Parallel, Robust Geometric Multigrid for Adaptive High-Order Meshes and Highly Heterogeneous, Nonlinear Stokes Flow of Earth’s Mantle

Johann Rudi

Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, USA

Advisor: Omar Ghattas

Main collaborators: Georg Stadler (NYU), Mike Gurnis (Caltech), Tobin Isaac (UT)
Introduction
Introduction to mantle convection & plate tectonics

- Mantle convection is the thermal convection in the Earth’s upper \(~3000\) km
- It controls the thermal and geological evolution of the Earth
- Solid rock in the mantle moves like viscous incompressible fluid on time scales of millions of years

Central open questions:

- Energy dissipation in hinge zones
- Main drivers of plate motion: negative buoyancy forces or convective shear traction
- Role of slab geometries
- Accuracy of rheology extrapolations from experiments
Computational challenges of global-scale mantle flow

- Severe nonlinearity, heterogeneity, and anisotropy of the Earth’s rheology with a wide range of spatial scales
- Highly localized features with respect to Earth’s radius (~6371 km), like plate thickness ~50 km and shearing zones at plate boundaries ~5 km
- 6 orders of magnitude viscosity contrast within ~5 km thin plate boundaries
- Resolution down to ~1 km at plate boundaries (uniform mesh of Earth’s mantle would result in computationally prohibitive $O(10^{12})$ degrees of freedom). Enabled by: adaptive mesh refinement
- Velocity approximation with high accuracy and local mass conservation. Enabled by: high-order discretizations
Mantle convection modeled as nonlinear Stokes flow

Rock in the mantle moves like a viscous, incompressible fluid (over millions of years) and can be modeled as a nonlinear Stokes system:

\[-\nabla \cdot \left[ \mu(T, u) \left( \nabla u + \nabla u^\top \right) \right] + \nabla p = f(T)\]
\[\nabla \cdot u = 0\]

The viscosity \( \mu \) depends exponentially on the temperature, on a power of the second invariant of the strain rate tensor, incorporates plastic yielding and lower/upper bounds:

\[\mu(T, u) = \max\left(\mu_{\min}, \min\left(\frac{\tau_{\text{yield}}}{2\dot{\varepsilon}(u)}, w \min\left(\mu_{\max}, a(T) \dot{\varepsilon}(u)^{\frac{1-n}{n}}\right)\right)\right)\]
Mantle convection modeled as nonlinear Stokes flow

Rock in the mantle moves like a viscous, incompressible fluid (over millions of years) and can be modeled as a nonlinear Stokes system:

\[-\nabla \cdot \left[ \mu(T, \mathbf{u}) \left( \nabla \mathbf{u} + \nabla \mathbf{u}^\top \right) \right] + \nabla p = \mathbf{f}(T)\]

\[\nabla \cdot \mathbf{u} = 0\]

The viscosity \( \mu \) depends exponentially on the temperature, on a power of the second invariant of the strain rate tensor, incorporates plastic yielding and lower/upper bounds:

\[\mu(T, \mathbf{u}) = \max \left( \mu_{\min}, \min \left( \frac{\tau_{\text{yield}}}{2\hat{\varepsilon}(\mathbf{u})}, w \min \left( \mu_{\max}, a(T) \hat{\varepsilon}(\mathbf{u})^{\frac{1-n}{n}} \right) \right) \right)\]

The Newton update \((\tilde{\mathbf{u}}, \tilde{p})\) is computed as the inexact solution of:

\[-\nabla \cdot \left[ \left( \mu \mathbf{I} + \hat{\varepsilon} \frac{\partial \mu}{\partial \hat{\varepsilon}} \frac{(\nabla \mathbf{u} + \nabla \mathbf{u}^\top) \otimes (\nabla \mathbf{u} + \nabla \mathbf{u}^\top)}{\| (\nabla \mathbf{u} + \nabla \mathbf{u}^\top) \|_F^2} \right) (\nabla \tilde{\mathbf{u}} + \nabla \tilde{\mathbf{u}}^\top) \right] + \nabla \tilde{p} = -\mathbf{r}_{\text{mom}}\]

\[\nabla \cdot \tilde{\mathbf{u}} = -\mathbf{r}_{\text{mass}}\]
Methods & Algorithms
Solving the discretized Stokes system

Finite element discretization:

- **High-order, inf-sup stable** velocity-pressure pairings: \( \mathbb{Q}_k \times \mathbb{P}^{\text{disc}}_{k-1} \)
- **Local mass conservation** at the element level, discont. pressure
Solving the discretized Stokes system

Finite element discretization:
- **High-order, inf-sup stable** velocity-pressure pairings: $\mathbb{Q}_k \times \mathbb{P}^{\text{disc}}_{k-1}$
- **Local mass conservation** at the element level, discont. pressure

Coupled iterative solver with upper triangular block preconditioning:

\[
\begin{bmatrix}
A & B^\top \\
B & 0
\end{bmatrix}
\begin{bmatrix}
\tilde{A} & B^\top \\
0 & \tilde{S}
\end{bmatrix}^{-1}
\begin{bmatrix}
u' \\
p'
\end{bmatrix} = 
\begin{bmatrix}
f \\
0
\end{bmatrix}
\]

(Stokes operator \hspace{1cm} preconditioner)

Requires: (i) approx. inverse of the viscous stress block, $\tilde{A}^{-1} \approx A^{-1}$
(ii) approx. inverse of the Schur complement, $\tilde{S}^{-1} \approx (BA^{-1}B^\top)^{-1}$
Solving the discretized Stokes system

Finite element discretization:

- High-order, inf-sup stable velocity-pressure pairings: $Q_k \times P^\text{disc}_{k-1}$
- Local mass conservation at the element level, discont. pressure

Coupled iterative solver with upper triangular block preconditioning:

\[
\begin{bmatrix}
A & B^T \\
B & 0
\end{bmatrix}
\begin{bmatrix}
\tilde{A} & B^T \\
0 & \tilde{S}
\end{bmatrix}^{-1}
\begin{bmatrix}
u' \\
p'
\end{bmatrix}
= 
\begin{bmatrix}
f \\
0
\end{bmatrix}
\]

Stokes operator
preconditioner

Requires: (i) approx. inverse of the viscous stress block, $\tilde{A}^{-1} \approx A^{-1}$  
(ii) approx. inverse of the Schur complement, $\tilde{S}^{-1} \approx (BA^{-1}B^T)^{-1}$

\textbf{BFBT / Least Squares Commutator (LSC) method:}

\[
\tilde{S}^{-1} = (BD^{-1}B^T)^{-1}(BD^{-1}AD^{-1}B^T)(BD^{-1}B^T)^{-1}
\]

with diagonal scaling, $D^{-1} := \text{diag}(A)^{-1}$. 
Comparison to state of the art for unstructured meshes

Solve $Au = f$

Solve Stokes system
Review: BFBT/LSC methods for Schur complement $\tilde{S}^{-1}$

BFBT method [Elman, 1999]: pseudoinverse

$$\tilde{S}^{-1} = (B A^{-1} B^T)^+ = (B B^T)^{-1} (B A B^T) (B B^T)^{-1}$$

Least Squares Commutators (LSC) [Elman, et al., 2006]:
Find commutator matrix $X$ s.t. $(A B^T - B^T X) \approx 0$, by solving the least squares problem:

Find columns $x_j$ of $X$ s.t. $\min_{x_j} \| [A B^T]_j - B^T x_j \|_2^2$

$$\Rightarrow X = (B B^T)^{-1} (B A B^T)$$

$(A B^T - B^T X) \approx 0 \Rightarrow (B A^{-1} B^T)^{-1} \approx (B B^T)^{-1} (B A B^T) (B B^T)^{-1}$

LSC gives same result for $\tilde{S}^{-1}$ as pseudoinverse.

Q: Does this work for FE discretizations?...
Review: BFBT/LSC methods for Schur complement $\tilde{S}^{-1}$

**BFBT method** [Elman, 1999]: pseudoinverse

$$\tilde{S}^{-1} = (BA^{-1}B^T)^+ = (BB^T)^{-1}(BAB^T)(BB^T)^{-1}$$

**Least Squares Commutators (LSC)** [Elman, et al., 2006]:
Find commutator matrix $X$ s.t. $(AB^T - B^TX) \approx 0$, by solving the least squares problem:

Find columns $x_j$ of $X$ s.t. $\min_{x_j} \| [AB^T]_j - B^T x_j \|_2^2$

$$\Rightarrow \quad X = (BB^T)^{-1}(BAB^T)$$

$$(AB^T - B^TX) \approx 0 \Rightarrow (BA^{-1}B^T)^{-1} \approx (BB^T)^{-1}(BAB^T)(BB^T)^{-1}$$

LSC gives same result for $\tilde{S}^{-1}$ as pseudoinverse.

Q: Does this work for FE discretizations?... no
Review: BFBT/LSC methods for Schur complement $\tilde{S}^{-1}$

**Diagonally scaled BFBT method** [Elman, et al., 2006]:

Find columns $x_j$ of $X$ s.t. $\min_{x_j} \left\| M_1^{-1/2} [AM_2^{-1} B^\top]_j - M_1^{-1/2} B^\top x_j \right\|_2^2$

$$\Rightarrow X = (BM_1^{-1} B^\top)^{-1} (BM_1^{-1} AM_2^{-1} B^\top)$$

$$\Rightarrow \tilde{S}^{-1} = (BM_1^{-1} B^\top)^{-1} (BM_1^{-1} AM_2^{-1} B^\top) (BM_2^{-1} B^\top)^{-1}$$

Proposed scaling: For FE, use “diagonalized” velocity mass matrix,

- diagonal: $M_1 = M_2 = \text{diag}(M_u)$
- lumped: $M_1 = M_2 = \tilde{M}_u$

Since $BM_1^{-1} B^\top$ can be understood as a Laplace operator for the pressure, approximate $(BM_1^{-1} B^\top)^{-1}$ by a multigrid V-cycle.

Q: Is mass scaled BFBT effective for high viscosity variations?...
Review: BFBT/LSC methods for Schur complement $\tilde{S}^{-1}$

**Diagonally scaled BFBT method** [Elman, et al., 2006]:

Find columns $x_j$ of $X$ s.t.

$$\min_{x_j} \left\| M_1^{-1/2} [A M_2^{-1} B^\top]_j - M_1^{-1/2} B^\top x_j \right\|_2^2$$

$$\Rightarrow X = (B M_1^{-1} B^\top)^{-1} (B M_1^{-1} A M_2^{-1} B^\top)$$

$$\Rightarrow \tilde{S}^{-1} = (B M_1^{-1} B^\top)^{-1} (B M_1^{-1} A M_2^{-1} B^\top)(B M_2^{-1} B^\top)^{-1}$$

Proposed scaling: For FE, use “diagonalized” velocity mass matrix,

- diagonal: $M_1 = M_2 = \text{diag}(M_u)$
- lumped: $M_1 = M_2 = \tilde{M}_u$

Since $B M_1^{-1} B^\top$ can be understood as a Laplace operator for the pressure, approximate $(B M_1^{-1} B^\top)^{-1}$ by a multigrid V-cycle.

**Q:** Is mass scaled BFBT effective for high viscosity variations?... no
Review: BFBT/LSC methods for Schur complement $\tilde{S}^{-1}$

**BFBT for scaled Stokes systems** that arise in geodynamics [May, Moresi, 2008]:

$$\begin{bmatrix}
D_u^{-1/2} & 0 \\
0 & D_p^{-1/2}
\end{bmatrix}
\begin{bmatrix}
A & B^\top \\
B & 0
\end{bmatrix}
\begin{bmatrix}
D_u^{-1/2} & 0 \\
0 & D_p^{-1/2}
\end{bmatrix}$$

Then the standard BFBT method yields its scaled version,

$$\Rightarrow \quad \tilde{S}^{-1} = (BD_u^{-1}B^\top)^{-1}(BD_u^{-1}AD_u^{-1}B^\top)(BD_u^{-1}B^\top)^{-1}$$

Proposed scaling: heuristic, motivated by scaling of dimensional systems

$$[D_u]_{i,i} = \max_j |[A]_{i,j}|$$

**Q:** Is BFBT with this scaling effective for high viscosity variations?...
Review: BFBT/LSC methods for Schur complement $\tilde{S}^{-1}$

BFBT for scaled Stokes systems that arise in geodynamics [May, Moresi, 2008]:

$$
\begin{bmatrix}
D_u^{-1/2} & 0 \\
0 & D_p^{-1/2}
\end{bmatrix}
\begin{bmatrix}
A & B^\top \\
B & 0
\end{bmatrix}
\begin{bmatrix}
D_u^{-1/2} & 0 \\
0 & D_p^{-1/2}
\end{bmatrix}
$$

Then the standard BFBT method yields its scaled version,

$$
\Rightarrow \quad \tilde{S}^{-1} = (BD_u^{-1}B^\top)^{-1}(BD_u^{-1}AD_u^{-1}B^\top)(BD_u^{-1}B^\top)^{-1}
$$

Proposed scaling: heuristic, motivated by scaling of dimensional systems

$$
[D_u]_{i,i} = \max_j |[A]_{i,j}|
$$

Q: Is BFBT with this scaling effective for high viscosity variations?...yes
New view on BFBT/LSC methods (1)

Let $C$ be symm. pos. def. and let $D$ be arbitrary,

Find $X$ s.t. $\min_X \left\| AD^{-1}B^T e_j - B^T X e_j \right\|_{C^{-1}}^2$ for all $j$

$\Rightarrow X = \left( BC^{-1}B^T \right)^{-1} \left( BC^{-1}AD^{-1}B^T \right)$

And we have a $C^{-1}$-orthogonal projection, i.e., the residual satisfies

$\left\langle B^T e_i, (AD^{-1}B^T - B^T X) e_j \right\rangle_{C^{-1}} = 0$ for all $i, j$,

therefore

$\left( AD^{-1}B^T - B^T X \right) e_j \perp_{C^{-1}} \text{Ran}(B^T)$ for all $j$
New view on BFBT/LSC methods (2)

**Goal:** Effective and robust preconditioning of the Schur complement in Stokes systems with high viscosity variations

**Note:** Condition for optimal preconditioning \((B\tilde{A}^{-1}B^\top)\tilde{S}^{-1} = I\).

By choosing \(C = \tilde{A}\), we obtain equivalence between orthogonality and the condition for optimal preconditioning:

\[
\left\langle B^\top e_i, (AD^{-1}B^\top - B^\top X)e_j \right\rangle_{\tilde{A}^{-1}} = 0 \quad \forall i, j \quad \iff \quad \tilde{S} = B\tilde{A}^{-1}B^\top
\]
New view on BFBT/LSC methods (2)

**Goal:** Effective and robust preconditioning of the Schur complement in Stokes systems with high viscosity variations

**Note:** Condition for optimal preconditioning \((B\tilde{A}^{-1}B^\top)\tilde{S}^{-1} = I\).

By choosing \(C = \tilde{A}\), we obtain equivalence between orthogonality and the condition for optimal preconditioning:

\[
\left\langle B^\top e_i, (AD^{-1}B^\top - B^\top X)e_j \right\rangle_{\tilde{A}^{-1}} = 0 \quad \forall i, j \iff \tilde{S} = B\tilde{A}^{-1}B^\top
\]

Choices of \(C, D\) that are computationally feasible are limited. Our choice: \(C = D := \text{diag}(A)\), thus

\[
\tilde{S}^{-1} = (BD^{-1}B^\top)^{-1}(BD^{-1}AD^{-1}B^\top)(BD^{-1}B^\top)^{-1}
\]
Parallel geometric multigrid (GMG)

MG hierarchy: Viscous Stress

- $p$-coarsening
- Geometric $h$-coarsening
- Algebraic coarsening

MG hierarchy: Pressure Poisson

- $p$-GMG
- $h$-GMG
- AMG
- Direct

- Discontinuous modal pressure space
- Continuous nodal function space
- Geometric $h$-coarsening
- Algebraic coarsening

- High-order F.E.
- Trilinear F.E.
- Decreasing #cores
- Small #cores and small MPI communicator
- Single core

- Accurate high-order $L^2$-projection operators for restriction and interpolation during V-cycles, and for coarsening of the viscosity
- Coarsening of full fourth-order tensor coefficient of Jacobian
- Chebyshev accelerated point-Jacobi smoothers
- Velocity null spaces are projected out to establish stable convergence
\( h \)-dependence of GMG components & Stokes precond.

Solve \( Au = f \)

\( \text{Solve } (BD^{-1}B^T) \ p = g \)

Solve Stokes system
\(p\)-dependence of GMG components & Stokes precond.

Solve \(Au = f\)

\[
\begin{array}{c}
0 & 50 & 100 & 150 & 200 & 250 \\
10^{-6} & 10^{-4} & 10^{-2} & 10^0
\end{array}
\]

Solve \((BD^{-1}B^T)p = g\)

\[
\begin{array}{c}
0 & 50 & 100 & 150 & 200 & 250 \\
10^{-6} & 10^{-4} & 10^{-2} & 10^0
\end{array}
\]

Solve Stokes system

\[
\begin{array}{c}
0 & 50 & 100 & 150 & 200 & 250 \\
10^{-6} & 10^{-4} & 10^{-2} & 10^0
\end{array}
\]
Computational results using real Earth data
### Computational results: Solver Robustness

#### Robustness of linear Stokes solver w.r.t. plate boundary thickness

<table>
<thead>
<tr>
<th>Plate boundary thickness (km)</th>
<th>DOF for solving $Au = f$</th>
<th>GMRES #iter for solving Stokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1.16B</td>
<td>115</td>
</tr>
<tr>
<td>10</td>
<td>1.41B</td>
<td>129</td>
</tr>
<tr>
<td>5</td>
<td>3.01B</td>
<td>123</td>
</tr>
</tbody>
</table>

#### Robustness of inexact Newton-Krylov nonlinear solver w.r.t plate boundary thickness

<table>
<thead>
<tr>
<th>Plate boundary thickness (km)</th>
<th>DOF (at nl. solution)</th>
<th>Newton #steps</th>
<th>GMRES #iter (for all Newton steps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1.00B</td>
<td>25</td>
<td>3915</td>
</tr>
<tr>
<td>10</td>
<td>1.63B</td>
<td>27</td>
<td>4645</td>
</tr>
<tr>
<td>5</td>
<td>4.78B</td>
<td>29</td>
<td>6112</td>
</tr>
</tbody>
</table>
Computational results: Algorithmic scalability

(Fix problem parameters and refine the mesh)

Algorithmic scalability of linear and nonlinear solver

Degrees of freedom

Billions of DOF / GMRES iter

10^-4

10^-3

10^-2

10^-1

1B

2B

4B

ideal

lin. solver

nl. solver
Computational results: Weak scalability

(Increase core count and problem size simultaneously)
Computational results: Strong scalability
(Increase core count while problem size stays fixed)
Computational results: Contribution to science

Effective viscosity at nonlinear solution and surface velocity

Normal stress at the surface and surface velocity
Summary of key contributions

- Parallel geometric multigrid for the viscous stress block on adaptive meshes (my impl.; based on AMG-only solver and parallel AMR library)

- $p$- and $h$-independent multigrid convergence through improvement of projection operators (my dev.)

- Stable convergence in presence of rotation null spaces (my dev.)

- Stable coarsening of anisotropic fourth-order tensor coefficient in Jacobian (my dev.)

- Geometric multigrid based BFBT; first matrix-free implementation of BFBT (my dev.)

- Inexact Newton-Krylov for complex Earth rheology with dynamic mesh refinement (my impl.; ideas in collaboration with G. Stadler)

- Parallel optimizations, e.g., MPI communicator reduction, OpenMP threading (my impl.)

- Parallel scalability on full system JUQUEEN supercomputer with 458,752 CPU cores (lead dev. in collaboration with IBM Research – Zürich)