An Optimization-Based Perturbed Newton Method for Viscoplastic Fluids with von Mises Yielding

Johann Rudi[1] Georg Stadler[2] Omar Ghattas[3]

[1]Mathematics and Computer Science Division, Argonne National Laboratory

[2]Courant Institute of Mathematical Sciences, New York University

[3]Oden Institute, Jackson School of Geosciences, and Dept. of Mechanical Engineering, The University of Texas at Austin

Outline

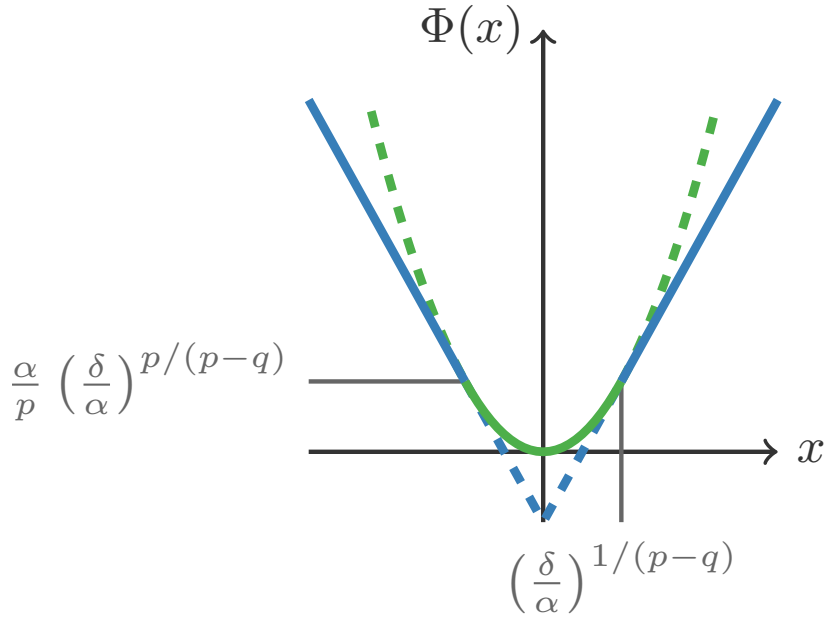
Introduction: Minimization problem

Find the minimizer $x^* \in \mathbb{R}$ of the problem

$$\min_x \Phi(x) := \begin{cases} \frac{\alpha}{p} |x|^p , & \alpha |x|^{p-q} \leq \delta, \\ \frac{\delta}{q} |x|^q + \gamma, & \text{otherwise}, \end{cases}$$

with given

- exponents $1 \leq q < p < \infty$
- parameters $\alpha, \delta > 0$
- $\gamma := \alpha \left(\frac{\delta}{\alpha}\right)^{p/(p-q)} \left(\frac{1}{p} - \frac{1}{q}\right)$

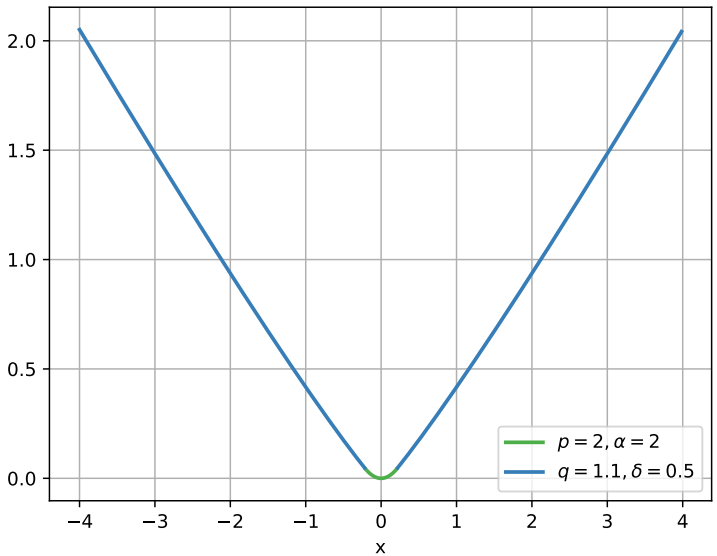


Introduction: Objective, gradient, and Hessian

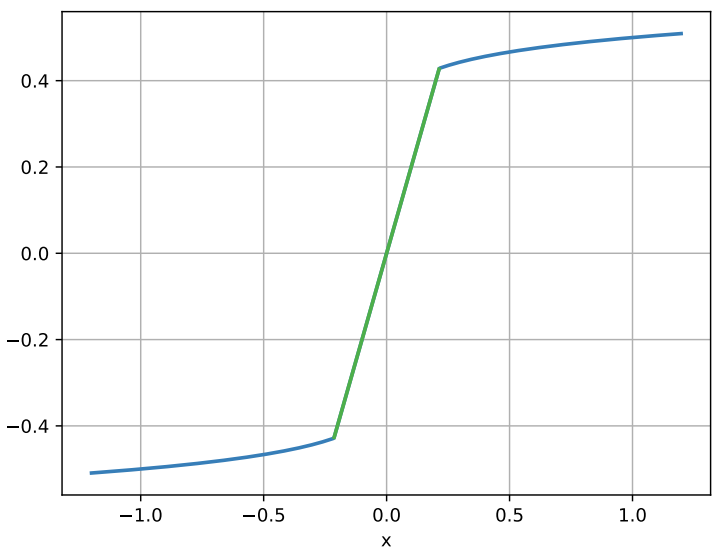
Parameters:

- $p = 2$
- $q = 1.1$
- $\alpha = 2$
- $\delta = 0.5$

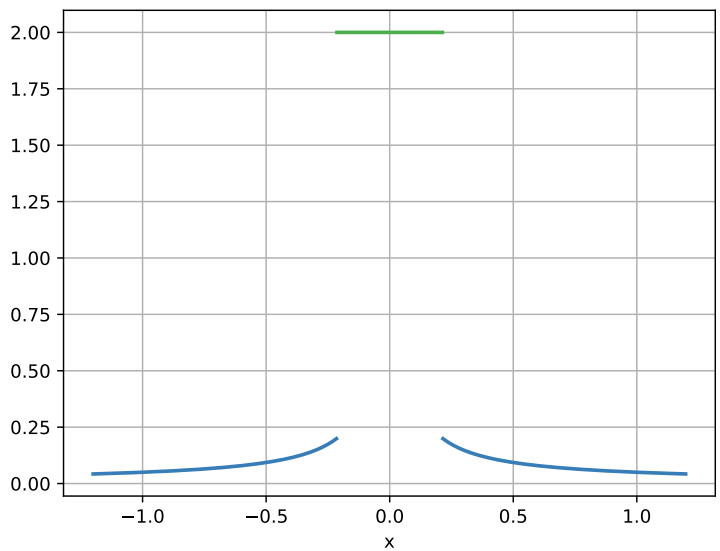
Objective Φ



Gradient Φ'

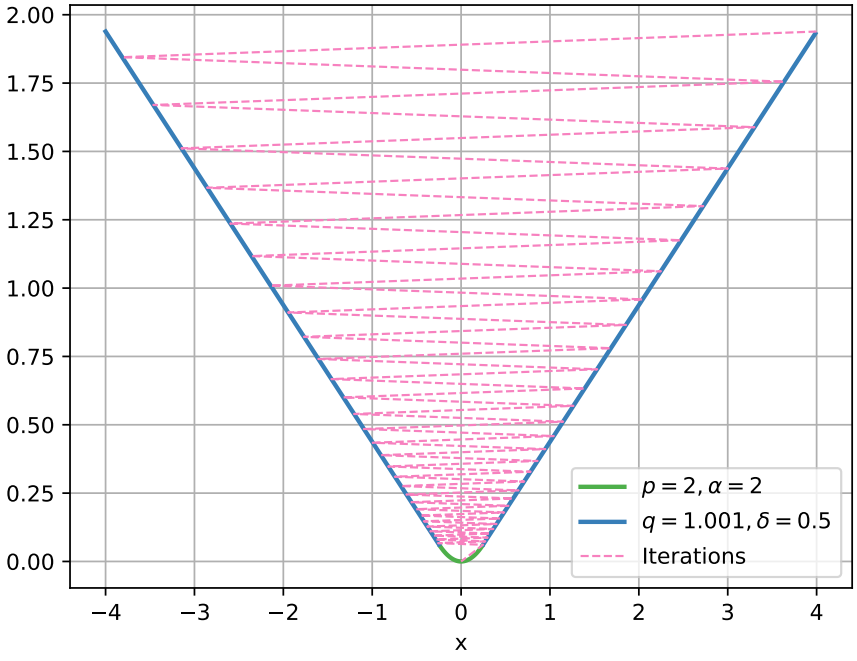


Hessian Φ''

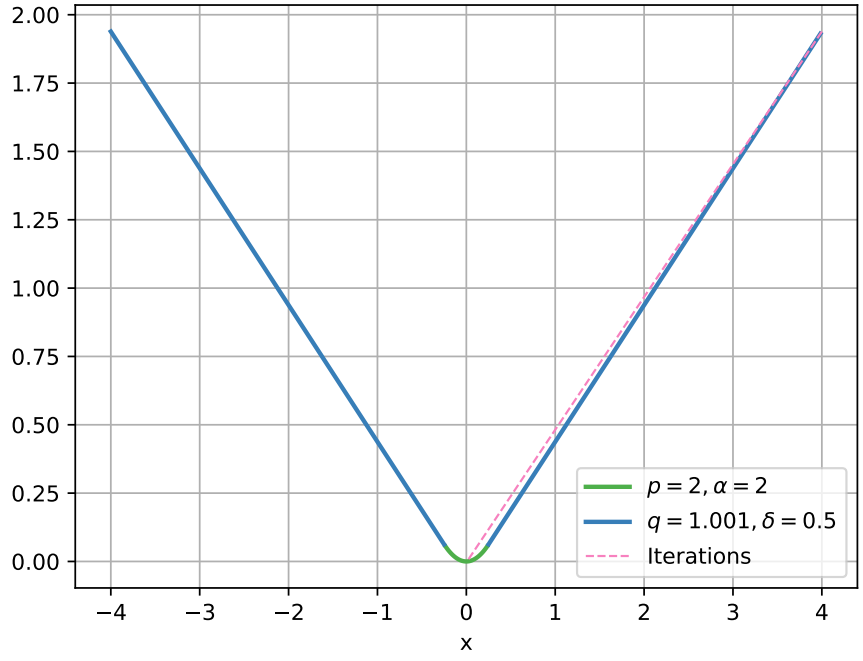


Introduction: Numerical experiments

Standard approach:
Newton: 59 iterations



Want:
Steepest Descend + Newton: 2 iterations



Parameters: $p = 2$, $q = 1.001$, $\alpha = 2$, $\delta = 0.5$

Outline

Minimization problem

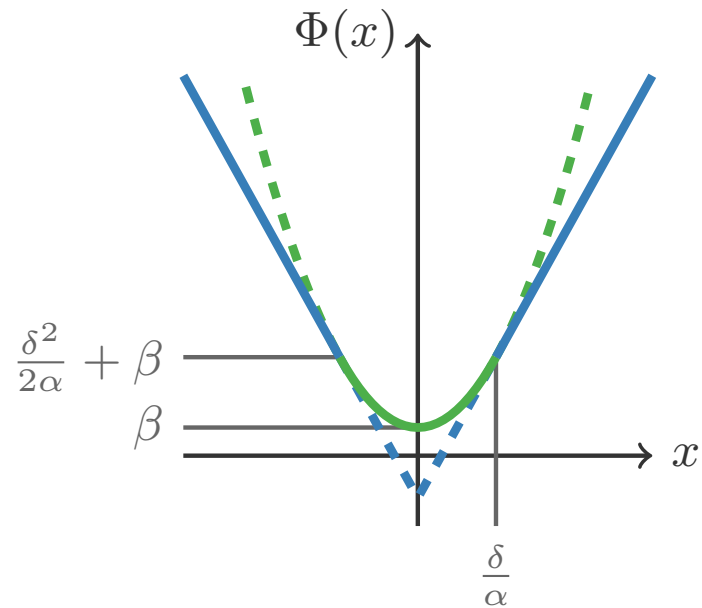
Given dimensions $d \in \mathbb{N}$ and $n \in \{1, d, d \times d\}$, domain $\Omega \subseteq \mathbb{R}^d$, and a linear and bounded functional $F : L^2(\Omega)^n \to \mathbb{R}$,

$$\text{find minimizer} \quad U^* : \Omega \to \mathbb{R}^n \quad \text{of} \quad \min_U J(U) := \int_\Omega \Phi(U) - F(U)$$

using a generalization of the *Huber loss:*

$$\Phi(U) := \beta + \begin{cases} \frac{\alpha}{2} |U|^2, & \alpha |U| \leq \delta, \\ \delta |U| - \frac{\delta^2}{2\alpha}, & \text{otherwise}, \end{cases}$$

with parameters $\alpha, \delta > 0$ and $\beta \in \mathbb{R}$.



Newton's method: A quadratic model of the nonlinearity

Goal: Construct a model of the nonlinear objective $J(U)$ locally around U. Newton's method approximates J with a Taylor expansion, truncated at the second-order term: (g and H are gradient and Hessian resp.)

$$J(U+\hat{U}) \approx J(U) + g(U)\hat{U} + \frac{1}{2}\Big(H(U)\hat{U}\,,\hat{U}\Big)$$

One Newton iteration involves:

1. Solve the linearized system for step $\hat{U}$ (in variational form):

$$\Big(H(U)\hat{U}\,,\tilde{U}\Big) = -g(U)\tilde{U} \quad \text{for all } \tilde{U}$$

2. Backtracking/line search algorithm to find $\theta \in (0,1]$ such that

$$J(U+\theta\hat{U}) < J(U) - \epsilon, \quad \text{for some fixed } \epsilon > 0$$

3. Update the solution: $U \leftarrow U + \theta\hat{U}$

Standard Newton linearization

The gradient and Hessian are the 1st- and 2nd-order variations of J:

$$g(U)\tilde{U} \coloneqq \int_\Omega \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\right)\left\langle U\,,\tilde{U}\right\rangle - F(\tilde{U}), \quad \chi \coloneqq \begin{cases} 1, & \alpha\,|U| \leq \delta, \\ 0, & \text{o.w.}, \end{cases}$$

$$\left(H(U)\hat{U}\,,\tilde{U}\right) \coloneqq \int_\Omega \left\langle \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U\otimes U}{|U|^2}\right)\right)\hat{U}\,,\tilde{U}\right\rangle$$

The standard Newton linearization requires to solve for Newton step $\hat{U}$ in

$$\int_\Omega \left\langle \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U\otimes U}{|U|^2}\right)\right)\hat{U}\,,\tilde{U}\right\rangle = -g(U)\tilde{U}$$

The outer product term is computationally challenging because:

- Coefficient $\left(\mathbf{I} - \frac{U\otimes U}{|U|^2}\right)$ in Hessian represents an orthogonal projector
- Hessian has a zero eigenvalue associated to eigenvector $(1-\chi)U$

Locally perturbed Newton linearization

Introduce for $U, S \in L^2(\Omega)^n$

$$\text{Error term:} \qquad E(U,S) \coloneqq \frac{U}{|U|} - S$$

$$\text{Perturbation:} \quad D(U,S) \coloneqq U - |U|\, S$$

We augment the previous gradient

$$g(U)\tilde{U} = \int_\Omega \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|} \right) \left\langle U \,, \tilde{U} \right\rangle - F(\tilde{U})$$

by a model perturbation, $D(U,S) \neq 0$,

$$g(U,S)\tilde{U} = \int_\Omega \left\langle \chi\alpha U + (1-\chi)\delta S \,, \tilde{U} \right\rangle - F(\tilde{U})$$

$$\left(D(U,S) \,, \tilde{S} \right) = \int_\Omega \left\langle U - |U|\, S \,, \tilde{S} \right\rangle$$

Locally perturbed and reduced Newton linearization

The perturbed Newton linearization for step $(\hat{U}, \hat{S})$ is

$$\int_\Omega \left\langle \chi\alpha\hat{U} + (1-\chi)\delta\hat{S}\,,\tilde{U}\right\rangle = -g(U,S)\tilde{U}$$

$$\int_\Omega \left\langle \left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\hat{U} - |U|\,\hat{S}\,,\tilde{S}\right\rangle = -\left(D(U,S)\,,\tilde{S}\right)$$

and has an explicit expression for the dual step

$$\hat{S} = \frac{U}{|U|} - S + \frac{1}{|U|}\left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\hat{U}$$

Substitute to get the perturbed and reduced Newton linearization for $\hat{U}$

$$\int_\Omega \left\langle \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\right)\hat{U}\,,\tilde{U}\right\rangle = -g(U)\tilde{U}$$

Improvements gained from perturbation

$$\int_\Omega \left\langle \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U \otimes U}{|U|^2} \right) \right) \hat{U} \, , \tilde{U} \right\rangle = -g(U)\tilde{U}$$

$$\int_\Omega \left\langle \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U \otimes S}{|U|} \right) \right) \hat{U} \, , \tilde{U} \right\rangle = -g(U)\tilde{U}$$

- Perturbation results in model error dependent regularization
- It maintains fast Newton convergence close to the solution

$$\mathbf{I} - \frac{U \otimes S}{|U|} = \mathbf{I} - \frac{U \otimes U}{|U|^2} + \frac{U \otimes E(U,S)}{|U|} \to \mathbf{I} - \frac{U \otimes U}{|U|^2} \quad \text{as} \quad E(U,S) \to 0$$

- It acts as a nonlinear preconditioner far from the solution
- Simple and computationally cheap update of the dual variable S given $\hat{U}$

$$S \leftarrow S + \hat{S} = \frac{U}{|U|} + \frac{1}{|U|}\left(\mathbf{I} - \frac{U \otimes S}{|U|} \right) \hat{U}$$

Outline

Earth's nonlinear rheology modeling the mantle & plates

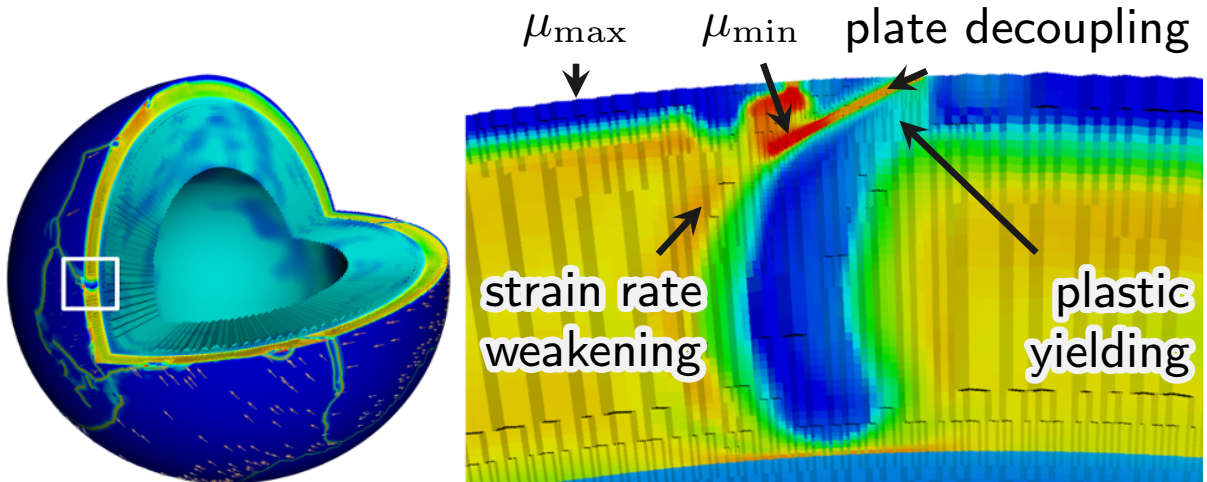
Nonlinear constitutive relationship / rheology due to:

- Strain rate weakening exponent $n \geq 1$ ($\dot{\varepsilon}_{\text{II}}(\boldsymbol{u})$ is 2nd invariant of strain rate)
- Yield strength $\tau_{\text{yield}} > 0$ causing plastic yielding

Additional heterogeneity is introduced via:

- Exponential temperature dependence $a(T)$ (Arrhenius relationship)
- Plate decoupling factor $0 < w(\boldsymbol{x}) \leq 1$ with orders-of-magnitude contrasts

$$\mu_{\text{eff}}(T, \dot{\varepsilon}_{\text{II}}(\boldsymbol{u})) := \mu_{\text{min}} + \min\left(\frac{\tau_{\text{yield}}}{2\dot{\varepsilon}_{\text{II}}(\boldsymbol{u})} \, , \, w(\boldsymbol{x}) \min\left(\mu_{\text{max}} \, , \, a(T)\, \dot{\varepsilon}_{\text{II}}(\boldsymbol{u})^{\frac{1}{n}-1} \right) \right)$$



Mantle flow governed by incompressible Stokes equations

Nonlinear incompressible Stokes PDE (w/ free-slip & no-normal flow BC):

$$-\nabla \cdot \left[2\mu(\dot{\varepsilon}_{\mathrm{II}}(\boldsymbol{u}))\,\dot{\varepsilon}(\boldsymbol{u})\right] + \nabla p = \boldsymbol{f} \qquad \text{viscosity } \mu,\ \text{RHS forcing } \boldsymbol{f}$$

$$-\nabla \cdot \boldsymbol{u} = 0 \qquad \text{seek: velocity } \boldsymbol{u},\ \text{pressure } p$$

Linearization (with Newton), then discretization (with inf-sup stable F.E.):

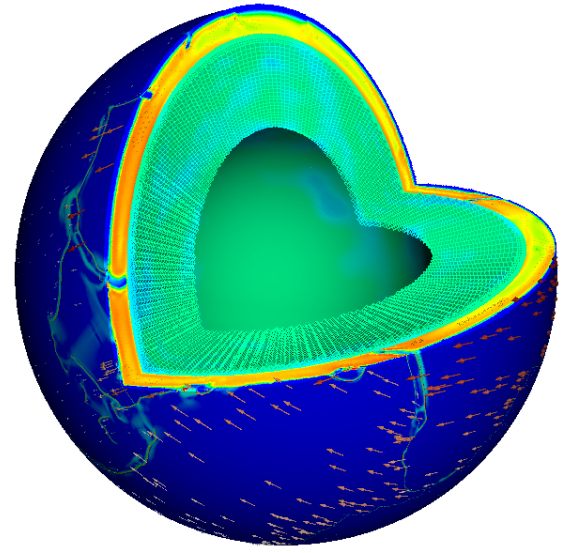
$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathsf{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix}$$

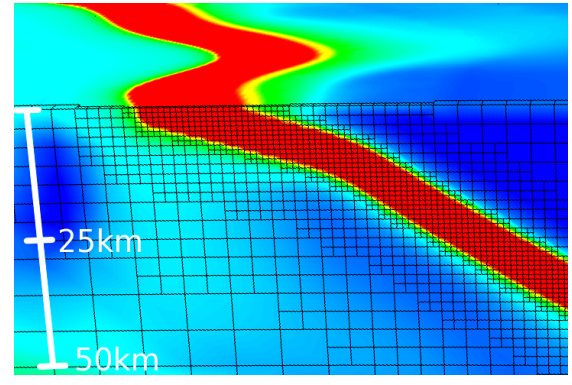
- High-order finite element shape functions
- Inf-sup stable velocity–pressure pairings: $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\mathrm{disc}}$ with order $k \geq 2$
- Locally mass conservative due to discontinuous, modal pressure
- Non-conforming hexahedral meshes with "hanging nodes"
- Adaptive mesh refinement resolving fine-scale features of mantle

Severe challenges for parallel scalable implicit solvers

. . . arising in global mantle convection:

- Severe nonlinearity and heterogeneity of Earth's rheology and anisotropy induced by it
- Sharp viscosity gradients in narrow regions (6 orders of magnitude drop in $\sim$5 km)
- Wide range of spatial scales and highly localized features, e.g., plate boundaries of size $\mathcal{O}(1\,\text{km})$ influence plate motion at continental scales of $\mathcal{O}(1000\,\text{km})$
- Adaptive mesh refinement is essential
- High-order finite elements $\mathbb{Q}_k \times \mathbb{P}_{k-1}^{\text{disc}}$, order $k \geq 2$, with local mass conservation; yields a difficult to deal with discontinuous, modal pressure approximation





Viscosity *(colors)* and locally refined mesh.

Outline

Stokes flow with yielding rheology: Minimization problem

Given viscosity $\mu > 0$, yield strength $\tau_{\text{yield}} > 0$, the 2nd invariant of the strain rate tensor, $\dot{\varepsilon}_{\text{II}} := \frac{1}{\sqrt{2}} |\nabla_{\text{s}} \boldsymbol{u}|$ with $\nabla_{\text{s}} \boldsymbol{u} := \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathsf{T}})$, and

$$\Phi(\nabla_{\text{s}} \boldsymbol{u}) = \begin{cases} 2\mu \dot{\varepsilon}_{\text{II}}^2 + \dfrac{\tau_{\text{yield}}^2}{2\mu}, & 2\mu \dot{\varepsilon}_{\text{II}} \leq \tau_{\text{yield}}, \\ 2\tau_{\text{yield}} \dot{\varepsilon}_{\text{II}}, & \text{otherwise}, \end{cases}$$

with Huber parameters $\alpha = 2\mu$, $\delta = \sqrt{2}\tau_{\text{yield}}$, $\beta = \frac{\delta^2}{2\alpha} = \frac{\tau_{\text{yield}}^2}{2\mu}$; consider

$$\text{find minimizer} \quad \boldsymbol{u}^* : (0,1)^3 \to \mathbb{R}^3 \quad \text{of}$$

$$\min_{\boldsymbol{u}} J(\boldsymbol{u}) := \int_\Omega \Phi(\nabla_{\text{s}} \boldsymbol{u}) - \boldsymbol{f} \cdot \boldsymbol{u} \quad \text{s.t.} \quad \nabla \cdot \boldsymbol{u} = 0$$

Standard v.s. perturbed Newton linearization

Standard Newton linearization yields the system for step $(\hat{\boldsymbol{u}}, \hat{p})$:
(Note: The viscosity in the Newton step is an anisotropic 4^{th}-order tensor.)

$$-\nabla \cdot \left[2 \left(\chi\mu + (1-\chi) \frac{\tau_{\text{yield}}}{\sqrt{2}\,|\nabla_{\text{s}} \boldsymbol{u}|} \left(\mathbf{I} - \frac{\nabla_{\text{s}} \boldsymbol{u} \otimes \nabla_{\text{s}} \boldsymbol{u}}{|\nabla_{\text{s}} \boldsymbol{u}|^2} \right) \right) \nabla_{\text{s}} \hat{\boldsymbol{u}} \right] + \nabla \hat{p} = -\boldsymbol{r}_{\text{mom}}$$

$$-\nabla \cdot \hat{\boldsymbol{u}} = -r_{\text{mass}}$$

Perturbed and reduced Newton linearization of the momentum equation:

$$-\nabla \cdot \left[2 \left(\chi\mu + (1-\chi) \frac{\tau_{\text{yield}}}{\sqrt{2}\,|\nabla_{\text{s}} \boldsymbol{u}|} \left(\mathbf{I} - \frac{\nabla_{\text{s}} \boldsymbol{u} \otimes \boldsymbol{S}}{|\nabla_{\text{s}} \boldsymbol{u}|} \right) \right) \nabla_{\text{s}} \hat{\boldsymbol{u}} \right] + \nabla \hat{p} = -\boldsymbol{r}_{\text{mom}}$$

where we defined, for $\boldsymbol{u} \in H^1(\Omega)^d$ and $\boldsymbol{S} \in L^\infty(\Omega)^{d \times d}$, $\|\boldsymbol{S}\|_{L^\infty} \leq 1$,

error: $E(\boldsymbol{u}, \boldsymbol{S}) := \dfrac{\nabla_{\text{s}} \boldsymbol{u}}{|\nabla_{\text{s}} \boldsymbol{u}|} - \boldsymbol{S}$, perturbation: $D(\boldsymbol{u}, \boldsymbol{S}) := \nabla_{\text{s}} \boldsymbol{u} - |\nabla_{\text{s}} \boldsymbol{u}|\, \boldsymbol{S}$

Numerical results for Stokes flow with yielding rheology

Numerical comparison of solver convergence for 3D nonlinear Stokes flow.

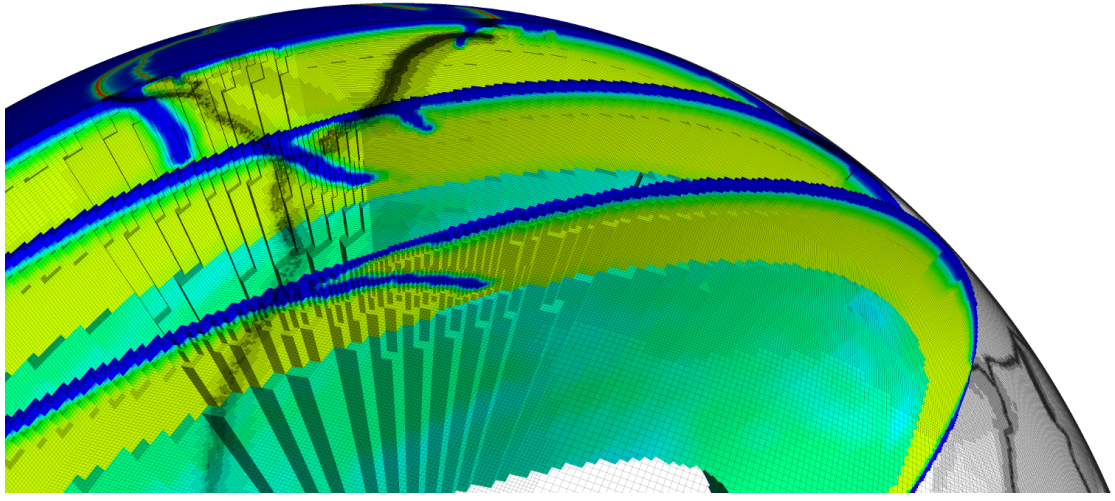
The model problem smoothly incorporates low-viscosity plumes in a high-viscosity background medium.

Yielding volume	**Mesh level** ℓ	**Standard Newton**			**Perturbed Newton**		
		It. Newton	#backtr.	It. GMRES	It. Newton	#backtr.	It. GMRES
∼45%	4	33	20	1469	10	0	379
∼45%	5	36	25	2255	12	0	664
∼45%	6	57	49	4255	13	0	876
∼65%	4	29	21	1559	18	10	965
∼65%	5	37	26	2464	17	9	1245
∼65%	6	48	39	3892	20	9	1707
∼90%	4	35	25	1505	19	11	872
∼90%	5	40	32	2147	21	11	1267
∼90%	6	32	21	2312	23	11	1811

Outline

Inexact Newton–Krylov with AMR for global mantle models

- Robust Newton convergence w.r.t. nonlinear parameters that is largely independent of the discretization; allows treatment of mantle models in an inverse problem setting.
- Grid continuation at initial Newton steps: Adaptive mesh refinement to resolve increasing viscosity variations arising from the nonlinear rheology.
- Velocity residual is measured in H^{-1}-norm for backtracking line search; this avoids overly conservative update steps $\ll 1$.

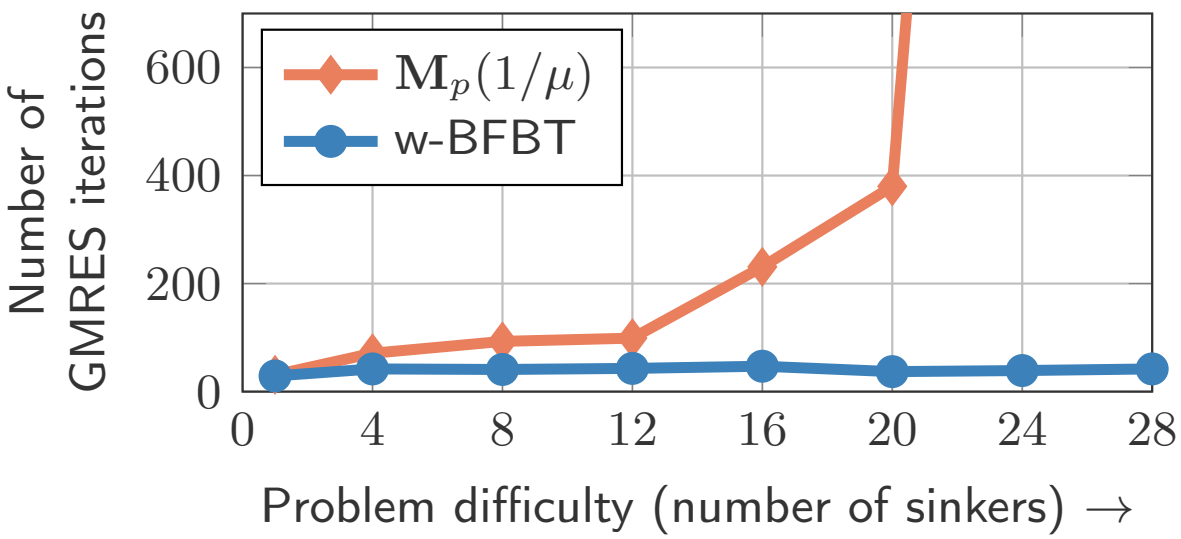


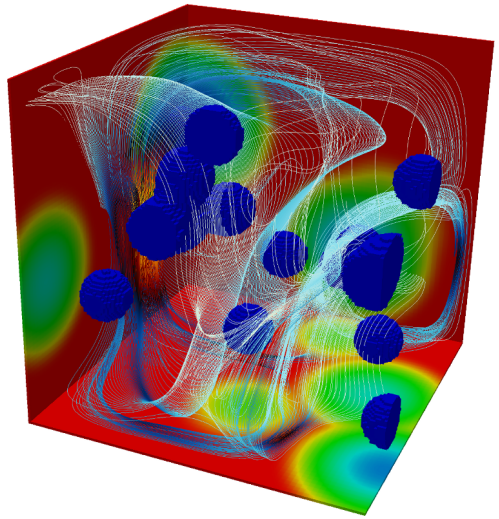
w-BFBT: Robust inverse Schur complement approximation

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^\mathsf{T} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{B}^\mathsf{T} \\ \mathbf{0} & \tilde{\mathbf{S}} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\hat{\mathbf{u}}} \\ \tilde{\hat{\mathbf{p}}} \end{bmatrix} = \begin{bmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \end{bmatrix} \qquad \begin{aligned} \tilde{\mathbf{A}}^{-1} &\approx \mathbf{A}^{-1} \;\rightarrow\; \text{MG V-cycle} \\ \tilde{\mathbf{S}}^{-1} &\approx (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\mathsf{T})^{-1} \end{aligned}$$

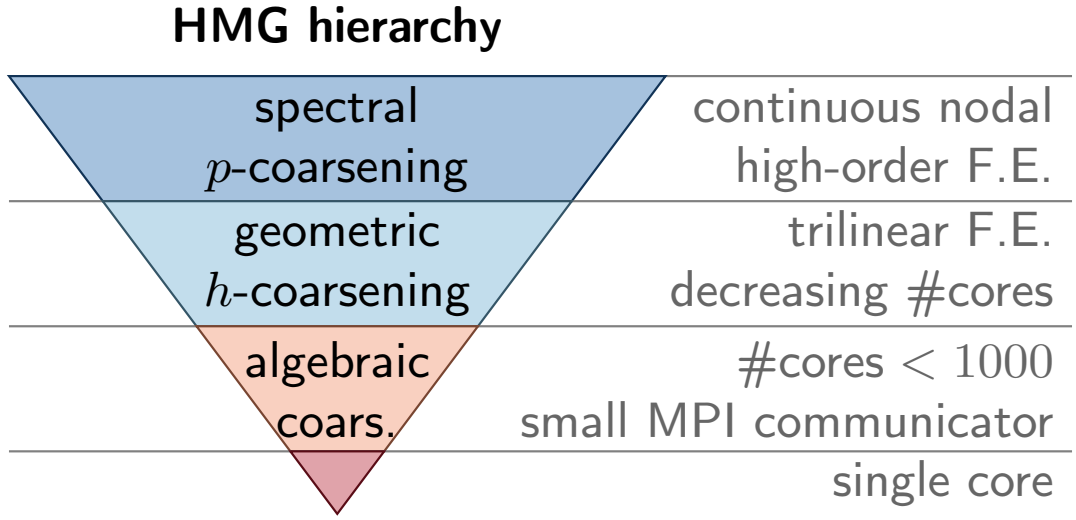
$$\tilde{\mathbf{S}}^{-1}_{\text{w-BFBT}} := \underbrace{\left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{B}^\mathsf{T}\right)^{-1}}_{\rightarrow \text{ MG V-cycle}} \left(\mathbf{B}\mathbf{C}_w^{-1}\mathbf{A}\mathbf{D}_w^{-1}\mathbf{B}^\mathsf{T}\right) \underbrace{\left(\mathbf{B}\mathbf{D}_w^{-1}\mathbf{B}^\mathsf{T}\right)^{-1}}_{\rightarrow \text{ MG V-cycle}}$$

Choice of diagonal weighting matrices $\mathbf{C}_w = \mathbf{D}_w$ is critical for efficacy & robustness [Rudi, Stadler, Ghattas, 2017].



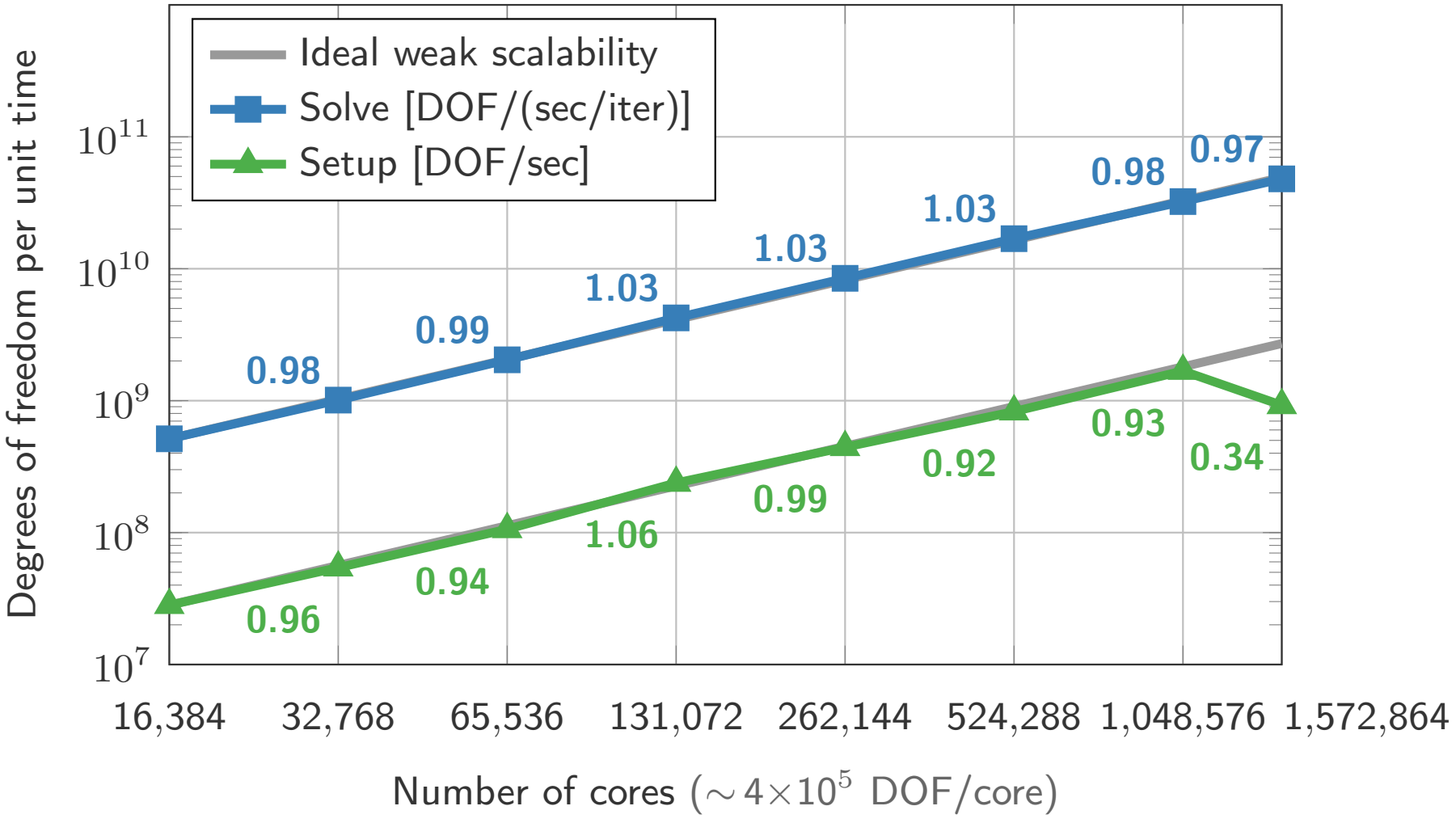


HMG: Hybrid spectral–geometric–algebraic multigrid



- Multigrid hierarchy of nested meshes is generated from an adaptively refined octree-based mesh via spectral–geometric coarsening
- Parallel repartitioning of coarser meshes for load-balancing (crucial for AMR); sufficiently coarse meshes occupy only subsets of cores
- High-order L^2-projection onto coarser/finer levels
- Chebyshev accelerated Jacobi smoother with tensorized matrix-free high-order stiffness apply; assembly of high-order diagonal only

Extreme weak scalability on LLNL's Sequoia supercomputer



[Rudi, Malossi, Isaac, Stadler, Gurnis, Staar, Ineichen, Bekas, Curioni, Ghattas, 2015]

Given: Observational data [Collab. with M. Gurnis, Caltech]

- Current plate motion from GPS and magnetic anomalies
- Topography indicating normal traction at Earth's surface
- Plate deformation obtained from dense GPS networks
- Average viscosity in regions affected by post-glacial rebound

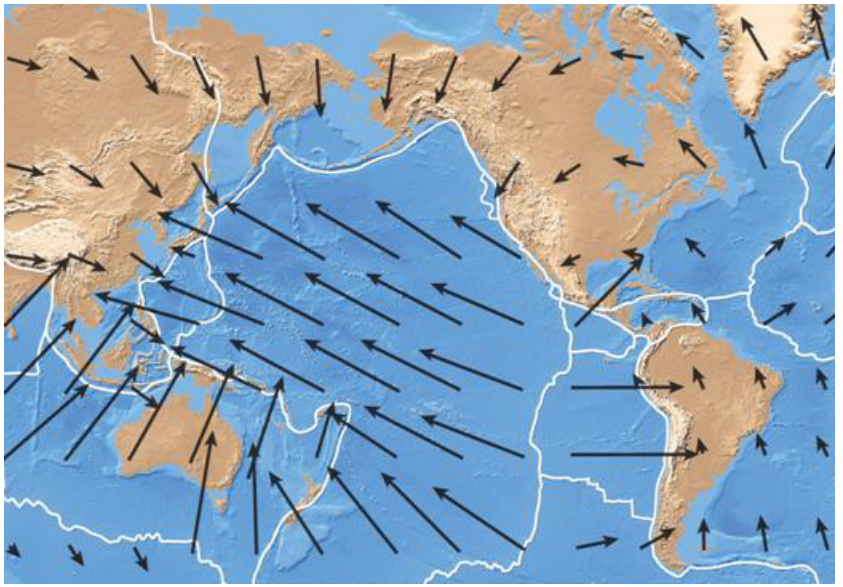


Plate motion (Credit: Pearson Prentice Hall, Inc.)

Additional knowledge contributing to mantle flow models:

- Location and geometry of plates, plate boundaries, and subducting slabs (from seismicity)
- Images of present-day Earth structure (by correlating seismic wave speed with temperature)
- Rock rheology extrapolated from laboratory experiments

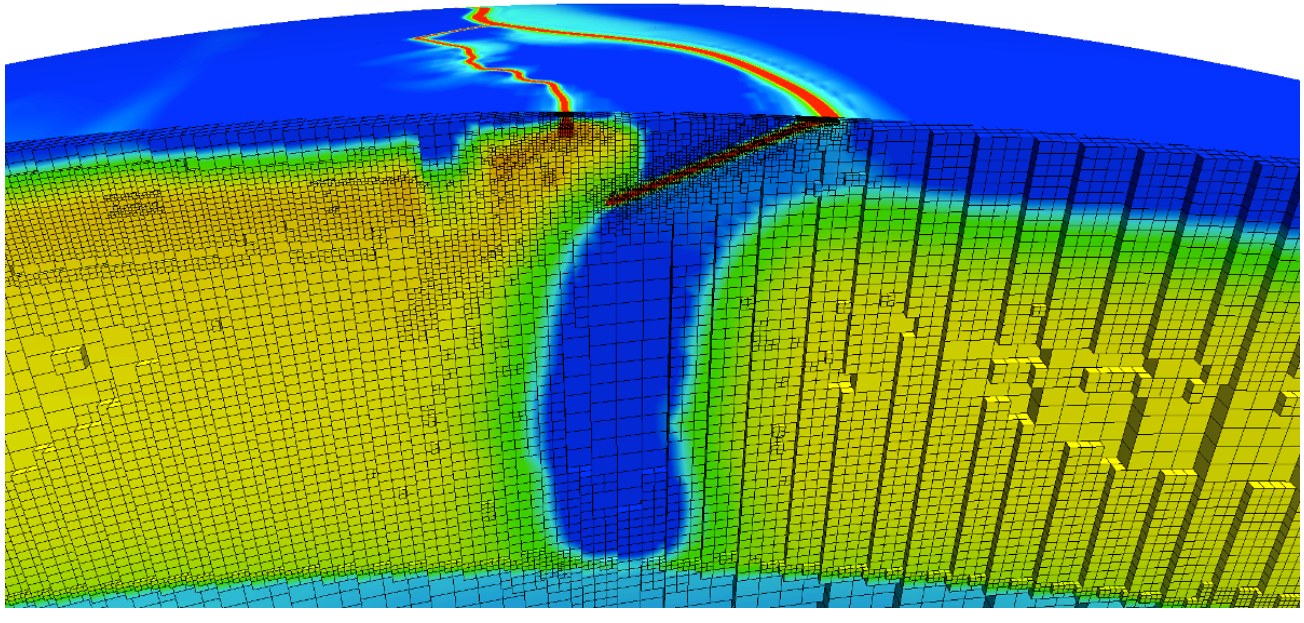
Want: Inversion to constrain parameters of mantle models

Constant rheological parameters affecting viscosity and nonlinearity:

- Scaling factor of the upper mantle viscosity (down to ∼660 km depth)
- Stress exponent controlling severity of strain rate weakening
- Yield strength governing plastic yielding phenomena

Spatially varying parameters modeling geometry of plate boundaries:

- Coupling strength / energy dissipation between plates



(Credit: Alisic)

Summary

Focus in today's talk:

- Our challenge: nonlinear applications modelled by optimization problems with singularities in the Hessian due to a L^1-norm type term in the objective.
- Hessian exhibits a null space upon linearization with Newton's method that is problematic for convergence.
- Applications include inverse problems with total variation regularization as well as viscoplastic/multi-viscosity flows like Earth's mantle convection.
- We analyze issues with the standard Newton linearization theoretically and through numerical experiments.
- We propose an improved linearization based on a perturbation of an otherwise implicitly assumed equality constraint.
- We achieve robust and fast Newton convergence largely independent of the discretization.

Global-scale mantle solver:

- Hybrid spectral–geometric–algebraic multigrid (builds on p4est library; extended by a coarsening correction to enable coarsening across core boundaries). [Rudi, Malossi, Isaac, et al., 2015]
- Weighted BFBT preconditioner for the for the Schur complement; scalable HMG-based BFBT algorithms, heterogeneity-robust weighting of BFBT and theoretical foundation. [Rudi, Stadler, Ghattas, 2017]
- Inexact Newton–Krylov with nonlinear preconditioning and grid continuation for highly nonlinear mantle rheology.
- Optimal algorithmic performance w.r.t. mesh refinement; nearly optimal w.r.t. higher discretization order.
- Parallel scalability of solvers to 1.6M cores (collaboration with IBM Research, Zurich). [Rudi, Malossi, Isaac, et al., 2015]