf{b}\cdot\nabla)\mathbf{u} + 2\nabla p \\ 2\nabla\cdot\mathbf{u} \end{bmatrix} + \begin{bmatrix} (\sigma-\nu)\Delta\mathbf{u} - \nabla p \\ -\nabla\cdot\mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}$$

"Easy" Oseen **Stokes**

Each subproblem is weakly coercive on $U\times V$: the split satisfies our assumptions

Specialization to the Oseen equations

Sandia National Laboratories

Optimization reformulation

$$\begin{cases} \text{minimize} \quad J(\mathbf{u}_1, p_1; \mathbf{u}_2, p_2) = \frac{1}{2}\left(\|\nabla \mathbf{u}_1 - \nabla \mathbf{u}_2\|_0^2 + \|p_1 - p_2\|_0^2 \right) \\ \text{s.t.} \begin{cases} Q_1(\mathbf{u}_1, p_1; \mathbf{v}_1, q_1) - (\xi, \mathbf{v}_1)_1 - (r, q_1)_0 = \langle f, \mathbf{v}_1 \rangle \quad \forall \mathbf{v}_1, q_1 \in U \\ Q_2(\mathbf{u}_2, p_2; \mathbf{v}_2, q_2) + (\xi, \mathbf{v}_2)_1 + (r, q_2)_0 = 0 \qquad \forall \mathbf{v}_2, q_2 \in U \end{cases} \end{cases}$$

Virtual distributed control

$$\theta = \{\xi, r\} \in H_0^1(\Omega) \times L_0^2(\Omega)$$

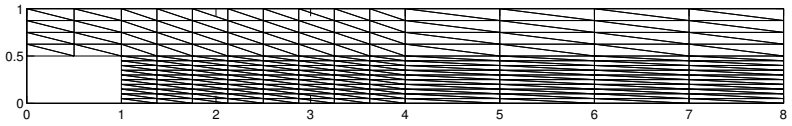
Interpretation via auxiliary variables

$$\begin{bmatrix} -\sigma\Delta\mathbf{u} + (\mathbf{b}\cdot\nabla)\mathbf{u} + 2\nabla p - \Delta\xi \\ 2\nabla\cdot\mathbf{u} - r \end{bmatrix} + \begin{bmatrix} (\sigma - \nu)\Delta\mathbf{u} - \nabla p + \Delta\xi \\ -\nabla\cdot\mathbf{u} + r \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}$$

As in the abstract case

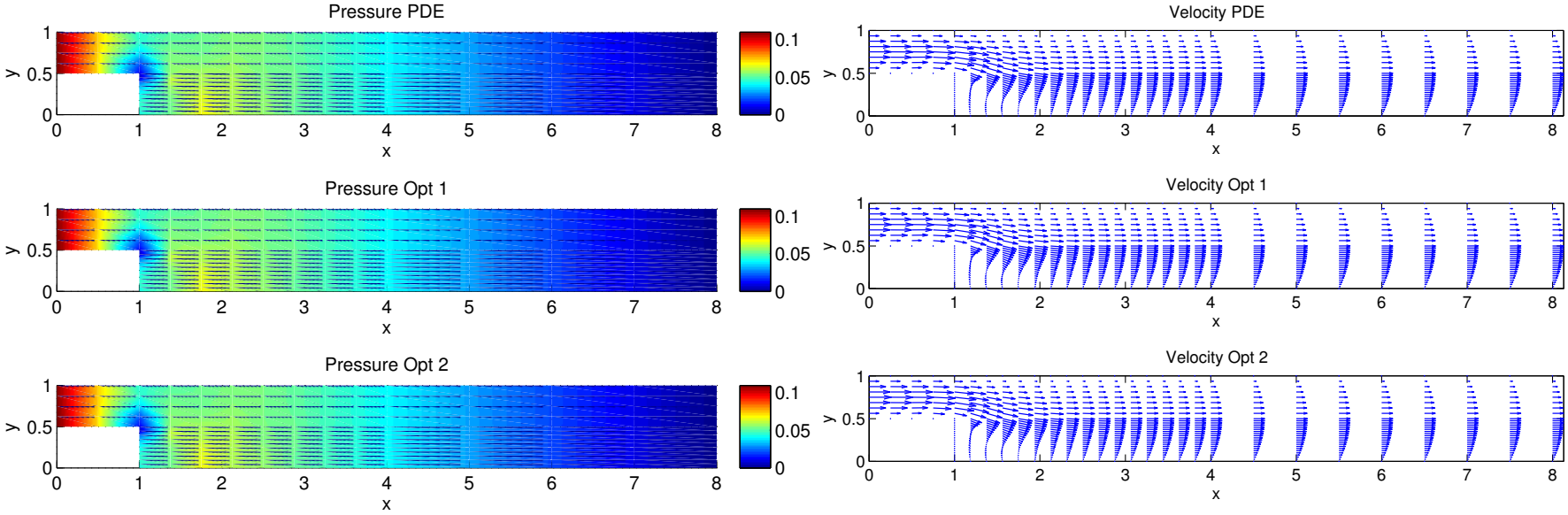
- Derivation of this system is not obvious at first!
- The variational setting and its discretization are left to a guesswork and serendipity.
- Optimization automates and formalizes the discovery of decompositions.

Numerical examples



Backward facing step channel geometry and a typical mesh

Equivalence of reformulated and monolithic problems



$$\sum_{i=1,2} \left\| \mathbf{u}^h - \mathbf{u}_i^h \right\|_0^2 + \left\| p^h - p_i^h \right\|_0^2 = 1.0 \times 10^{-7}$$

Solutions of optimization-based decomposition match monolithic solution to within the GMRES tolerance set to 1.0E-08

Numerical results

Sandia National Laboratories

Optimization-based solver is independent of the mesh size

Mesh level	# cells	#DoF	#GMRES
1	352	1,727	**37**
2	1,408	6,619	**39**
4	5,632	25,907	**40**
8	22,528	102,499	**40**
16	90,112	407,747	**38**

Parameters

GMRES tol=10^{-6}

$\nu = 5 \times 10^{-3}$

$\sigma = 1$

Optimization-based solver is mildly dependent on viscosity

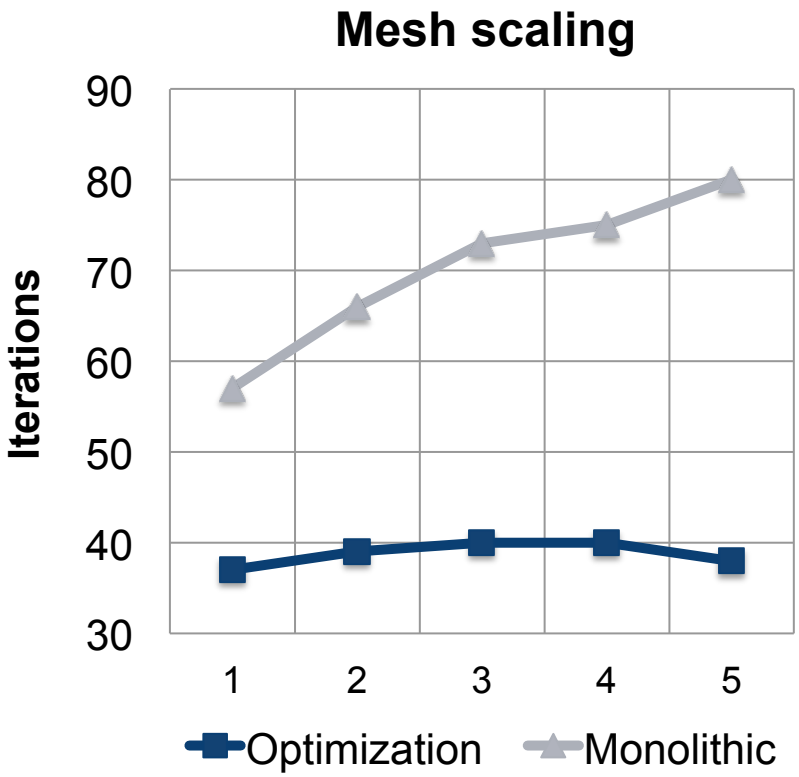
Visc.	1E+2	1E+1	1E-1	1E-2	5E-3
#GMRES	4	4	6	22	40

Viscosity **decreases** by **five** orders of magnitude.
Iterations **increase** by a **single** order of magnitude.

Numerical results

Sandia National Laboratories

Optimization-based vs. preconditioned monolithic solver



Parameters

GMRES tol=10^{-6}, $\nu = 5\times10^{-3}$, $\sigma = 1$.

Monolithic solver

$\left(\mathbf{Q}_1 + \mathbf{Q}_2\right)$ preconditioned by $\mathbf{Q}_1$:

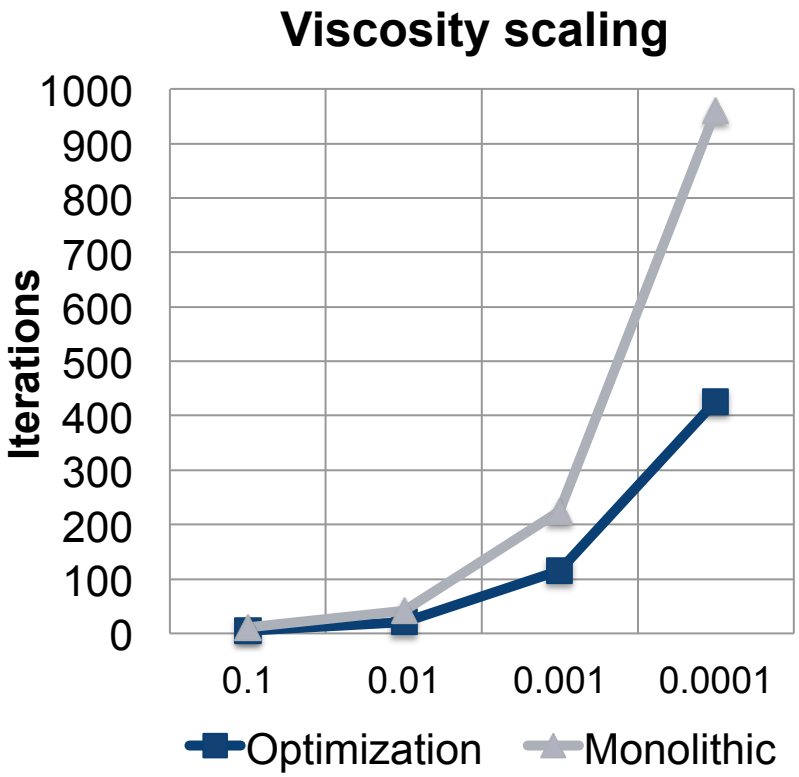
$$\mathbf{Q}_1 \sim \begin{bmatrix} -\sigma\Delta\mathbf{u} + (\mathbf{b}\cdot\nabla)\mathbf{u} + 2\nabla p \\ 2\nabla\cdot\mathbf{u} \end{bmatrix}$$

"Easy" Oseen

Numerical results

Sandia National Laboratories

Optimization-based vs. preconditioned monolithic solver



Parameters

GMRES tol=10^{-6}, $\sigma = 1$, Level 8 mesh.

Monolithic solver

$\left(\mathbf{Q}_1 + \mathbf{Q}_2\right)$ preconditioned by $\mathbf{Q}_1$:

$$\mathbf{Q}_1 \sim \begin{bmatrix} -\sigma\Delta\mathbf{u} + (\mathbf{b}\cdot\nabla)\mathbf{u} + 2\nabla p \\ 2\nabla\cdot\mathbf{u} \end{bmatrix}$$

"Easy" Oseen

These preliminary results do not explore further tuning of the optimization solver by virtue of the splitting parameter

References

1. D. Olson, M. Luskin, A. Shapeev and P. Bochev, Analysis of an optimization-based atomistic-to-continuum coupling method for point defects. *ESAIM, 2015*. SAND2014-18401J.
2. D. Olson, P. Bochev, M. Luskin and A. Shapeev. An optimization-based atomistic-to-continuum coupling method. *SIAM. J. Num. Anal*.. Vol. 52, Issue 4, pp.2183-2204 (2014)
3. D. Olson, P. Bochev, M. Luskin and A. Shapeev. Development of an optimization-based atomistic-to-continuum coupling method. In. Lirkov, Wasniewski, editors, Large-Scale Scientific Computing, Vol. 8353, LNCS, pp. 33-44, Springer Berlin, Heidelberg, 2014.
4. M. D'Elia and P. B. Bochev. Optimization-based local-to-nonlocal coupling method, Sandia Technical Report, No. SAND2014-17373J (2014).
5. M. D'Elia and P. B. Bochev, Optimization-Based Coupling of Nonlocal and Local Diffusion Models. *Materials Research Society Proceeding*, (2014) .
6. M. D'Elia and P. Bochev, Formulation, analysis and computation of an optimization-based local-to-nonlocal coupling method. Submitted to *SIAM. J. Num. Anal*. SAND2014-17373J.
7. P. Bochev and D. Ridzal, *An optimization-based approach for the design of robust solution algorithms*. SIAM J. Num. Anal., vol. 47, No. 5, pp.3938-3955, 2009.
8. P. Bochev and D. Ridzal, *Additive Operator Decomposition and Optimization-Based Reconnection with Applications.* Proceedings of LSSC 2009, Springer Lecture Notes in Computer Science, LNCS 5910, 2010
9. P. Bochev and D. Ridzal, *Optimization-based additive operator decomposition of weakly coercive problems with applications,* CAMWA, 2016.

Sandia National Laboratories

Part 2

Case study 2: Transport schemes

In Part 2 we apply optimization ideas to develop property-preserving methods for transport of passive tracers in climate models.

Minimize

Coupling mismatch $\left\|u_1^h - u_2^h\right\|^2$ + $\sum_i \left\|u_i^h - u_i^T\right\|^2$ **Target mismatch**

Subject to

Target definition $L_1^h(u_1^T, \theta) = f_1$ $\quad \underline{C}_i \le C u_i^h \le \overline{C}_i$ **Physical properties**

$L_2^h(u_2^T, \theta) = f_2$ $\quad B u_i^h = b$

This case study highlights application of optimization ideas for the preservation of relevant physical properties in numerical methods.

P. Bochev, D. Ridzal, M. Shashkov, Fast optimization-based conservative remap of scalar fields, J. Comp. Phys. 246 (2013)

P. Bochev, D. Ridzal, K. Peterson, Optimization-based remap and transport: A divide and conquer strategy for feature- preserving discretizations, J. Comp. Phys. 257, (2014) 1113 – 1139.

Transport of passive tracers

Sandia National Laboratories

An ubiquitous problem in geosciences and climate modeling

$$\left.\begin{aligned} &\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{u} = 0 \\ &\frac{\partial \rho q}{\partial t} + \nabla \cdot \rho q \mathbf{u} = 0 \end{aligned}\right\} \Rightarrow \quad \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0$$

where:

- ρ - density
- q - tracer mixing ratio
- $\mathbf{u}$ - velocity

Key requirements

1. Conservation of mass and total tracer: $M = \int_\Omega \rho \, dV \qquad Q = \int_\Omega \rho q \, dV$
2. Preservation of local bounds for q and ρ: $\rho_i^{\min} \le \rho_i \le \rho_i^{\max} \qquad q_i^{\min} \le q_i \le q_i^{\max}$
3. Preservation of linear correlations between tracers: $q_1(x) = a q_2(x) + b$
4. Preservation of constant tracers, i.e., “compatibility”.

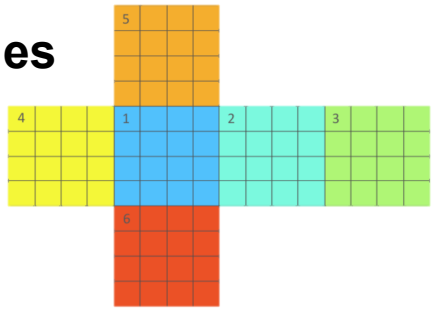
Semi-Lagrangian schemes are the method of choice in these communities because they allow for time steps much larger than the CFL-limited time steps in Explicit Eulerian methods. This is even more critical for recent high-order nodal schemes deployed in climate models.

Why SL + SE?

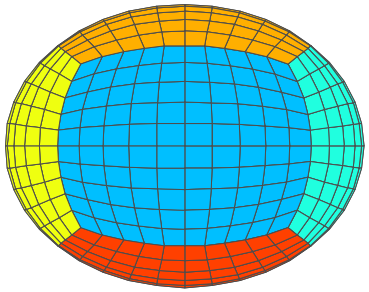
Sandia National Laboratories

We begin Part 2 by developing a new scheme, which combines

- **Spectral elements** (SE) for spatial discretization.
- **Semi-Lagrangian** (SL) approach for time stepping.
- **Optimization** for enforcing conservation and local bounds.



cubed-sphere mesh



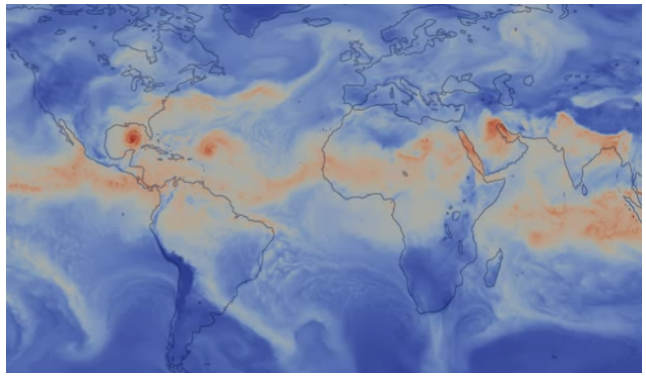
Advantages

- **SE: Diagonal** mass matrix + **Spectral** accuracy
- **SL: avoids severe CFL** restrictions of high-order methods
- **SL+SE: Simple, efficient and accurate!**
- **HOMME** (High Order Modeling Environment) uses SE and DG on fully unstructured quadrilateral meshes on the sphere

HOMME is a community model supported by the NSF and the DOE with contributions from NCAR, DOE laboratories and universities.

HOMME is the default dynamical core of the Community Atmosphere Model (CAM) and the Community Earth System Model (CESM)

The new SL-SE scheme for tracers is motivated by and implemented in HOMME.



Dennis J, Edwards J, Evans K, Guba O, Lauritzen P, Mirin A, St.-Cyr A, Taylor M, Worley P. 2012. CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. IJHPCA. 26:74-89.

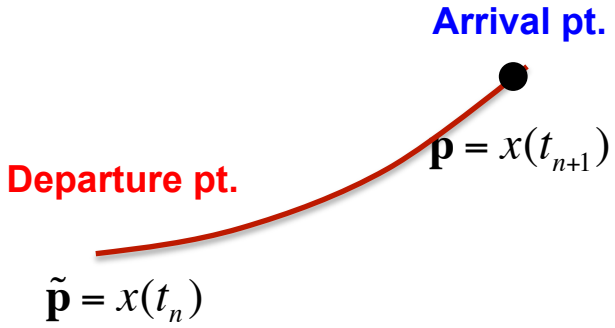
A generic nodal SL transport scheme

Key idea: convert PDEs into ODEs along Lagrangian particle paths

$$\left\{\begin{array}{l}\dfrac{\partial \rho}{\partial t}+\nabla\cdot\rho\mathbf{u}=0 \rightarrow \dfrac{\partial \rho}{\partial t}+\mathbf{u}\cdot\nabla\rho=-\rho\nabla\cdot\mathbf{u}\\ \dfrac{\partial \rho q}{\partial t}+\nabla\cdot\rho q\mathbf{u}=0 \rightarrow \dfrac{\partial q}{\partial t}+\mathbf{u}\cdot\nabla q=0\end{array}\right\} \rightarrow \frac{dx}{dt}=\mathbf{u}(x(t),t) \rightarrow \left\{\begin{array}{l}\dfrac{D\rho}{Dt}=-\rho\nabla\cdot\mathbf{u}\\ \dfrac{Dq}{Dt}=0\end{array}\right.$$

Step 1: solve the "final value" problem in $[t_n,t_{n+1}]$:

$$\frac{dx}{dt}=\mathbf{u}(x(t),t) \text{ and } x(t_{n+1})=\mathbf{p} \rightarrow \tilde{\mathbf{p}}=x(t_n)$$



Step 2: solve the initial value problems in $[t_n,t_{n+1}]$:

$$\frac{D\rho}{Dt}=-\rho\nabla\cdot\mathbf{u} \text{ and } \rho(t_n)=\rho_h(\tilde{\mathbf{p}},t_n) \rightarrow \rho_h(\mathbf{p},t_{n+1})=\rho(t_{n+1})$$

ODE solution at t_{n+1} = PDE solution at **arrival pt.**

$$\frac{Dq}{Dt}=0 \text{ and } q(t_n)=q_h(\tilde{\mathbf{p}},t_n) \rightarrow q_h(\mathbf{p},t_{n+1})=q(t_{n+1})$$

Initial value at t_n = PDE solution at **departure pt.**

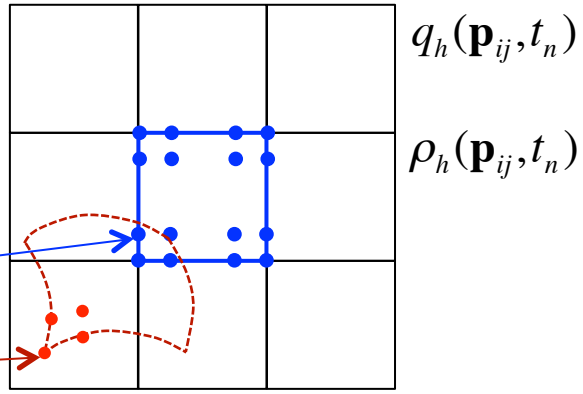
Combine with SE reconstruction

Step 1: solve the "final value" problem in $[t_n, t_{n+1}]$:

$$\frac{dx}{dt} = \mathbf{u}(x(t),t) \quad \text{and} \quad x(t_{n+1}) = \mathbf{p}_{ij}$$

$\{\mathbf{p}_{ij}\}$ → Arrival points = Gauss-Lobatto points

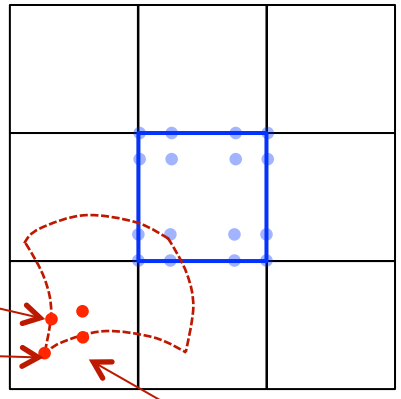
$\{\tilde{\mathbf{p}}_{ij}\}$ → Departure points



Step 2: solve the initial value problems in $[t_n, t_{n+1}]$:

$$\frac{D\rho}{Dt} = -\rho \nabla \cdot \mathbf{u} \quad \text{and} \quad \rho(t_n) = \rho_h(\tilde{\mathbf{p}}_{ij}, t_n)$$

$$\frac{Dq}{Dt} = 0 \quad \text{and} \quad q(t_n) = q_h(\tilde{\mathbf{p}}_{ij}, t_n)$$



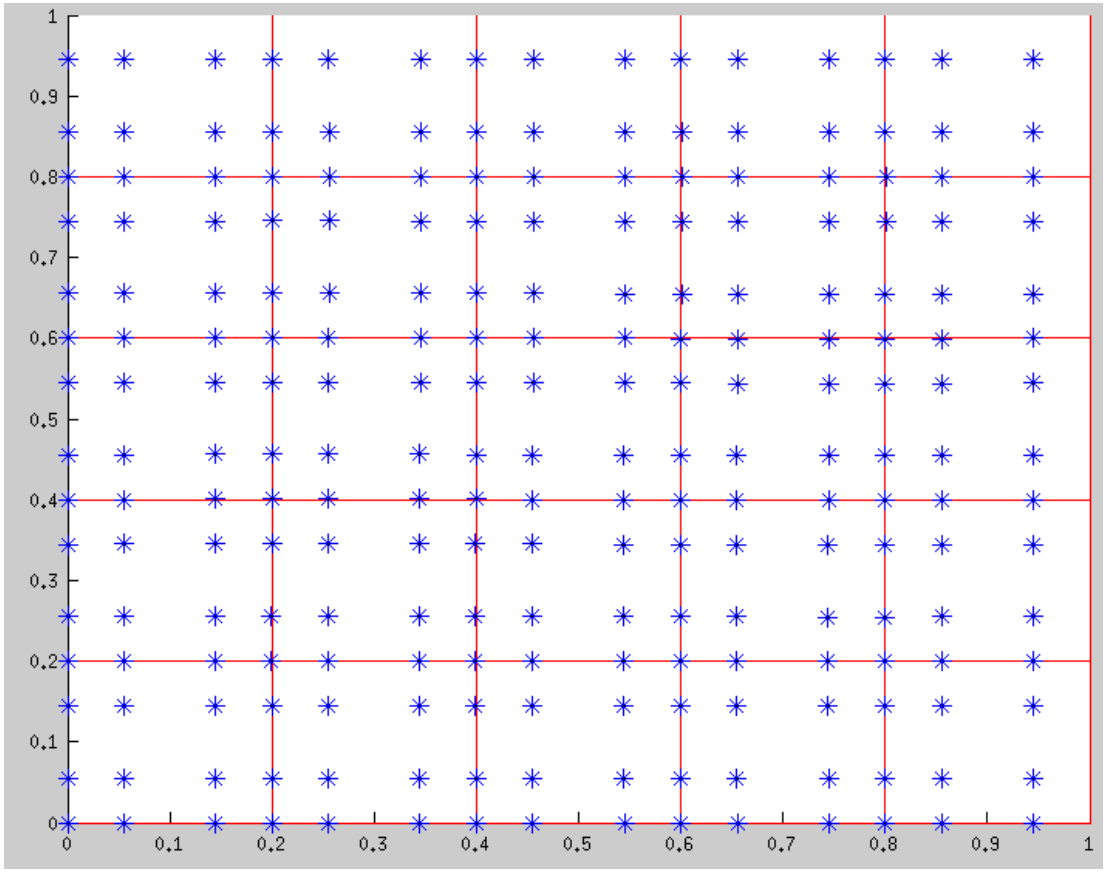
Initial values = spectral element reconstruction at Gauss-Lobatto **departure points**

Example: rotation

$$\frac{dx}{dt} = \mathbf{u}(x(t),t) \quad \text{and} \quad x(t_{n+1}) = \mathbf{p}_{ij}$$

$$\mathbf{u}(x(t),t) = \begin{pmatrix} 0.5 - y \\ 0.5 - x \end{pmatrix}$$

Solved by RK4

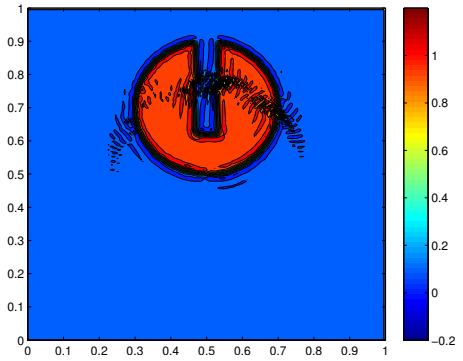


Generic SE+SL scheme scorecard

Sandia National Laboratories

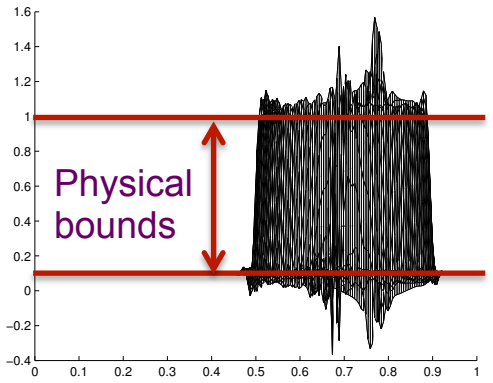
Recall the advantages:

- **Diagonal** mass matrix
- **Spectral** accuracy
- **Avoids severe CFL** restrictions of high-order methods
- **Simple!!** (compare, e.g., to tent-pitching schemes)



However, the generic scheme

- **Does not conserve mass and total tracer**
- **Does not preserve local solution bounds**



Critical for physically consistent tracer transport, since high-order spatial schemes are prone to unphysical oscillations:

Solution: combine the generic SE+SL scheme with optimization to

- **Conserve mass and total tracer**
- **Preserve local solution bounds**

Optimization-based SE-SL scheme

Sandia National Laboratories

Start with a generic SE+SL scheme:

1. Determine GL departure points → $\tilde{\mathbf{p}}_{ij} = x(t_n)$
2. Determine solution at arrival points → $\rho_h(\mathbf{p}_{ij}, t_{n+1}) = \rho(t_{n+1})$ and $q_h(\mathbf{p}_{ij}, t_{n+1}) = q(t_{n+1})$

Then proceed as follows to find the tracer at t_{n+1} (density is similar)

3. Set optimization target to SE+SL solution: $\hat{q} := q_h(\mathbf{p}_{ij}, t_{n+1})$
4. Determine local solution bounds: $q_{ij}^{\min} \le q(\mathbf{p}_{ij}, t_{n+1}) \le q_{ij}^{\max}$ → **TBD later!**
5. Set solution at the new time step by solving

$$q_{n+1}^{*} = \underset{q \in Q^r}{\operatorname{argmin}} \left\| q - \hat{q} \right\|_0^2 \quad \text{subject to} \quad \begin{cases} \int_\Omega q\,dx = \int_\Omega q_n\,dx & \leftarrow \text{Conservation} \\ q_{ij}^{\min} \le q_{ij} \le q_{ij}^{\max} & \leftarrow \text{Local bounds} \end{cases}$$

The optimization problem

Sandia National Laboratories

Algebraic form

$$\mathbf{q}_{n+1} = \underset{\mathbf{q}}{\operatorname{argmin}}\, \mathbf{q}^T\mathbf{M}\mathbf{q} + \mathbf{c}^T\mathbf{q} + \mathbf{c}_0 \quad \text{subject to} \quad \begin{cases} \mathbf{w}^T\mathbf{q} = \mathbf{w}^T\mathbf{q}_n & \leftarrow \text{Conservation} \\ \mathbf{q}^{\min} \le \mathbf{q} \le \mathbf{q}^{\max} & \leftarrow \text{Local bounds} \end{cases}$$

$$\mathbf{M} = \int_\Omega \phi_{ij}\phi_{kl}\, dx = diag(M_{ij}); \quad \mathbf{c} = -2\mathbf{M}\hat{\mathbf{q}}; \quad \mathbf{c}_0 = \hat{\mathbf{q}}^T\mathbf{M}\hat{\mathbf{q}}; \quad \mathbf{w} \rightarrow \text{Gauss-Lobato weights}$$

- ☞ Example of a "*singly linearly constrained QP with simple bounds*"
- ☞ QP structure admits a fast *O(N)* optimization algorithm.

Theorem (Existence of optimal solutions)

The feasible set of the optimization problem for the solution transfer is non-empty. The problem has a unique optimal solution.

Fast Optimization Algorithm

Sandia National Laboratories

Without the equality constraint the QP splits into *N* one-dimensional QPs with simple bounds:

$$q_{ij,n+1} = \underset{q_{ij}}{\operatorname{argmin}} M_{ij}\left(q_{ij} - \hat{q}_{ij}\right)^2 \quad \text{subject to} \quad q_{ij}^{\min} \le q_{ij} \le q_{ij}^{\max}$$

$$\Rightarrow \quad q_{ij,n+1} = med(q_{ij}^{\min}, \hat{q}_{ij}, q_{ij}^{\max})$$

The Lagrangian

$$L(\mathbf{q}, \lambda, \mu_1, \mu_2) = \sum_{node} M_{ij}\left(q_{ij} - \hat{q}_{ij}\right)^2 - \lambda \sum_{node} w_{ij}\left(q_{ij} - q_{ij,n}\right) - \sum_{node} \mu_{1,ij}(q_{ij} - q_{ij}^{\min}) - \sum_{node} \mu_{2,ij}(q_{ij} - q_{ij}^{\max})$$

The Karush-Kuhn-Tucker (KKT) conditions

$$\begin{cases} q_{ij} = \hat{q}_{ij} + \lambda + \mu_{1,ij} - \mu_{2,ij} \\ q_{ij}^{\min} \le q_{ij} \le q_{ij}^{\max} \\ \mu_{1,ij} \ge 0, \quad \mu_{2,ij} \ge 0 \\ \mu_{1,ij}(q_{ij} - q_{ij}^{\min}) = 0, \\ \mu_{2,ij}(q_{ij} - q_{ij}^{\max}) = 0 \end{cases} \quad \text{and} \quad \sum_{node} w_{ij}\left(q_{ij} - q_{ij,n}\right) = 0$$

Without the equality constraint the KKT conditions are fully separable and can be solved for any fixed value of λ.

Fast Optimization Algorithm

Sandia National Laboratories

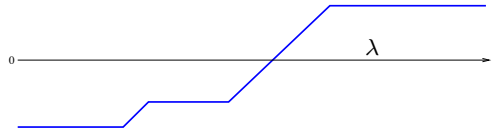
Step 1: solve the first set of KKT conditions to find *q* as a function of *λ*

$$\begin{cases} q_{ij} = \tilde{q}_{ij} + \lambda;\ \mu_{1,ij} = 0;\ \mu_{2,ij} = 0 & \text{if}\quad q_{ij}^{\min} \le \tilde{q}_{ij} + \lambda \le q_{ij}^{\max} \\ q_{ij} = q_{ij}^{\min};\quad \mu_{2,ij} = 0;\ \mu_{1,ij} = q_{ij} - \tilde{q}_{ij} - \lambda & \text{if}\quad q_{ij}^{\min} \ge \tilde{q}_{ij} + \lambda \\ q_{ij} = q_{ij}^{\max};\quad \mu_{1,ij} = 0;\ \mu_{2,ij} = \tilde{q}_{ij} - q_{ij} + \lambda & \text{if}\quad \tilde{q}_{ij} + \lambda \ge q_{ij}^{\max} \end{cases} \Rightarrow q_{ij}(\lambda) = med(q_{ij}^{\min}, \tilde{q}_{ij} + \lambda, q_{ij}^{\max});$$

Trivial, communication-free O(N) computation

Step 2: solve the single equality constraint for *λ*

$$\text{Solve} \quad \sum_{node} w_{ij}\left(q_{ij}(\lambda) - q_{ij,n}\right) = 0$$



- Piecewise linear, monotonically increasing function of single scalar variable *λ*
- Can solve to machine precision by a simple secant method
- Globalization is unnecessary: $\lambda_0 = 0$ is an excellent initial guess: $q_{ij}(\lambda_0) = med(q_{ij}^{\min}, \tilde{q}_{ij}, q_{ij}^{\max});$
- $q_{ij}(\lambda_0)$ solves the QP without the equality constraint, i.e., "almost" a solution
- Locality $\Rightarrow q_{ij}(\lambda_0)$ barely violates the mass conservation constraint

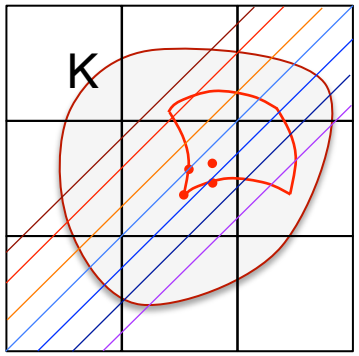
Local solution bounds: $\nabla \cdot \mathbf{u} = 0$

For solenoidal fields local bounds are easy:

$$\frac{D\rho}{Dt} = 0 \quad \Rightarrow \quad \rho(x(t),t) = const \quad \rightarrow \quad \rho_h(\mathbf{p}_{ij}, t_{n+1}) = \rho(t_{n+1}) = \rho(t_n) = \rho_h(\tilde{\mathbf{p}}_{ij}, t_n)$$

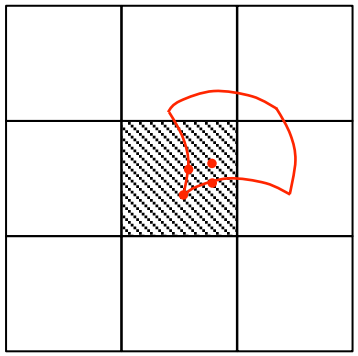
$$\frac{Dq}{Dt} = 0 \quad \Rightarrow \quad q(x(t),t) = const \quad \rightarrow \quad q_h(\mathbf{p}_{ij}, t_{n+1}) = q(t_{n+1}) = q(t_n) = q_h(\tilde{\mathbf{p}}_{ij}, t_n)$$

Solution is constant along Lagrangian paths ⇒ taking min/max in a neighborhood of the departure points is sufficient to determine solution bounds:

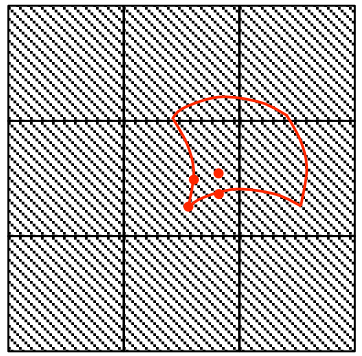


$$q_{ij}^{\min} = \min_{\mathbf{p} \in K} q(\mathbf{p}_{ij}, t_n)$$

$$q_{ij}^{\max} = \max_{\mathbf{p} \in K} q(\mathbf{p}_{ij}, t_n)$$



Tight bounds

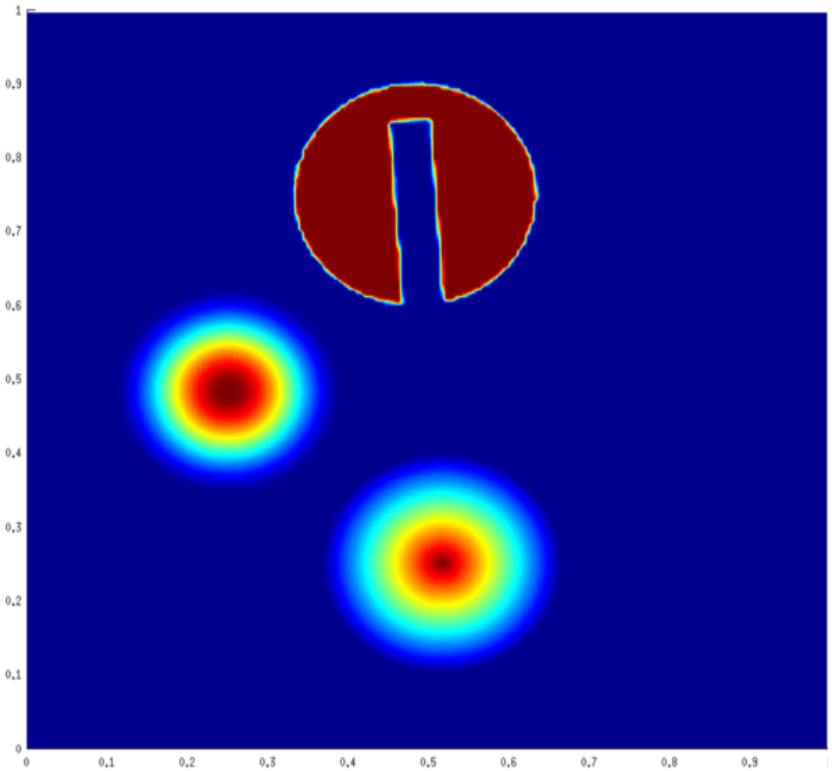


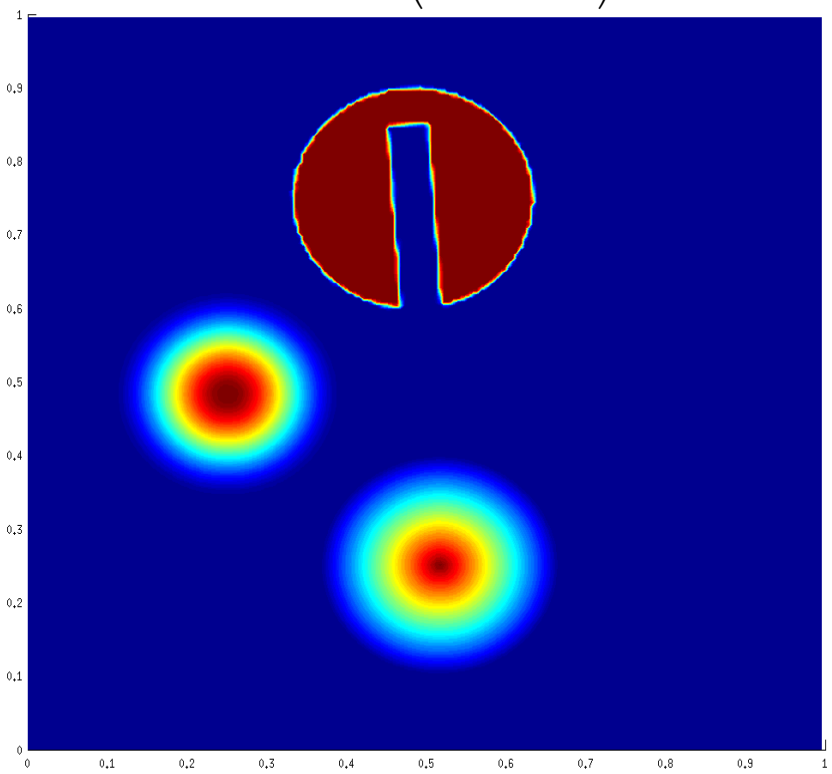
Loose bounds

Rotational flow: LeVeque's combo

Zalesak cylinder, cone and a smooth hump

$$\mathbf{u}(\mathbf{p},t) = \begin{pmatrix} 0.5 - y \\ 0.5 - x \end{pmatrix}$$





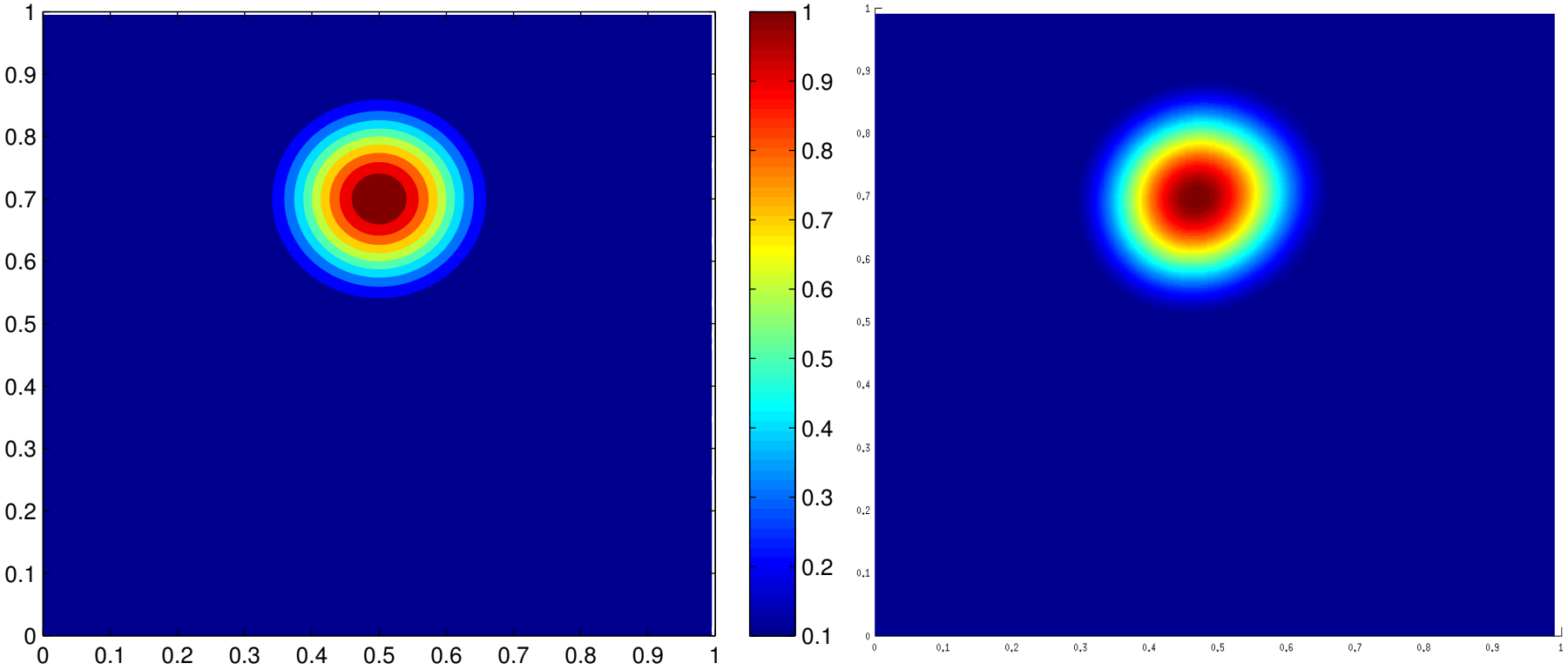
80x80 bi-cubic elements; CFL=0.7

R. J. LeVeque, High-resolution conservative algorithms for advection in incompressible flow, SINUM 33 (1996) 627–665.

Deformational flow: cosine bell

$$q(x,t) = 0.5(1.0 + \cos(\pi r_1)); \quad r_1 = \frac{\min(r, r_0)}{r_0}$$

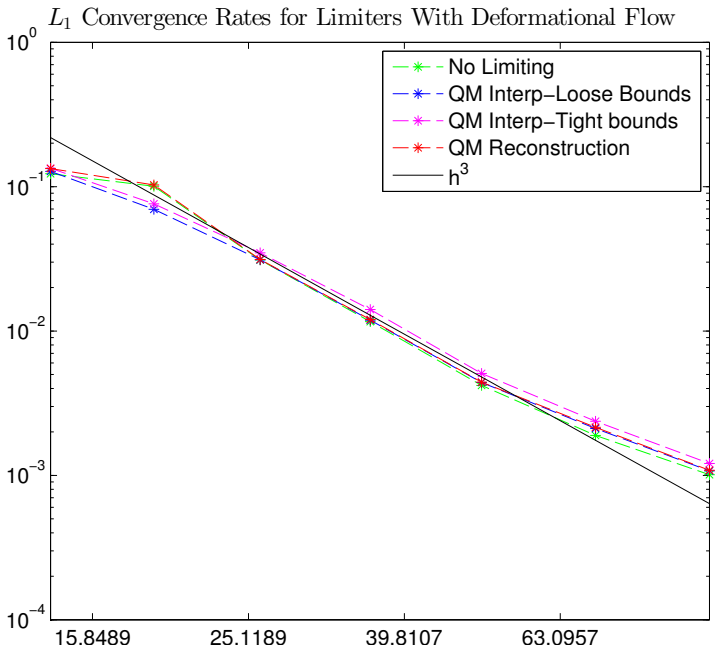
$$\mathbf{u}(\mathbf{p},t) = \begin{pmatrix} \sin(\pi x)^2 \sin(2\pi y)\cos(\pi t / T) \\ -\sin(\pi y)^2 \sin(2\pi x)\cos(\pi t / T) \end{pmatrix}$$

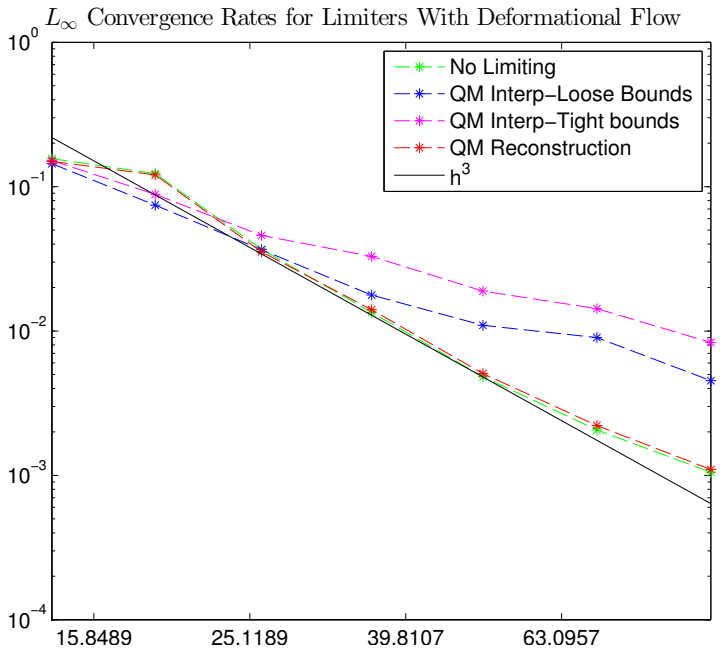


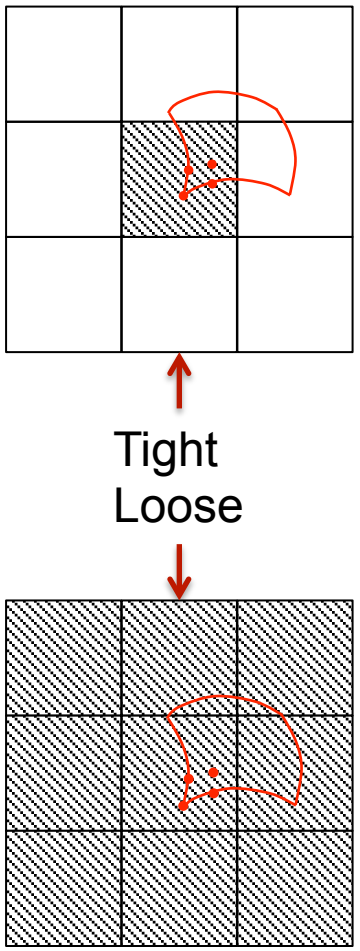
80x80 bi-cubic elements; CFL=0.7

R. J. LeVeque, High-resolution conservative algorithms for advection in incompressible flow, SINUM 33 (1996) 627–665.

Convergence rates: Gaussian hill



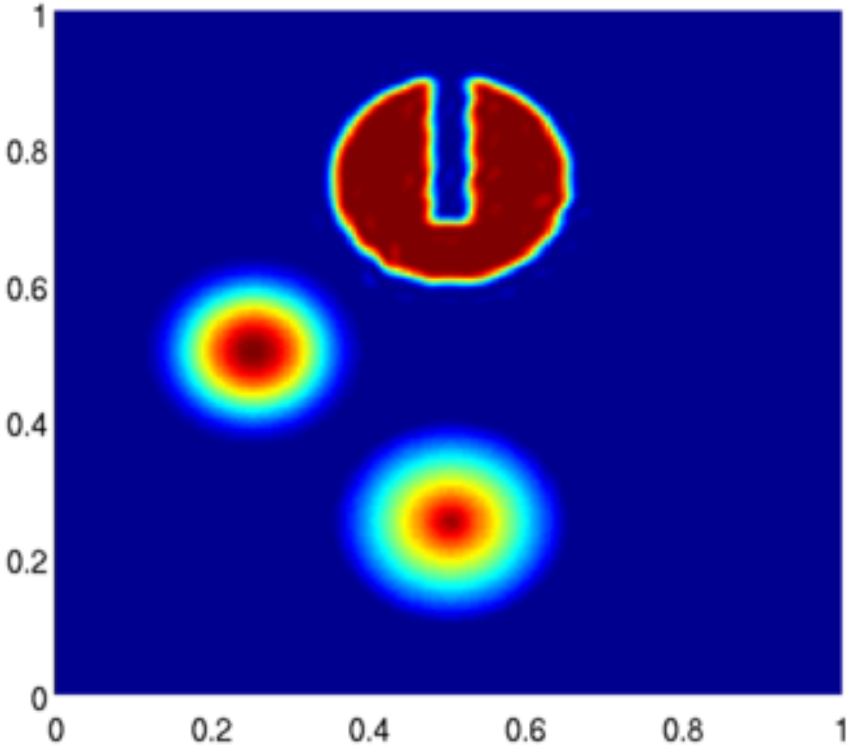




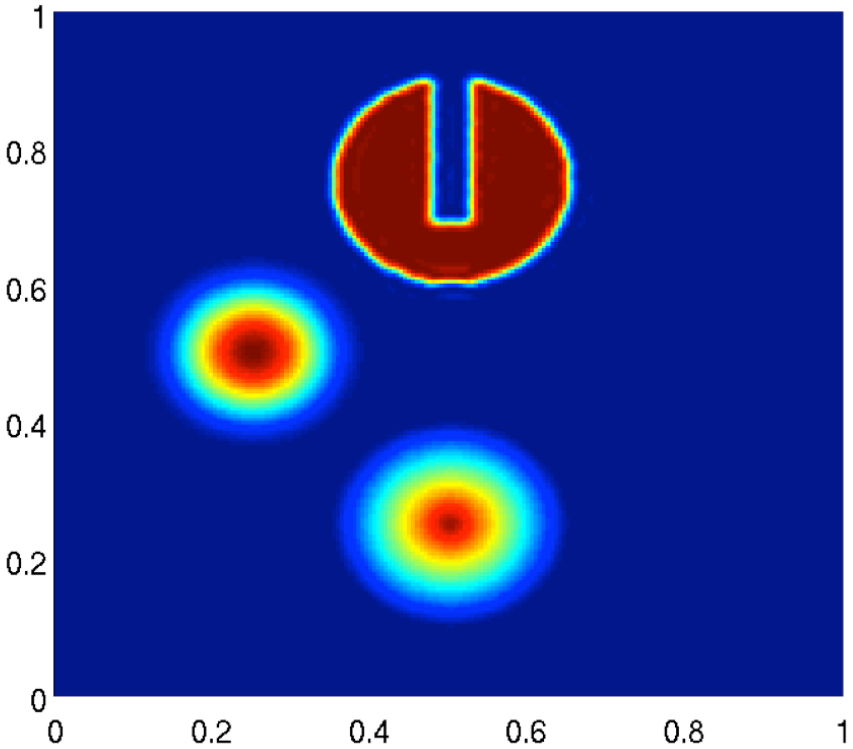
A SE+SL method with limiters would typically truncate the order of convergence to 2 even for L1 errors.

We see essentially no degradation in the 3rd order L1 error rate (compared to “raw” solution convergence).

Long-time accuracy



CFL=7.04



CFL=14.08

60x60 bi-cubic elements; 20 full revolutions.

Local solution bounds: $\nabla \cdot \mathbf{u} \neq 0$

Sandia National Laboratories

For $\nabla \cdot \mathbf{u} \neq 0$ the density equation is a balance rather than a conservation law.

⇒ The density is not constant along Lagrangian paths.

⇒ Taking min/max in a neighborhood of the departure points is not appropriate.

Solution: combining the Geometric Conservation Law and the balance law

$$\left.\begin{aligned} &\textbf{GC Law:} && \frac{\partial V}{\partial t} + \mathbf{u}\cdot\nabla V = V\nabla\cdot\mathbf{u} \\ &\textbf{Balance law:} && \frac{\partial \rho}{\partial t} + \mathbf{u}\cdot\nabla\rho = -\rho\nabla\cdot\mathbf{u} \end{aligned}\right\} \Rightarrow \frac{\partial \rho V}{\partial t} + \mathbf{u}\cdot\nabla(\rho V) = 0 \Rightarrow \boxed{\frac{D(\rho V)}{Dt} = 0}$$

yields a **new conservation law** for the point "mass" distribution $M := \rho V$.

The idea is to associate and track an arbitrary initial volume V_0 and "mass" with every GLL point and use these quantities to provide bounds for the density. This resembles what we do in a finite volume semi-Lagrangian scheme (next topic).

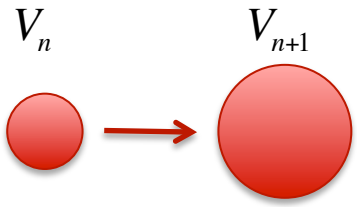
Local solution bounds: $\nabla \cdot \mathbf{u} \neq 0$

Sandia National Laboratories

Assume V_n and M_n are given at t_n :

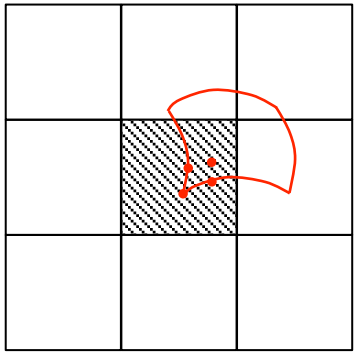
4a. Solve the GCL in $[t_n, t_{n+1}]$

$$\frac{DV}{Dt} = V \nabla \cdot \mathbf{u} \quad \text{and} \quad V(t_n) = V_n \quad \rightarrow \quad \boxed{V_{n+1} = V(t_{n+1})}$$



4b. Determine local bounds for the point masses:

$$M_{ij}^{\min} = \min_{\mathbf{p} \in K} M(\mathbf{p}_{ij}, t_n) \qquad M_{ij}^{\max} = \max_{\mathbf{p} \in K} M(\mathbf{p}_{ij}, t_n)$$



4c. Determine local bounds for the density:

$$\rho^{\min} = \frac{M^{\min}}{V_{n+1}} \qquad \rho^{\max} = \frac{M^{\max}}{V_{n+1}}$$

$$M_{ij}^{\max} = \max_{\mathbf{p} \in K} M(\mathbf{p}_{ij}, t_n)$$

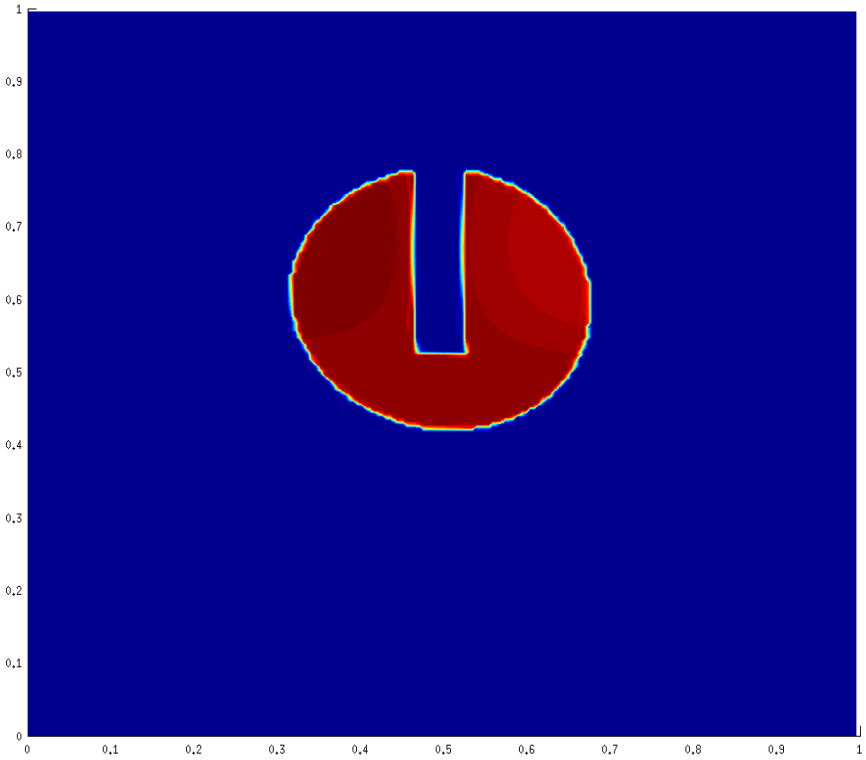
$$M_{ij}^{\min} = \min_{\mathbf{p} \in K} M(\mathbf{p}_{ij}, t_n)$$

4d. Solve the mass law in $[t_n, t_{n+1}]$

$$\frac{DM}{Dt} = 0 \quad \text{and} \quad M(t_n) = M_n$$

Divergent flow

$$\mathbf{u}(\mathbf{p},t) = \begin{pmatrix} -\sin(\pi x)^2 \sin(2\pi(y-0.5))\cos(\pi(y-0.5)^2 \cos(\pi t/T) \\ \frac{1}{2}\sin(\pi x)\cos(\pi(y-0.5))^3 \cos(\pi t/T) \end{pmatrix}$$



80x80 bi-cubic elements; CFL=0.7

A finite volume semi-Lagrangian (SL) scheme

Sandia National Laboratories

Why do we care about finite volume schemes?

- Cell-centered schemes are ubiquitous in DOE codes. However,
  - These schemes use monotone reconstruction, i.e., limiters to control bounds.
  - Limiters use local "worst case" scenarios when enforcing the bounds.
  - Limiters entangle accuracy with preservation of bounds, which obscures sources of discretization errors.
- Besides getting a better scheme we will have another chance to showcase the use of optimization to preserve physical properties!

Cell-centered discretization of density and tracer

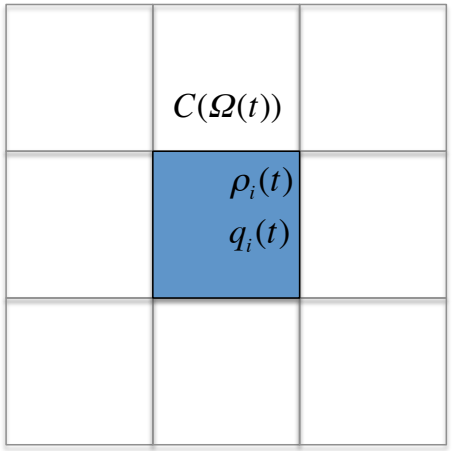
$$\mu_i = \int_{C_i} dx \quad \text{Cell area}$$

$$m_i = \int_{C_i} \rho\, dx \quad \text{Cell mass}$$

$$Q_i = \int_{C_i} \rho q\, dx \quad \text{Cell tracer}$$

$$\rho_i = \frac{m_i}{\mu_i} \quad \text{Cell average density}$$

$$q_i = \frac{Q_i}{m_i} \quad \text{Cell average tracer}$$



A generic finite volume SL scheme

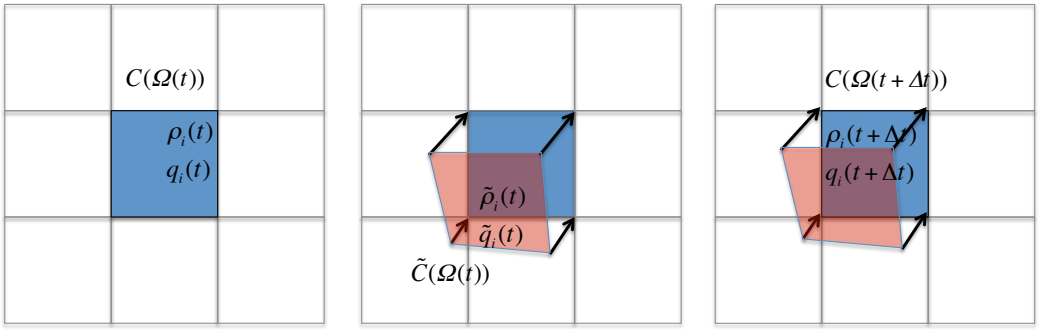
Sandia National Laboratories

For Lagrangian volumes

Dukowicz and Baumgardner (2000) JCP

$$\frac{d}{dt}\int_{C_i} \rho\, dx = 0 \longrightarrow m_i(t+\Delta t) = m_i(t)$$

$$\frac{d}{dt}\int_{C_i} \rho q\, dx = 0 \longrightarrow Q_i(t+\Delta t) = Q_i(t)$$



Step 1: **Trace back** cell vertices to find the Lagrangian (departure) grid $\tilde{C}(\Omega(t))$

Step 2: Remap Lagrangian quantities from arrival to departure grid:

- Reconstruct $\tilde{\rho}_i$ such that $\rho_i^{\min} \le \tilde{\rho}_i \le \rho_i^{\max}$
- Reconstruct $\tilde{q}_i$ such that $q_i^{\min} \le \tilde{q}_i \le q_i^{\max}$

$\longrightarrow$ **Lagrangian quantities**

$$\begin{cases} \tilde{m}_i = \int_{\tilde{C}_i} \tilde{\rho}_i\, dx \\ \tilde{Q}_i = \int_{\tilde{C}_i} \tilde{\rho}_i \tilde{q}_i\, dx \end{cases}$$

Step 3: **Update** values on the Eulerian (arrival) grid $\tilde{C}(\Omega(t))$

$$m_i(t+\Delta t) = \tilde{m}_i \longrightarrow \rho_i = \frac{\tilde{m}_i}{\mu_i}$$

$$Q_i(t+\Delta t) = \tilde{Q}_i \longrightarrow q_i = \frac{\tilde{Q}_i}{m_i}$$

Optimization-based finite volume SL scheme

Sandia National Laboratories

Step 1: **Trace back** cell vertices to find the Lagrangian (departure) grid $\tilde{C}(\Omega(t))$

Step 2: Remap Lagrangian quantities from arrival to departure grid:

- Reconstruct $\tilde{\rho}_i$ without applying bounds
- Reconstruct $\tilde{q}_i$ without applying bounds

$\Big\}$ → **Lagrangian targets**

$$\begin{cases} \tilde{m}_i^T = \int_{\tilde{C}_i} \tilde{\rho}_i \, dx \\ \tilde{Q}_i^T = \int_{\tilde{C}_i} \tilde{\rho}_i \tilde{q}_i \, dx \end{cases}$$

- **Solve two quadratic programs (QP) for the Lagrangian quantities:**

$$\min_{\tilde{m}_i} \sum_{C_i} \left(\tilde{m}_i - \tilde{m}_i^T \right)^2 \quad \text{subject to}$$
$$\sum_{C_i} \tilde{m}_i = M; \quad \text{and} \quad m_i^{\min} \le \tilde{m}_i \le m_i^{\max}$$

$$\min_{\tilde{Q}_i} \sum_{C_i} \left(\tilde{Q}_i - \tilde{Q}_i^T \right)^2 \quad \text{subject to}$$
$$\sum_{C_i} \tilde{Q}_i = Q; \quad \text{and} \quad Q_i^{\min} \le \tilde{Q}_i \le Q_i^{\max}$$

Step 3: **Update** values on the Eulerian (arrival) grid $\tilde{C}(\Omega(t))$

$$m_i(t+\Delta t) = \tilde{m}_i \longrightarrow \rho_i = \frac{\tilde{m}_i}{\mu_i}$$

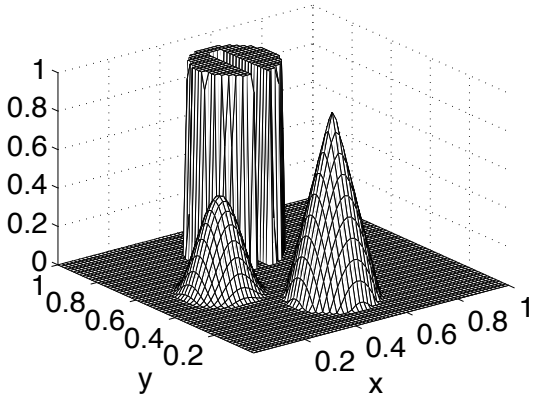
$$Q_i(t+\Delta t) = \tilde{Q}_i \longrightarrow q_i = \frac{\tilde{Q}_i}{m_i}$$

Advantages

- The solution is a **globally optimal state** that also satisfies the bounds:
  - By definition it is the **best possible solution** satisfying the bounds!
- The solution **provably preserves linear tracer correlations.**
- The two QPs have the exact **same structure** as in the SE-SL case:
  - We have a **fast, scalable optimization algorithm**!
  - Solution times are **essentially the same** as for conventional limiters:

Timings for Leveque's combo example.

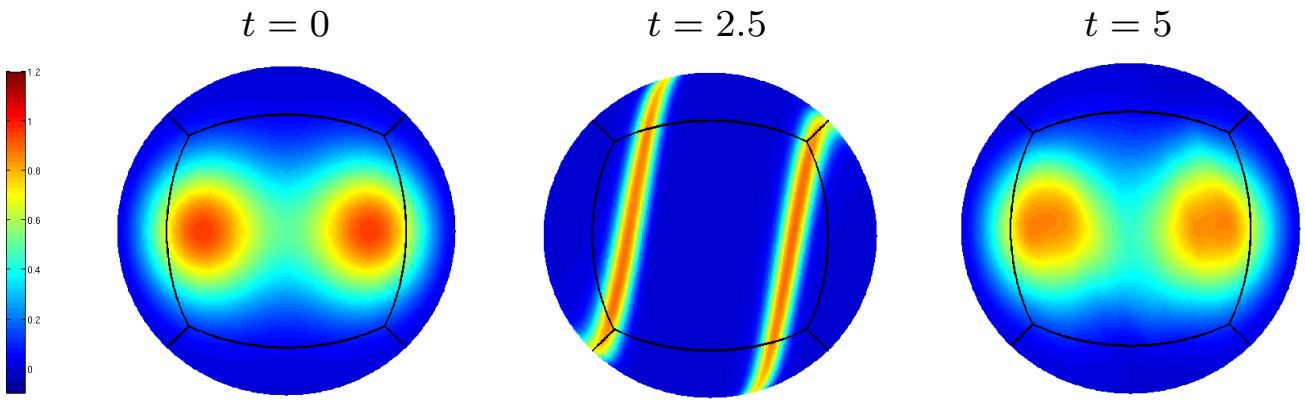
Cells	Time steps	FCT (sec)	Van Leer	OB-SL	Ratio
64x64	400	4.51	4.55	**4.98**	**1.1**
128x128	810	47.60	48.35	**48.78**	**1.0**
256x256	1,610	390.47	399.15	**405.92**	**1.0**
512x512	3,220	5802.05	5804.66	**5655.00**	**0.9**



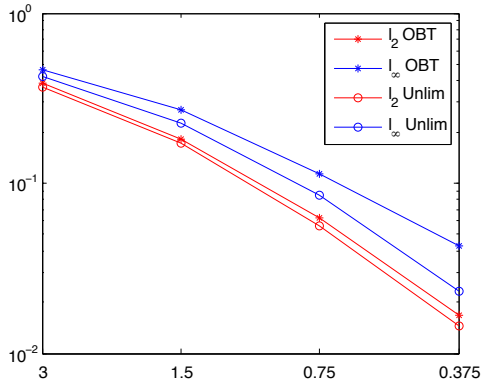
Vectorized Matlab code: **wall-clock times on a 3.06GHz Intel Core Duo MacBook Pro**

Convergence test:

Smooth Gaussian hills on a cubed sphere mesh



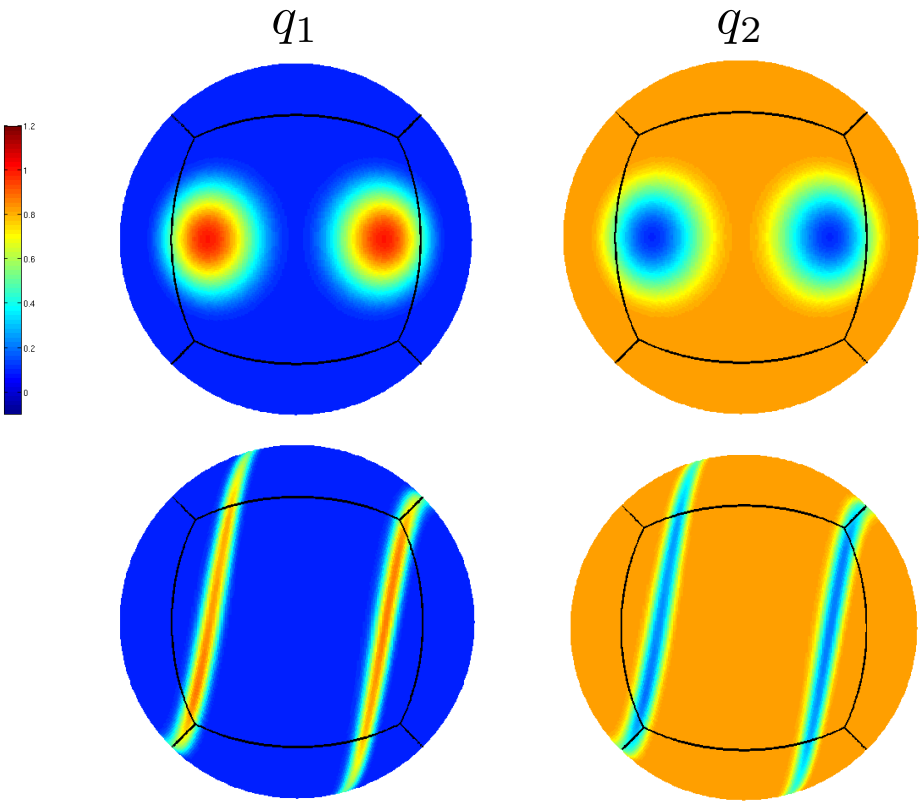
		OBT*		**Unlimited**	
mesh	steps	l_2	l_∞	l_2	l_∞
$3°$	600	0.386	0.465	0.368	0.425
$1.5°$	1200	0.182	0.268	0.172	0.225
$0.75°$	2400	0.0626	0.113	0.0559	0.0843
$0.375°$	4800	0.0167	0.0425	0.0144	0.0233
	Rate	*1.51*	*1.16*	*1.56*	*1.40*



Using optimization to enforce bounds does not lead to degradation of accuracy!

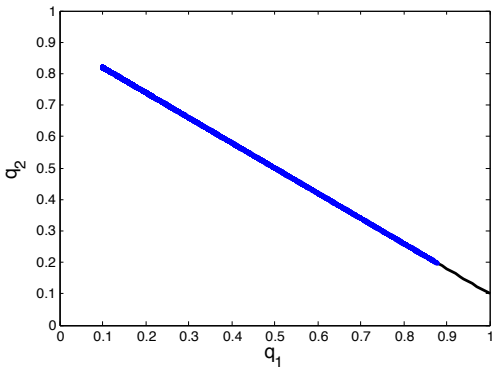
Linear tracer correlations

Initial tracer distributions: two linearly correlated cosine bells



$$q_2 = -0.8 q_1 + 0.9$$

Correlation $t = 2.5$



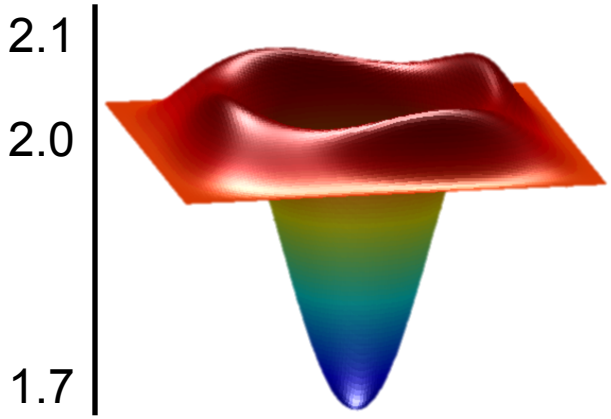
Optimization formulation provably preserves linear tracer correlations

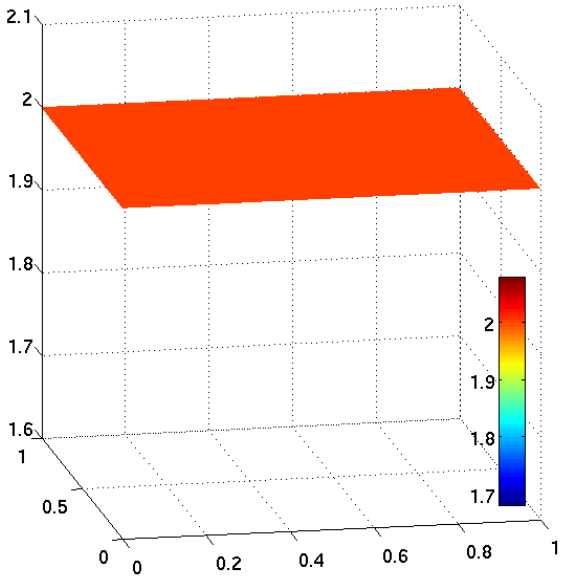
Sandia National Laboratories

Part 3

However, …

Do you think there's anything wrong with this result?



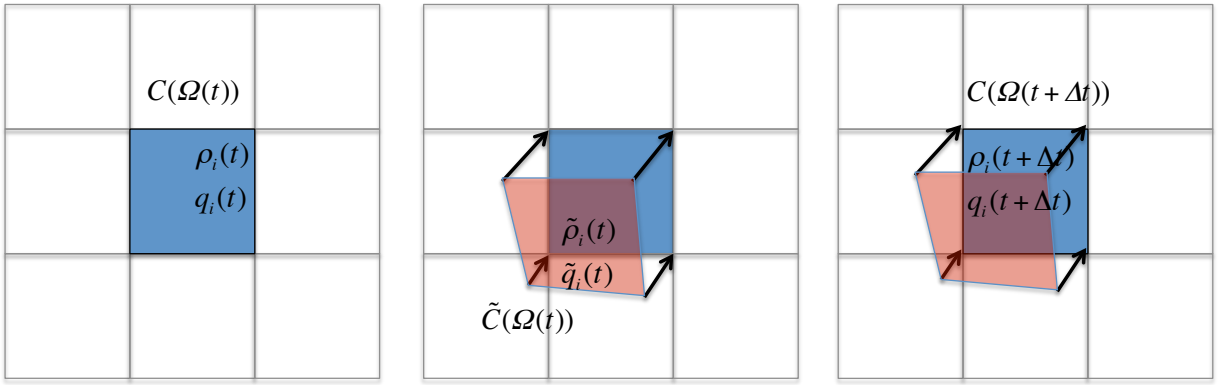


Everything! Exact solution (density) is constant in time!

All we did was switch from RK4 to a forward Euler. Clearly Euler is less accurate but it is still supposed to preserve constant in time functions. So what is causing such a dramatic deterioration in the solution?

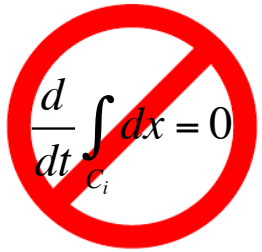
There's another physical property…

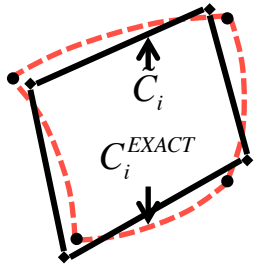
Let's take ρ=*const* and examine what happens during a single time step:



$$\rho_i(t) = \frac{m_i}{\mu_i} = \rho_{const} \qquad \tilde{\rho}_i = \rho_{const} \qquad \tilde{m}_i = \int_{\tilde{C}_i} \tilde{\rho}_i \, dx = \rho_{const} \, \tilde{\mu}_i \qquad m_i(t+\Delta t) = \tilde{m}_i \qquad \rho_i(t+\Delta t) = \frac{\tilde{m}_i}{\mu_i}$$

$$\rho_i(t+\Delta t) = \frac{\tilde{m}_i}{\mu_i} = \frac{\rho_{const}\tilde{\mu}_i}{\mu_i} = \rho_{const}\frac{\tilde{\mu}_i}{\mu_i} \neq \rho_{const}$$





Our departure grid approximates the true Lagrangian grid, hence it violates the property that **non-divergent Lagrangian flows preserve volumes!**

The Geometric Conservation Law

Sandia National Laboratories

Our scheme violates the Geometric Conservation Law (GCL), which is critical for methods involving any kind of moving grids:

$$\frac{d}{dt}\int_{C_i(t)} dx = \int_{\partial C_i(t)} \mathbf{u}\cdot\mathbf{n}\, ds$$

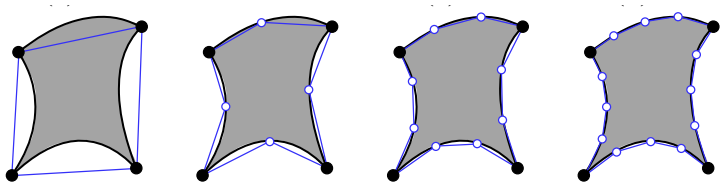
Thomas, Lombardi, AIAA 17, 1979

Some recent work on GCL:

Use more Lagrangian points.

- **Enforces GCL approximately.**

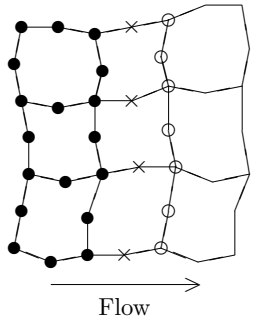
Lauritzen, Nair, Ullrich, A conservative semi-Lagrangian multi-tracer transport scheme on the cubed-sphere grid, JCP 229/5 (2010)



Heuristic mesh adjustment procedure:

- **No theoretical assurance of completion.**

Arbogast, Huang, A fully mass and volume conserving implementation of a characteristic method for transport problems, SISC 28 (6) (2006).



- ● Adjusted point to remain fixed at this stage.
- ○ Points adjusted simultaneously in the direction of the characteristic.
- × Points adjusted "sideways" to the flow.

Monge-Ampere trajectory correction

- **Requires** nontrivial solution of the nonlinear MAE
- **Approximate:** GCL ≈ accuracy of MAE scheme

Cossette, Smolarkiewicz, Charbonneau, The Monge–Ampere trajectory correction for semi-Lagrangian schemes, JCP, (2014) –

Correct departure points according to

$$\tilde{\mathbf{p}}_{ij}^{corr} = \tilde{\mathbf{p}}_{ij} + (t - t_n)\nabla\phi; \quad \det\frac{\partial \mathbf{p}_{ij}^{corr}}{\partial x} = 1$$

An optimization solution to the GCL

Sandia National Laboratories

Statement of the volume correction problem

Given: a source mesh $\tilde{C}(\Omega)$ and $\mathbf{c}_o \in \mathbf{R}^m$ such that $\sum_{C_i} c_{0,i} = |\Omega|$ and $c_{0,i} \geq 0 \quad \forall i$

Find: a volume compliant mesh $C(\Omega)$ such that:

a) $C(\Omega)$ has the same connectivity as the source mesh

b) The volumes of its cells match the volumes prescribed in $\boldsymbol{c}_0$

c) Every cell $C_i \in C(\Omega)$ is valid; or convex

d) Boundary points in $C(\Omega)$ correspond to boundary points in $\tilde{C}(\Omega)$

- **The volume correction problem may or may not have a solution!**
- **An important setting in which solution always exist is when**

The source mesh $\tilde{C}(\Omega)$ is transformation of another mesh $\breve{C}(\Omega)$ such that:

$$\forall \breve{C}_i \in \breve{C}(\Omega) \quad \text{is valid, or convex and} \quad |\breve{C}_i| = c_{0,i}$$

In this case $C(\Omega) = \breve{C}(\Omega)$ is a trivial solution of the volume correction problem

Volume correction as an optimization problem

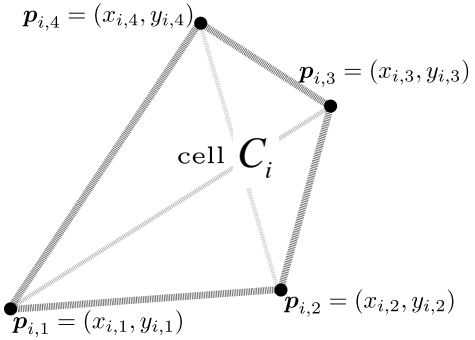
We consider quads (simplices are actually easier). We need few things:

Oriented volume of a quad cell:

$$\forall C_i \in C(\Omega), \quad |C_i| = \frac{1}{2}\left((x_{i,1} - x_{i,3})(y_{i,2} - y_{i,4}) + (x_{i,2} - x_{i,4})(y_{i,3} - y_{i,1})\right)$$

Partitioning of a quad into triangles:

$$T_{i,r} \in C_i, \quad T_{i,r} = (p_{a_r}, p_{b_r}, p_{b_r}) \qquad (a_r, b_r, c_r) = \begin{cases} (1,2,4) & r = 1 \\ (2,3,4) & r = 2 \\ (1,3,4) & r = 3 \\ (1,2,3) & r = 4 \end{cases}$$



Oriented volume of a triangle

$$T_{i,r} \in C_i, \quad |T_{i,r}| = \frac{1}{2}\left(x_{i,a_r}(y_{i,c_r} - y_{i,b_r}) - x_{i,b_r}(y_{i,a_r} - y_{i,c_r}) - x_{i,c_r}(y_{i,b_r} - y_{i,a_r})\right)$$

Convexity indicator for a quad cell:

C_i is convex, if the oriented areas of all its triangles are positive: $\forall T_{i,r} \in C_i, \ |T_{i,r}| > 0$

Volume correction as an optimization problem

Sandia National Laboratories

Optimization objective:

Mesh distance → $J_0(p,\tilde{p}) = \frac{1}{2} d(C(\Omega), \tilde{C}(\Omega))^2 = \left| p - \tilde{p} \right|_{\ell^2}^2$

Optimization constraints:

① **Volume equality** → $\forall C_i \in C(\Omega), \quad |C_i| = c_{0,i}$

② **Cell convexity** → $\forall C_i \in C(\Omega), \ \forall T_{i,r} \in C_i, \ |T_{i,r}| > 0$

③ **Boundary compliance** → $\forall p_j \in \partial\Omega, \quad \gamma(p_j) = 0$

Nonlinear programming problem (NLP)

$$p^* = \arg\min \left\{ J_0(p,\tilde{p}) \ \text{ subject to } \ (1), (2), \ \text{and } (3) \right\}$$

A simplified NLP formulation

Sandia National Laboratories

Consider a polygonal domain:

- Boundary compliance on polygonal Ω can be subsumed in the volume constraint
- Convexity can be enforced weakly by logarithmic barrier functions
- This leaves only the equality volume constraint and gives the simplified NLP:

$$p^* = \operatorname{argmin}\left\{ J(p) \;\text{ subject to }\; |C_i| = c_{0,i}\ \forall i \right\} \qquad J(p) = J_0(p) - \beta \sum_{C_i} \sum_{T_{i,r} \in C_i} \log |T_{i,r}|$$

Specialization to simplicial cells

A valid simplex is always convex ⟶ A simplex is valid if and only if $|C_i| > 0$

Since $c_{0,i} > 0$, the volume equality constraint $\forall C_i \in C(\Omega),\ |C_i| = c_{0,i}$ implies $|C_i| > 0$!

$$p^* = \operatorname{argmin}\left\{ J_0(p) \;\text{ subject to }\; |C_i| = c_{0,i}\ \forall i \right\}$$

A scalable optimization algorithm

Sandia National Laboratories

Based on the inexact trust region sequential quadratic programming (SQP) method of Ridzal and Heinkenschloss. Key properties of the inexact SQP approach:

- Fast local convergence, based on its relationship to Newton's method,
- Use of 'inexact' solvers enables an efficient solution of very large NLP.
- Key requirement in the method: **design of an efficient preconditioner**.

Given an optimization iterate p^k all linear systems involved are of the form

$$\begin{pmatrix} I & \nabla C(p^k)^T \\ \nabla C(p^k) & 0 \end{pmatrix} \begin{pmatrix} v^1 \\ v^v \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix}$$

$C(p^k)$ - polynomial matrix function of coordinates

Preconditioner

$$\pi^k = \begin{pmatrix} I & 0 \\ 0 & \left(\nabla C(p^k)\nabla C(p^k)^T + \varepsilon I\right)^{-1} \end{pmatrix}$$

- $\varepsilon > 0$ small parameter ~ $10^{-8}h$
- $\nabla C(p^k)\nabla C(p^k)^T + \varepsilon I$ formed explicitly
- Inverse: smoothed aggregation AMG – Trilinos

If $\nabla C(p^k)$ is full rank, preconditioned GMRES converges in 3 iterations (Golub et al SISC 21/6, 2000)

Algorithm scalability

Sandia National Laboratories

To challenge the algorithm we test performance as follows:

- Start with a uniform *nxn* mesh and advance to final time using the deformational velocity field.
- Apply algorithm to the deformed mesh at final time setting c_0 to initial mesh volumes.

Analytic action of $\nabla^2 J(p)$ but finite difference $\nabla C(p)$

n	SQP	CG	GMRES	GMRES av.	CPU	%ML
64	3	2	34	**2.3**	0.962	**59**
128	3	2	42	**2.8**	3.551	**75**
256	2	1	30	**3.0**	10.54	**82**
512	3	1	49	**3.5**	87.07	**88**

Analytic action of $\nabla^2 J(p)$

n	SQP	CG	GMRES	GMRES av.	CPU	%ML
64	3	2	28	**1.9**	0.860	**65**
128	3	2	36	**2.4**	3.115	**81**
256	2	1	25	**2.5**	8.787	**85**
512	3	1	43	**3.1**	73.775	**90**

- Almost constant GMRES iterations; average GMRES ~ theoretical bound of 3 (inexact ML solve!)
- The matrix $\nabla C(\cdot)\nabla C(\cdot)^T \approx \Delta$, hence the appropriateness of AMG for the preconditioner
- CPU per SQP iteration scales linearly with problem size & confirms choice of preconditioner
- SQP and inner CG iteration counts to achieve machine precision are mesh independent
- The algorithm inherits its scalability from the AMG solver.

Applications: Lagrangian mesh motion

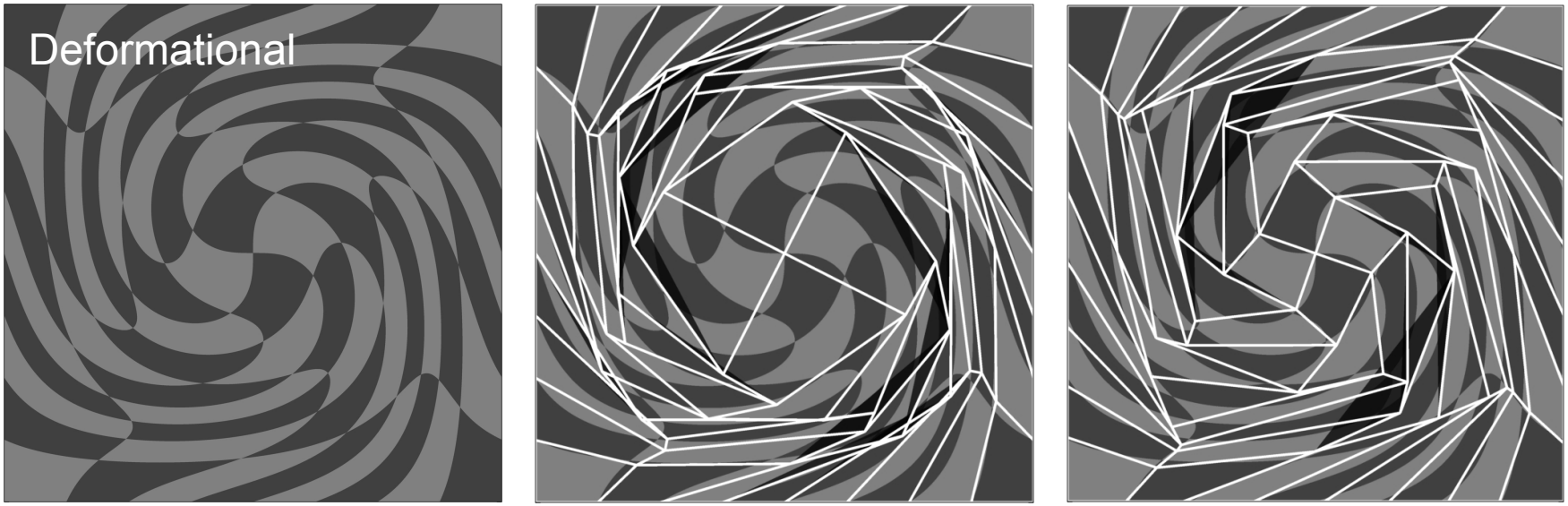
Sandia National Laboratories

Models the evolution of the computational mesh under a non-divergent velocity

$$\mathbf{u}(\mathbf{p},t) = \begin{pmatrix} \sin(\pi x)^2 \sin(2\pi y)\cos(\pi t/T) \\ -\sin(\pi y)^2 \sin(2\pi x)\cos(\pi t/T) \end{pmatrix}$$ ← Deformational

Rotational → $$\mathbf{u}(\mathbf{p},t) = \begin{pmatrix} 0.5 - y \\ 0.5 - x \end{pmatrix}$$

Exact | **Source (uncorrected)** | **Compliant (corrected)**



R. J. LeVeque, High-resolution conservative algorithms for advection in incompressible flow, SINUM 33 (1996)

Improvements in mesh geometry

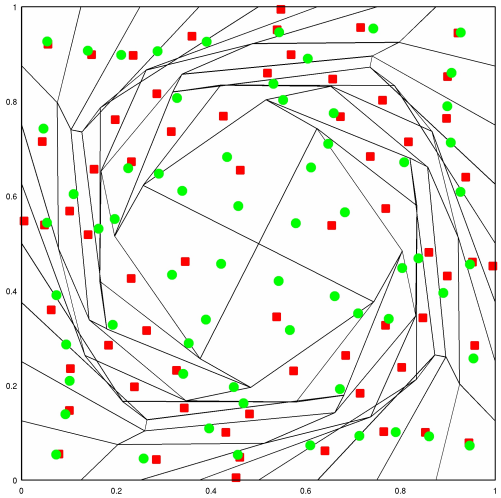
Sandia National Laboratories

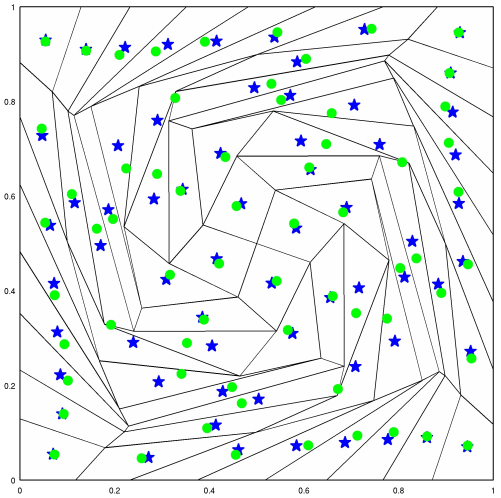
Cell barycenters

★ - exact Lagrangian mesh

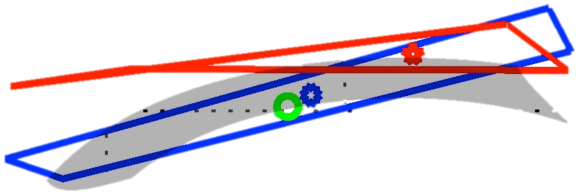
■ - source (uncorrected)

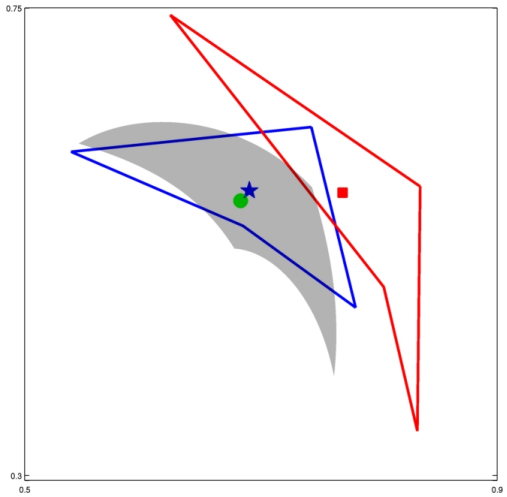
● - compliant (corrected)

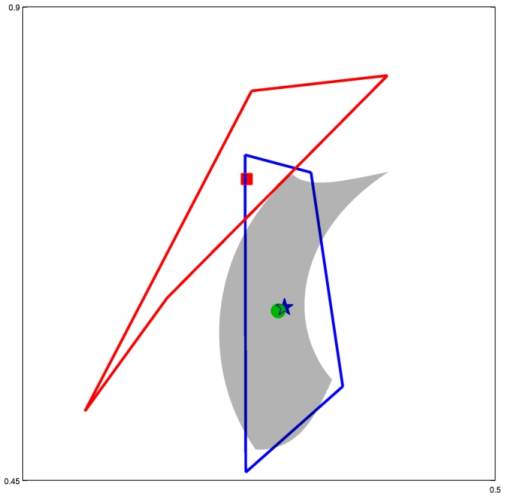




Invalid cell in the source mesh:





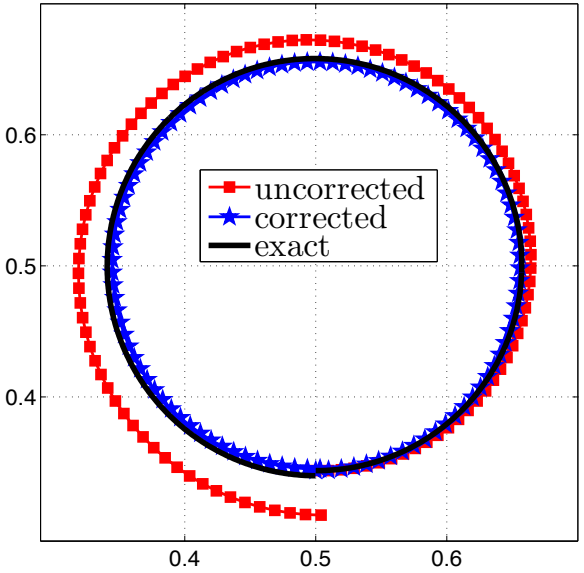


Improvements in mesh geometry

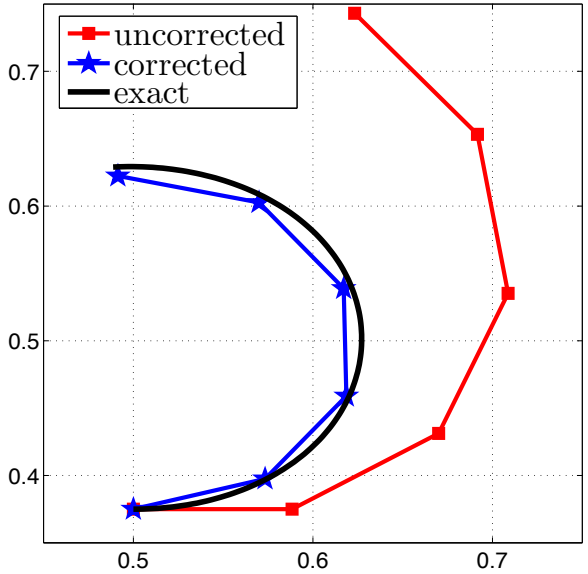
Sandia National Laboratories

Point trajectories

Rotational flow



Deformational flow



We observe significant improvements in the geometry of the corrected mesh:

- The shapes of the corrected cells are close to the exact Lagrangian shapes
- The barycenters of the corrected cells are very close to the exact barycenters
- The trajectories of the corrected points track the exact Lagrangian trajectories very closely

Applications: semi-Lagrangian transport

Sandia National Laboratories

Recall the cell-centered optimization-based semi-Lagrangian scheme:

Step 1: **Trace back** cell vertices to find the Lagrangian (departure) grid $\tilde{C}(\Omega(t))$

Step 2: Optimization-based remap of Lagrangian values from arrival to departure grid.

Step 3: **Update** values on the Eulerian (arrival) grid $\tilde{C}(\Omega(t))$

We modify it to include a volume correction step:

Step 1: **Trace back** cell vertices to find the Lagrangian (departure) grid $\tilde{C}(\Omega(t))$

Step 1$^+$: **Correct the departure grid to match the cell volumes of the arrival grid**

Step 2: Optimization-based remap of Lagrangian values from arrival to departure grid.

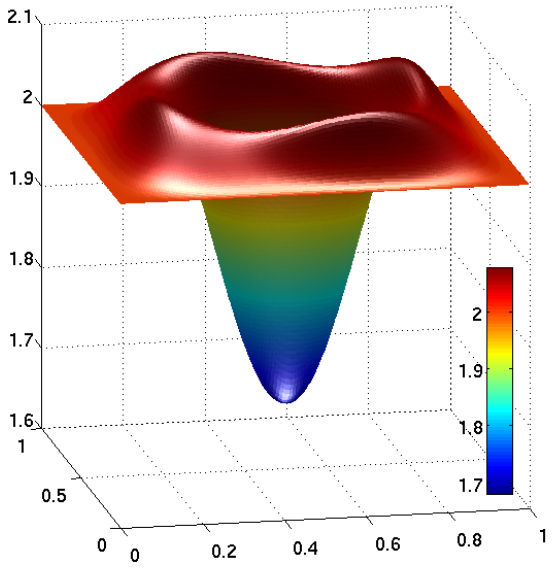
Step 3: **Update** values on the Eulerian (arrival) grid $\tilde{C}(\Omega(t))$

Applications: semi-Lagrangian transport

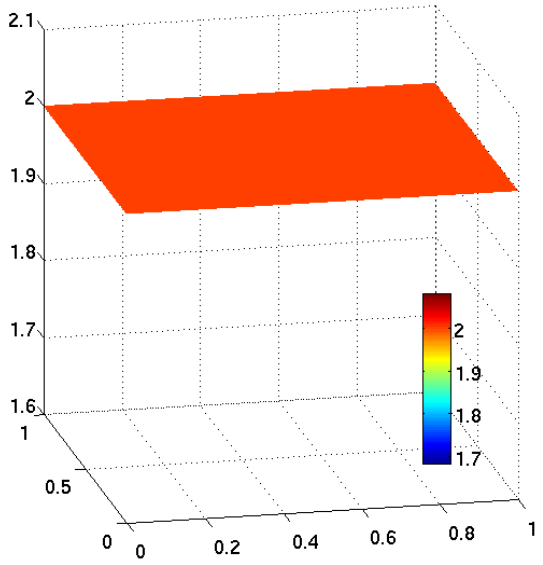
Sandia National Laboratories

Constant in time density: rotational flow

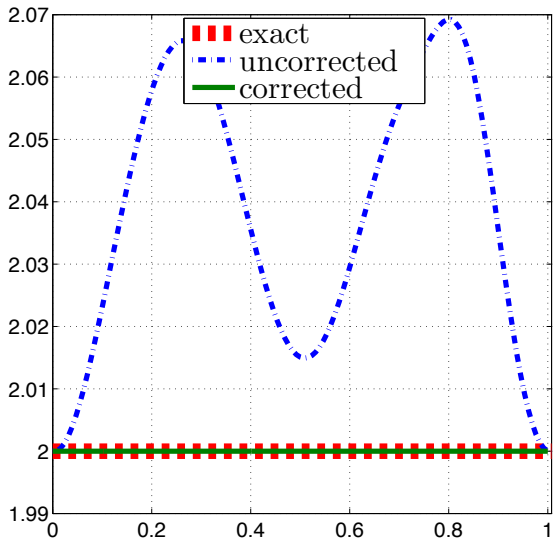
Uncorrected



Corrected



Comparison



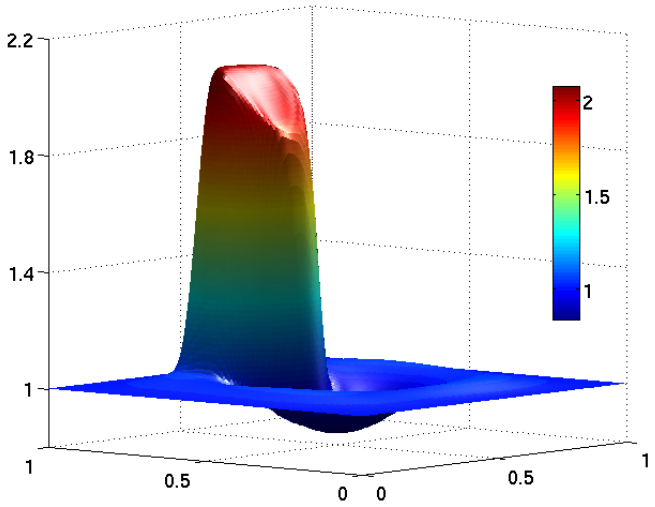
Plots of the density at time $t^N = 1.5$ for Forward Euler simulations with $\Delta t = 0.006$

Applications: semi-Lagrangian transport

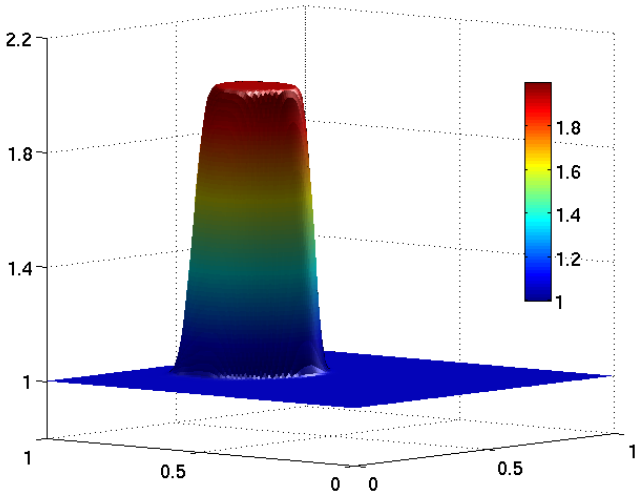
Sandia National Laboratories

Initial cylindrical density distribution: rotational flow

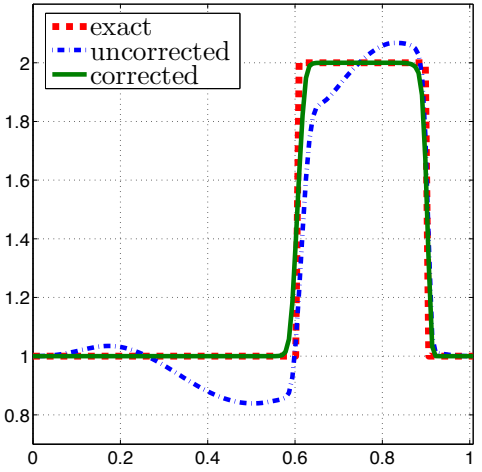
Uncorrected



Corrected



Comparison



Plots of the density at time $t^N = 1.5$ for Forward Euler simulations with $\Delta t = 0.006$

Conclusions

Sandia National Laboratories

Traditional approaches to devise stable and accurate numerical methods are reaching a point of diminishing returns for complex applications involving multiple mathematical models, requiring diverse, heterogeneous numerical methods.

The use of optimization ideas to couple hetergoeneous numerical methods and to preserve the relevant physical properties is very promising.

However, its success depends critically on the availability of efficient and scalable optimization algorithms to solve the resulting QPs and NLPs.

We've presented two examples where such algorithms are available and optimization leads to successful heterogenous numerical methods and property preserving schemes.

Development of fast, scalable optimization algorithms is likely to play the same role for Heterogeneous Numerical Methods as the development of fast, scalable linear solvers did in the past for conventional PDE methods.