Parallel High-Order Geometric Multigrid Methods on Adaptive Meshes for Highly Heterogeneous Nonlinear Stokes Flow Simulations of Earth's Mantle

Johann Rudi[1], Hari Sundar[2], Tobin Isaac[1], Georg Stadler[3], Michael Gurnis[4], and Omar Ghattas[1,5]

[1]Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, USA
[2]School of Computing, The University of Utah, USA
[3]Courant Institute of Mathematical Sciences, New York University, USA
[4]Seismological Laboratory, California Institute of Technology, USA
[5]Jackson School of Geosciences and Department of Mechanical Engineering, The University of Texas at Austin, USA

Summary of main results

I. Efficient methods/algorithms

- High-order finite elements
- Adaptive meshes, resolving viscosity variations
- Geometric multigrid (GMG) preconditioners for elliptic operators
- Novel GMG based BFBT/LSC pressure Schur complement preconditioner
- Inexact Newton-Krylov method
- H^{-1}-norm for velocity residual in Newton line search

II. Scalable parallel implementation

- Matrix-free stiffness/mass application and GMG smoothing
- Tensor product structure of finite element shape functions
- Octree algorithms for handling adaptive meshes in parallel
- Algebraic multigrid (AMG) as coarse solver for GMG avoids full AMG setup cost and large matrix assembly
- Parallel scalability results up to 16,384 CPU cores (MPI)

1. Mantle flow

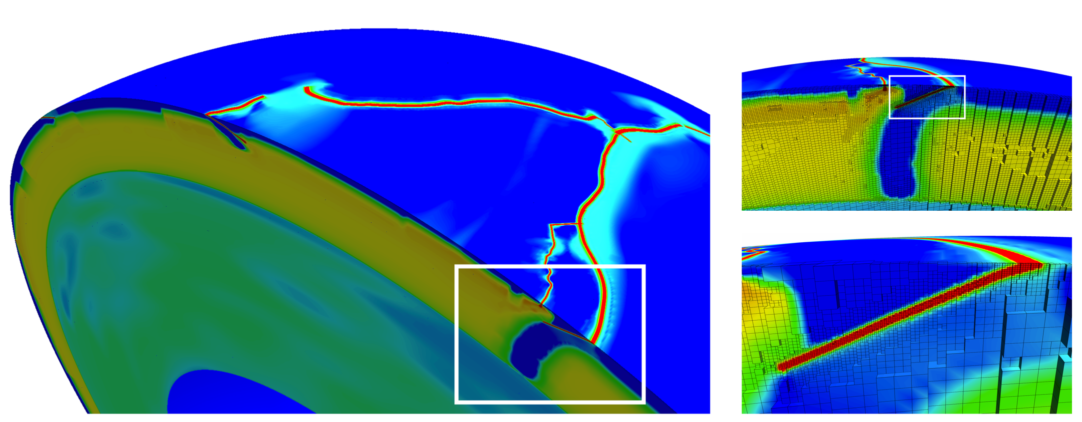
Model equations for Earth mantle flow

Rock in the mantle moves like a viscous, incompressible fluid on time scales of millions of years. From conservation of mass and momentum, we obtain that the flow velocity can be modeled as a nonlinear Stokes system.

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

$\boldsymbol{u}$ ... velocity
p ... pressure
T ... temperature
μ ... viscosity

The viscosity depends exponentially on the temperature (via an Arrhenius relationship), on a power of the second invariant of the strain rate tensor, and incorporates plastic yielding. The right-hand side forcing, $\boldsymbol{f}$, is derived from the Boussinesq approximation and depends on the temperature.



Effective viscosity field and adaptive mesh resolving narrow plate boundaries (shown in *red*).
(Visualization by L. Alisic)

Solver challenges

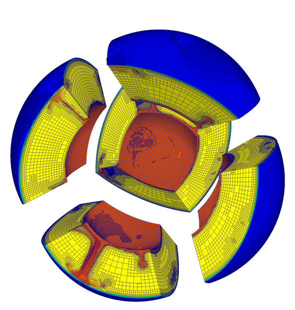
What causes the demand for scalable solvers for high-order discretizations on adaptive grids? — The severe nonlinearity, heterogeneity, and anisotropy of the Earth's rheology:

- Up to 6 orders of magnitude viscosity contrast; sharp viscosity gradients due to decoupling at plate boundaries
- Wide range of spatial scales and highly localized features with respect to Earth radius ($\sim$6371 km): plate thickness $\sim$50 km & shearing zones at plate boundaries $\sim$5 km
- Desired resolution of $\sim$1 km results in $O(10^{12})$ degrees of freedom on a uniform mesh of Earth's mantle, therefore adaptive mesh refinement is essential
- Demand for high accuracy leads to high-order discretizations

2. Scalable parallel Stokes solver

Parallel octree-based adaptive mesh refinement (p4est)

- Identify octree leaves with hexahedral elements
- Octree structure enables fast parallel adaptive octree/mesh refinement and coarsening
- Octrees and space filling curves enable fast neighbor search, repartitioning, and 2:1 balancing in parallel
- Algebraic constraints on non-conforming element faces with hanging nodes enforce global continuity of the velocity basis functions
- Demonstrated scalability to O(500K) cores (MPI)



High-order finite element discretization of the Stokes system

$$\left\{\begin{aligned}-\nabla\cdot\left[\mu\left(\nabla\boldsymbol{u}+\nabla\boldsymbol{u}^{\top}\right)\right]+\nabla p&=\boldsymbol{f}\\ \nabla\cdot\boldsymbol{u}&=0\end{aligned}\right.\ \xrightarrow[\text{high-order FE}]{\text{discretize with}}\ \begin{bmatrix}\mathrm{A}&\mathrm{B}^{\top}\\ \mathrm{B}&0\end{bmatrix}\begin{bmatrix}\mathrm{u}\\ \mathrm{p}\end{bmatrix}=\begin{bmatrix}\mathrm{f}\\ 0\end{bmatrix}$$

- High-order finite element shape functions
- Inf-sup stable velocity-pressure pairings: $\mathbb{Q}_k\times\mathbb{P}_{k-1}^{\mathrm{disc}}$ with $2\le k$
- Locally mass conservative due to discontinuous pressure space
- Fast, matrix-free application of stiffness and mass matrices
- Hexahedral elements allow exploiting the tensor product structure of basis functions for a high floating point to memory operations ratio

Linear solver: Preconditioned Krylov subspace method

Fully coupled iterative solver: GMRES with upper triangular block preconditioning

$$\underbrace{\begin{bmatrix}\mathrm{A}&\mathrm{B}^{\top}\\ \mathrm{B}&0\end{bmatrix}}_{\text{Stokes operator}}\underbrace{\begin{bmatrix}\tilde{\mathrm{A}}&\mathrm{B}^{\top}\\ 0&-\tilde{\mathrm{S}}\end{bmatrix}^{-1}}_{\text{preconditioner}}\begin{bmatrix}\mathrm{u}'\\ \mathrm{p}'\end{bmatrix}=\begin{bmatrix}\mathrm{f}\\ 0\end{bmatrix}$$

Approximating the inverse of the viscous stress block, $\tilde{\mathrm{A}}^{-1}\approx\mathrm{A}^{-1}$, is well suited for multigrid methods. Next, find an approximation for the inverse Schur complement, $\tilde{\mathrm{S}}^{-1}\approx\mathrm{S}^{-1}:=(\mathrm{BA}^{-1}\mathrm{B}^{\top})^{-1}$.

BFBT/LSC Schur complement approximation $\tilde{\mathrm{S}}^{-1}$

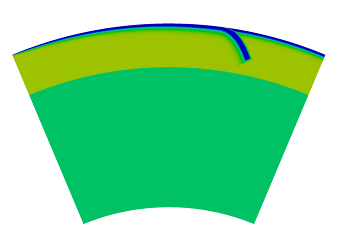
Improved BFBT / Least Squares Commutator (LSC) method:

$$\tilde{\mathrm{S}}^{-1}=(\mathrm{BD}^{-1}\mathrm{B}^{\top})^{-1}(\mathrm{BD}^{-1}\mathrm{AD}^{-1}\mathrm{B}^{\top})(\mathrm{BD}^{-1}\mathrm{B}^{\top})^{-1}$$

with diagonal scaling, $\mathrm{D}:=\mathrm{diag}(\mathrm{A})$. Derived from the least squares minimizer of a commutation relationship of A and $\mathrm{B}^{\top}$. Here, approximating the inverse of the discrete pressure Laplacian, $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$, is well suited for multigrid methods.

3. Stokes solver robustness with scaled BFBT Schur complement approximation

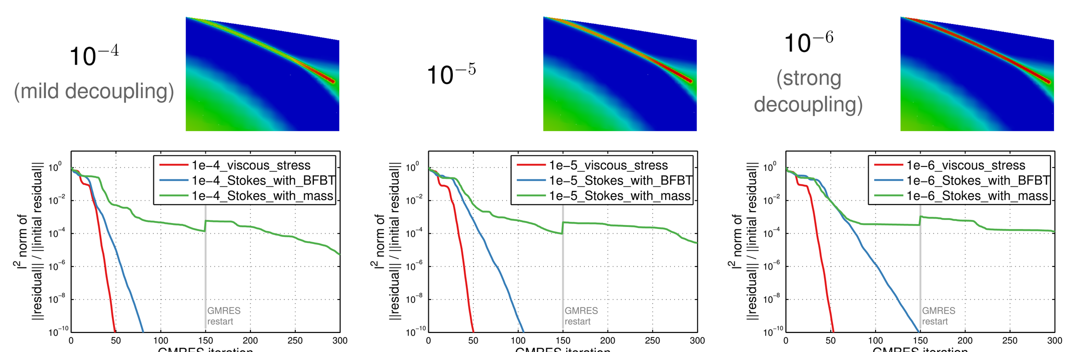
The subducting plate model problem on a cross section of the spherical Earth domain serves as a benchmark for solver robustness.



Subduction model viscosity field.

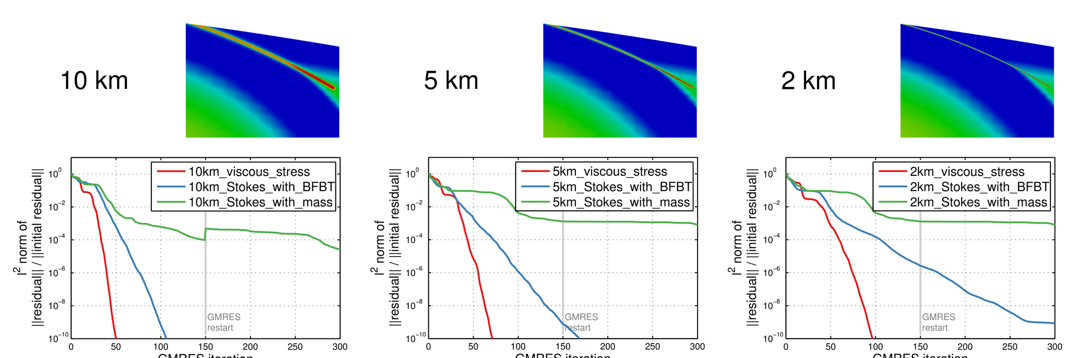
Multigrid parameters:
GMG for $\tilde{\mathrm{A}}$: 1 V-cycle, 3+3 smooth
AMG (PETSc's GAMG) for $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$:
3 V-cycles, 3+3 smooth

Robustness with respect to plate coupling strength



Convergence for solving $\mathrm{Au}=\mathrm{f}$ (*red*), Stokes system with BFBT (*blue*), Stokes system with viscosity weighted mass matrix as Schur complement approximation (*green*) for comparison to conventional preconditioning.

Robustness with respect to plate boundary thickness

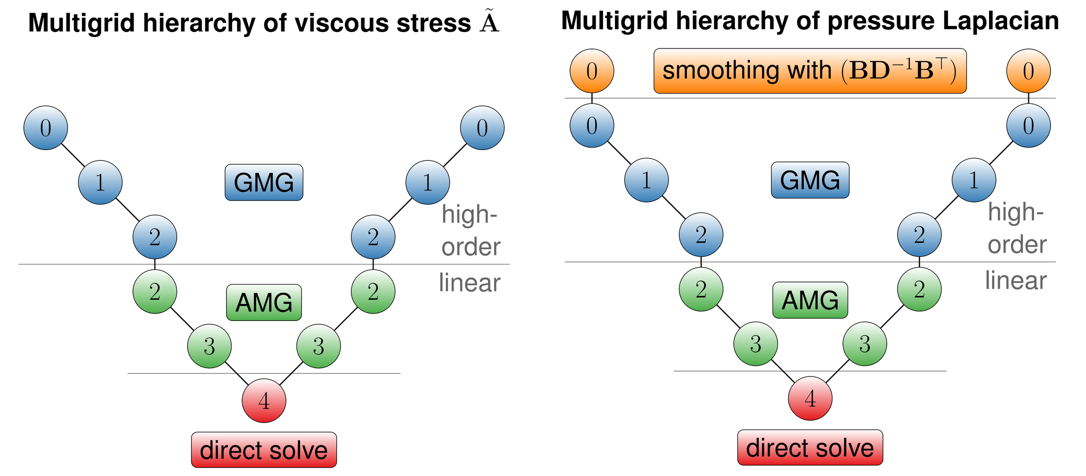


Convergence for solving $\mathrm{Au}=\mathrm{f}$ (*red*), Stokes system with BFBT (*blue*), Stokes system with viscosity weighted mass matrix as Schur complement approximation (*green*) for comparison to conventional preconditioning.

4. Parallel adaptive high-order geometric multigrid

The multigrid hierarchy of nested meshes is generated from an adaptively refined octree-based mesh via geometric coarsening:

- Parallel repartitioning of coarser meshes for load-balancing; repartitioning of sufficiently coarse meshes on subsets of cores
- High-order L^2-projection of coefficients onto coarser levels; re-discretization of differential equations at each coarser geometric multigrid level
- As the coarse solver for geometric multigrid, AMG (PETSc's GAMG) is invoked on only small core counts
- Geometric multigrid for the pressure Laplacian is problematic due to the discontinuous modal pressure discretization $\mathbb{P}_{k-1}^{\mathrm{disc}}$; here, a novel approach is taken by re-discretizing with continuous nodal $\mathbb{Q}_k$ basis functions

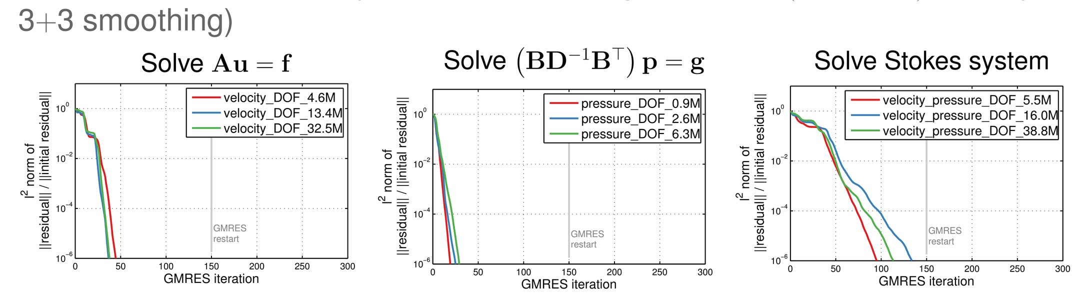


- GMG smoother: Chebyshev accelerated Jacobi (PETSc) with matrix-free high-order stiffness apply, assembly of high-order diagonal only
- GMG smoother for $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$, discontinuous modal: Chebyshev accelerated Jacobi (PETSc) with matrix-free apply and assembled diagonal
- GMG restriction & interpolation: High-order L^2-projection; restriction and interpolation operators are adjoints of each other in L^2-sense
- GMG restriction & interpolation for $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$: L^2-projection between discontinuous modal and continuous nodal spaces
- No collective communication in GMG cycles needed

5. Convergence dependence on mesh size and discretization order

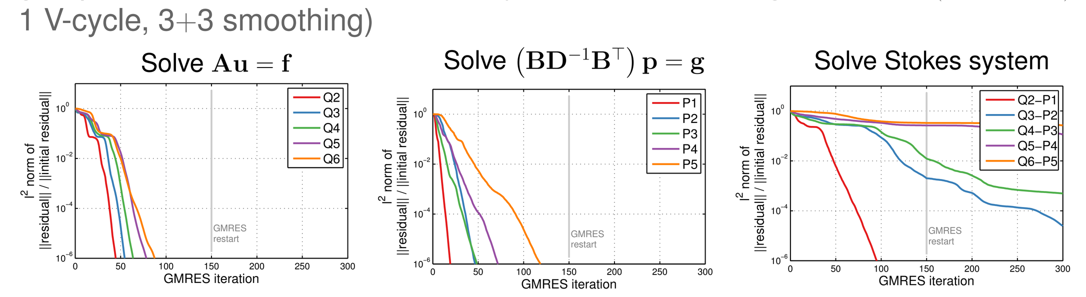
h-dependence using geometric multigrid for $\tilde{\mathrm{A}}$ and $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$

The mesh is increasingly refined while the discretization stays fixed to $\mathbb{Q}_2\times\mathbb{P}_1^{\mathrm{disc}}$. Performed with subducting plate model problem (see above). (Multigrid parameters: GMG for $\tilde{\mathrm{A}}$: 1 V-cycle, 3+3 smoothing; GMG for $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$: 1 V-cycle, 3+3 smoothing)



p-dependence using geometric multigrid for $\tilde{\mathrm{A}}$ and $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$

The discretization order of the finite element space increases while the mesh stays fixed. Performed with subducting plate model problem (see above). (Multigrid parameters: GMG for $\tilde{\mathrm{A}}$: 1 V-cycle, 3+3 smoothing; GMG for $(\mathrm{BD}^{-1}\mathrm{B}^{\top})$: 1 V-cycle, 3+3 smoothing)

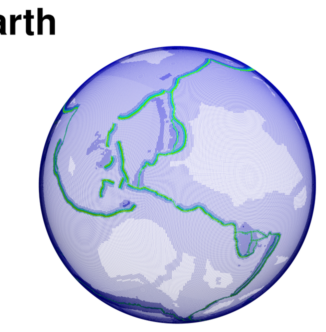


Remark: The deteriorating Stokes convergence with increasing order is due to a deteriorating approximation of the Schur complement by the BFBT method and not the multigrid components.

6. Parallel scalability of geometric multigrid

Global problem on adaptive mesh of the Earth

- Viscosity is generated from real Earth data
- Heterogeneous viscosity field exhibits 6 orders of magnitude variation
- Adaptively refined mesh (p4est library) down to $\sim$0.5 km local resolution; $\mathbb{Q}_2\times\mathbb{P}_1^{\mathrm{disc}}$ discretization
- Distributed memory parallelization with MPI



Stampede at the Texas Advanced Computing Center
16 CPU cores per node (2×8 core Intel Xeon E5-2680)
32GB main memory per node ($8\times$ 4GB DDR3-1600MHz)
6,400 nodes, 102,400 cores total, InfiniBand FDR network

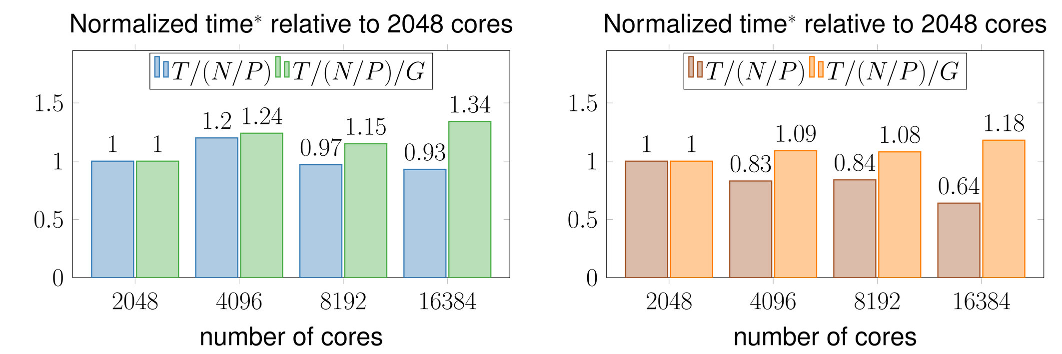
Weak scalability using adaptively refined Earth mesh

Normalized time* based on the setup and solve times for solving for velocity $\mathbf{u}$ in:

$$\mathbf{Au}=\mathbf{f}$$

Normalized time* based on the setup and solve times for solving for pressure $\mathbf{p}$ in:

$$(\mathbf{BD}^{-1}\mathbf{B}^{\top})\,\mathbf{p}=\mathbf{g}$$



*Normalization explanation:
$T/(N/P)$ ...... Scalability of algorithms & implementation
$T/(N/P)/G$ ... Scalability of implementation

T ... setup + solve time
N ... degrees of freedom (DOF)
P ... number of CPU cores
G ... number of GMRES iterations

Detailed timings for solving $\mathbf{Au}=\mathbf{f}$

#cores	velocity DOF	#levels GMG, AMG	setup time (s) GMG, AMG, total	solve time (s)	#iter
2048	637M	7, 4	10.2, 14.3, 24.6	2298.0	402
4096	1155M	7, 4	12.8, 28.6, 41.4	2482.5	389
8192	2437M	8, 4	15.2, 15.6, 30.8	2129.5	339
16384	5371M	8, 4	29.2, 51.0, 80.2	2198.4	279

Detailed timings for solving $(\mathbf{BD}^{-1}\mathbf{B}^{\top})\,\mathbf{p}=\mathbf{g}$

#cores	pressure DOF	#levels GMG, AMG	setup time (s) GMG, AMG, total	solve time (s)	#iter
2048	125M	7, 3	11.3, 0.9, 12.2	857.2	125
4096	227M	7, 4	12.3, 2.2, 14.5	638.0	95
8192	482M	8, 3	18.1, 1.5, 19.7	684.0	97
16384	1042M	8, 4	26.6, 9.1, 35.7	546.2	68

Remark: The number of GMRES iterations until convergence is reducing as the mesh is refined. This is due to an increasingly better resolution of the variations in the viscosity.

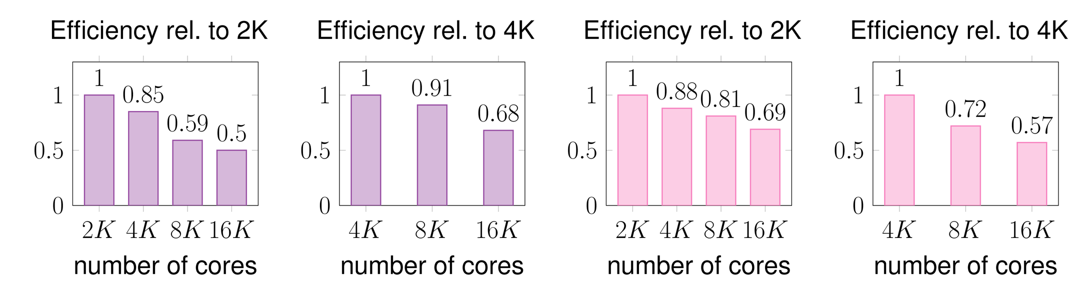
Strong scalability using fixed adaptive Earth mesh

Efficiency based on the setup and solve times for solving for velocity $\mathbf{u}$ in:

$$\mathbf{Au}=\mathbf{f}$$

Efficiency based on the setup and solve times for solving for pressure $\mathbf{p}$ in:

$$(\mathbf{BD}^{-1}\mathbf{B}^{\top})\,\mathbf{p}=\mathbf{g}$$



Problem size: 637M
#iterations: 401 ($\pm$1)

#	setup time (s) GMG, AMG, total	solve time (s)
2K	10.2, 14.3, 24.6	2298.0
4K	8.6, 16.4, 25.0	1327.5
8K	12.5, 27.2, 39.7	938.4
16K	11.4, 24.3, 35.8	544.9

Problem size: 1155M
#iterations: 388 ($\pm$1)

#	setup time (s) GMG, AMG, total	solve time (s)
2K	—	—
4K	12.8, 28.6, 41.4	2482.5
8K	10.2, 38.5, 48.7	1325.9
16K	17.0, 47.9, 64.9	858.6

Problem size: 125M
#iterations: 125 ($\pm$2)

#	setup time (s) GMG, AMG, total	solve time (s)
2K	11.3, 0.9, 12.2	857.2
4K	7.8, 1.1, 8.9	487.1
8K	7.7, 1.5, 9.2	256.3
16K	7.8, 2.4, 10.2	147.6

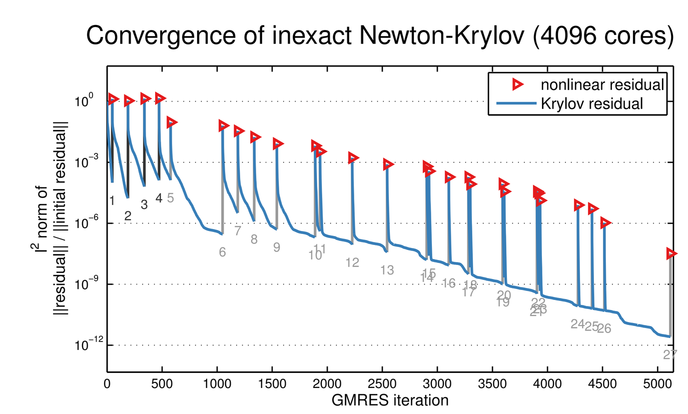
Problem size: 227M
#iterations: 96 ($\pm$1)

#	setup time (s) GMG, AMG, total	solve time (s)
2K	—	—
4K	12.3, 2.2, 14.5	638.0
8K	17.0, 9.9, 26.9	430.8
16K	13.5, 4.4, 17.9	269.1

7. Scalable nonlinear Stokes solver: Inexact Newton-Krylov method

- Newton update is computed inexactly via Krylov subspace iterative method
- Krylov tolerance decreases with subsequent Newton steps to guarantee superlinear convergence
- Number of Newton steps is independent of the mesh size
- Velocity residual is measured in H^{-1}-norm for backtracking line search; this avoids overly conservative update steps $\ll 1$ (evaluation of residual norm requires 3 scalar constant coefficient Laplace solves, which are performed by PCG with GMG preconditioning)
- Grid continuation at initial Newton steps: Adaptive mesh refinement to resolve increasing viscosity variations arising from the nonlinear dependence on the velocity

Convergence of inexact Newton-Krylov (4096 cores)



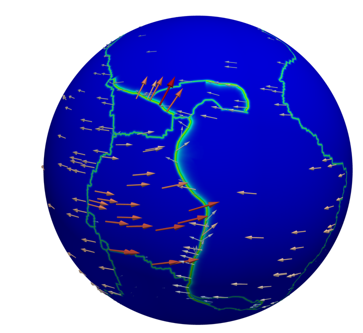


Plate velocities at nonlinear solution.

Note: Adaptive mesh refinements after the first four Newton steps are indicated by *black vertical lines*. 642M velocity & pressure DOF at solution, 473 min total runtime on 4096 cores.