Scalable Inference of Non-Newtonian Rheology Parameters in Earth’s Mantle on HPC Platforms

Johann Rudi\textsuperscript{1} Georg Stadler\textsuperscript{2} Jiashun Hu\textsuperscript{3} Micheal Gurnis\textsuperscript{4}

\textsuperscript{1}Department of Mathematics, Virginia Tech
\textsuperscript{2}Courant Institute of Mathematical Sciences, New York University
\textsuperscript{3}Department of Earth and Space Sciences, Southern University of Science and Technology (China)
\textsuperscript{4}Seismological Laboratory at the Division of Geological and Planetary Sciences, California Institute of Technology

Pronouns: he/him/his/himself
Outline

Background

Earth’s Mantle Convection – The Driving Application and Challenges

Inference & Uncertainty Quantification

Numerical Results
  Inference on a cross section of Earth’s mantle
  Inference on the full sphere of the Earth
Introducing forward & inverse problems

**Forward Problem:**
- $m$: input parameters
- $\mathcal{F}$: forward model
- $u$: model output

**Inverse Problem:**
- $m$: recovered parameters
- $\tilde{\mathcal{F}}^{-1}$: inverse operator
- $d = u + \varepsilon$ (noise): observational data
Introducing forward & inverse problems

Forward problem

\[ m \rightarrow \mathcal{F} \rightarrow u \]

- \( m \): input parameters
- \( \mathcal{F} \): forward model
- \( u \): model output

Inverse problem

\[ \tilde{\mathcal{F}}^{-1} \rightarrow d = u + \varepsilon \]

- \( \tilde{\mathcal{F}}^{-1} \): inverse operator
- \( d \): observational data

Three different ways to compute solutions of inverse problems

- **Sampling-based** methods use randomness to explore the posterior density; typically don’t need derivatives → Markov chain Monte Carlo (MCMC)
- **Adjoint derivative-based** methods use techniques from optimization thus require gradients/Hessians → PDE-constrained optimization
- **Deep learning-based** methods directly construct an inverse map, \( \tilde{\mathcal{F}}^{-1} \), from data to parameters (or even posterior densities) → artificial neural networks
Introducing forward & inverse problems

forward problem

\[
\begin{align*}
\text{input parameters} & \rightarrow \mathcal{F} & \text{forward model} & \rightarrow \text{model output} \\
m & \rightarrow \mathcal{F} & u & \rightarrow \text{output}
\end{align*}
\]

inverse problem

\[
\begin{align*}
\text{recovered parameters} & \leftarrow \tilde{\mathcal{F}}^{-1} & \text{inverse operator} & \leftarrow d = u + \varepsilon \text{ (noise)} & \text{observational data}
\end{align*}
\]

Three different ways to compute solutions of inverse problems

Sampling-based methods use randomness to explore the posterior density; typically don’t need derivatives → Markov chain Monte Carlo (MCMC)

Adjoint derivative-based methods use techniques from optimization thus require gradients/Hessians → PDE-constrained optimization

Deep learning-based methods directly construct an inverse map, \( \tilde{\mathcal{F}}^{-1} \), from data to parameters (or even posterior densities) → artificial neural networks
Outline

Background

Earth’s Mantle Convection – The Driving Application and Challenges

Inference & Uncertainty Quantification

Numerical Results
  - Inference on a cross section of Earth’s mantle
  - Inference on the full sphere of the Earth
Nonlinear Stokes PDE modeling Earth’s mantle

**Model:** Nonlinear incompressible Stokes (w/ free-slip & no-normal flow BC) models present-day instantaneous flow

\[-\nabla \cdot \left[ \mu(x, \dot{\varepsilon}_{II}(u)) \left( \nabla u + \nabla u^T \right) \right] + \nabla p = f \quad \text{viscosity } \mu, \ \text{RHS forcing } f\]
\[-\nabla \cdot u = 0 \quad \text{unknown: velocity } u, \ \text{pressure } p\]

**Rheology / effective viscosity:** Shear-thinning with plastic yielding and plate decoupling (or weakening) factor \( w(x) \)

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

Visualization from *rhea* code. Colors represent viscosity; Widths of plate decoupling are exaggerated.
Given: Observational data

- 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

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: Constrain parameters of mantle models

Global rheological parameters affecting viscosity and nonlinearity:

- **Scaling factor** of the upper mantle viscosity (down to $\sim 660$ 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
Model of Earth’s plate boundaries

Plate boundaries at Earth’s surface (*red lines*) and plate geometries obtained from MORVEL plate motion data set\(^1\)

\(^1\)DeMets, Gordon, and Argus 2010.
Parametric model for plate decoupling / weak zones

\[ w(x) := 1 - (1 - w_{\text{min}}) \exp \left( -\frac{\xi(x)^2}{2\sigma^2} \right) \in (0, 1] \]

\[ \xi(x) := \max(0, d(x) - d_{\text{min}}) \quad \text{and} \quad \sigma := \frac{d_w - d_{\text{min}}}{2} \]

Weak zone profile with width \( d_w = 20 \text{ km} \), plate boundary width \( d_{\text{min}} = 5 \text{ km} \), and weak zone factor \( w_{\text{min}} = 10^{-5} \).
Geometries of decoupling surfaces of subducting plates

Surfaces $d(x)$ of the subducting plates, where colors indicate depth (red is shallow and blue is deep, respectively)
Forward solver for the nonlinear Stokes PDE

Nonlinear incompressible Stokes PDE

\[-\nabla \cdot \left[ \mu(x, \dot{\varepsilon}_{II}(u)) \left( \nabla u + \nabla u^T \right) \right] + \nabla p = f\]
\[-\nabla \cdot u = 0\]

- Inexact Newton–Krylov method with grid continuation

Linearization with Newton’s method, then discretization yields

\[
\begin{bmatrix}
A & B^T \\
B & 0
\end{bmatrix}
\begin{bmatrix}
\hat{u} \\
\hat{p}
\end{bmatrix} = 
\begin{bmatrix}
-r_1 \\
-r_2
\end{bmatrix}
\]

Careful design of discretization with inf-sup stable Finite Elements

- High-order finite element shape functions \((Q_k \times P_{k-1}^{\text{disc}}, k \geq 2)\)
- Locally mass conservative due to discontinuous, modal pressure
- Non-conforming hexahedral meshes with “hanging nodes”
- Adaptive mesh refinement (AMR) 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 \( \sim 5 \text{ km} \))
- Wide range of spatial scales and highly localized features, e.g., plate boundaries of size \( O(1 \text{ km}) \) influence plate motion at continental scales of \( O(1000 \text{ km}) \)
- Adaptive mesh refinement is essential
- High-order finite elements \( \mathbb{Q}_k \times P_{k-1} \), order \( k \geq 2 \), with local mass conservation; yields a difficult to deal with discontinuous, modal pressure approximation

→ Developing scalable non-linear & linear solvers and preconditioners took several years: Rudi et al. (2015), Rudi, Stadler, and Ghattas (2017), Rudi, Shih, and Stadler (2020).
Outline

Background

Earth’s Mantle Convection – The Driving Application and Challenges

Inference & Uncertainty Quantification

Numerical Results
  Inference on a cross section of Earth’s mantle
  Inference on the full sphere of the Earth
Challenges of inferring parameters from observations

Data: Challenging because of limited amount

- Current plate motion of rigid plates from GPS and magnetic anomalies recorded in bands perpendicular to seafloor spreading
Challenges of inferring parameters from observations

**Data:** Challenging because of limited amount

- Current plate motion of rigid plates from GPS and magnetic anomalies recorded in bands perpendicular to seafloor spreading

**Model:** Challenging because of computational complexity and truncation/inexact solves

- Incompressible, nonlinear Stokes PDE:
  \[-\nabla \cdot [\mu(x, \dot{\varepsilon}_{II}(u)) (\nabla u + \nabla u^T)] + \nabla p = f, \quad -\nabla \cdot u = 0\]

- Effective viscosity:
  \[\mu(x, \dot{\varepsilon}_{II}(u)) := \mu_{\text{min}} + \min\left(\frac{\tau_{\text{yield}}}{2\dot{\varepsilon}_{II}(u)}, w(x) \min\left(\mu_{\text{max}}, a(T(x))^{\frac{1}{n}} \dot{\varepsilon}_{II}(u)^{\frac{1}{n}}\right)\right)\]
Challenges of inferring parameters from observations

**Data:** Challenging because of limited amount

- Current plate motion of rigid plates from GPS and magnetic anomalies recorded in bands perpendicular to seafloor spreading

**Model:** Challenging because of computational complexity and truncation/inexact solves

- Incompressible, nonlinear Stokes PDE:
  \[-\nabla \cdot [\mu(x, \dot{e}_{II}(u)) (\nabla u + \nabla u^T)] + \nabla p = f, \quad -\nabla \cdot u = 0\]

- Effective viscosity:
  \[\mu(x, \dot{e}_{II}(u)) := \mu_{\text{min}} + \min \left( \frac{\tau_{\text{yield}}}{2\dot{e}_{II}(u)}, w(x) \min \left( \mu_{\text{max}}, a(T(x)) \frac{1}{n} \dot{e}_{II}(u)^{\frac{1}{n}-1} \right) \right)\]

**Parameters:** Challenging because of vastly different scales of sensitivity

- Global parameters: scaling factors, activation energy in Arrhenius law \(a(T(x))\), stress exponent \(n\), yield strength \(\tau_{\text{yield}}\)

- Local **coupling strength** \(w_{\text{min}}\) (i.e., energy dissipation between plates)
Formulate inverse problem in a Bayesian setting

**Given:** (for simplicity $u$ now combines velocity and pressure)

- Model PDE (forward problem): $A(m, u) = f$ (here nonlinear Stokes PDE)
- Map of model output (dependent on parameters $m$) to observations: $F(u(m))$
- Assume data $d$ contains normally distributed additive noise, $N(0, \mathcal{C}_{\text{noise}})$
- Assume prior of the parameters $m$ is normally distributed, $N(m_{\text{pr}}, \mathcal{C}_{\text{pr}})$

**Want:** Description of the posterior density of the parameters (using Bayes’)

$$\pi_{\text{post}}(m) \propto \exp \left( -\frac{1}{2} \| d - F(u(m)) \|^2_{\mathcal{C}_{\text{noise}}^{-1}} - \frac{1}{2} \| m - m_{\text{pr}} \|^2_{\mathcal{C}_{\text{pr}}^{-1}} \right)$$
Formulate inverse problem in a Bayesian setting

**Given:** (for simplicity $u$ now combines velocity and pressure)

- Model PDE (forward problem): $A(m, u) = f$ (here nonlinear Stokes PDE)
- Map of model output (dependent on parameters $m$) to observations: $F(u(m))$
- Assume data $d$ contains normally distributed additive noise, $\mathcal{N}(0, \mathcal{C}_{\text{noise}})$
- Assume prior of the parameters $m$ is normally distributed, $\mathcal{N}(m_{\text{pr}}, \mathcal{C}_{\text{pr}})$

**Want:** Description of the posterior density of the parameters (using Bayes’)

$$
\pi_{\text{post}}(m) \propto \exp \left( -\frac{1}{2} \|d - F(u(m))\|_{\mathcal{C}_{\text{noise}}^{-1}}^2 - \frac{1}{2} \|m - m_{\text{pr}}\|_{\mathcal{C}_{\text{pr}}^{-1}}^2 \right)
$$

**Computationally feasible** (and thanks to Laplace approximation)

- Find the maximum of $\pi_{\text{post}}(m)$ by solving an optimization problem constraint by the model PDE:

  $$
  \arg \min_m \frac{1}{2} \|d - F(u)\|_{\mathcal{C}_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|m - m_{\text{pr}}\|_{\mathcal{C}_{\text{pr}}^{-1}}^2 \quad \text{such that} \quad A(m, u) = f
  $$

- Construct a Gaussian approximation of $\pi_{\text{post}}(m)$ around this maximum by approximating the Hessian of the optimization problem (Gauss–Newton)
Newton (outer loop): Ad joints for 1st & 2nd-order derivatives

Derivation of gradient equations, using a Lagrangian as Ansatz,

1. Solve the (nonlinear) forward problem for $u$: ($u$ combines velocity and pressure for simpler notation)

   $$ (A(m, u), \tilde{v}) = (f, \tilde{v}) \quad \text{for all } \tilde{v} $$

2. Solve the (linear) adjoint problem for $v$:

   $$ (\tilde{u}, \delta_u[A]^*v) = (\delta_u[F](\tilde{u}), d - F(m, u))_{\phi^{-1}_{noise}} \quad \text{for all } \tilde{u} $$

3. Compute the gradient with respect to parameters $m$:

   $$ G(\tilde{m}) = (\delta_m[A](\tilde{m}), v) + (\tilde{m}, m - m_{pr})_{\phi^{-1}_{pr}} - (\delta_m[F](\tilde{m}), d - F(m, u))_{\phi^{-1}_{noise}} $$

Computational complexity

- One (nonlinear) forward + one (linear) adjoint solve
- Independent of the dimension of parameters & data size → scalable
- Analogous approach is used to compute Hessians
Outline

Background

Earth’s Mantle Convection – The Driving Application and Challenges

Inference & Uncertainty Quantification

Numerical Results
  Inference on a cross section of Earth’s mantle
  Inference on the full sphere of the Earth
Setup: Cross section of Earth’s mantle²

Cross section (blue line); velocity vectors from MORVEL56 (green arrows). (Credit: M. Gurnis)

Inference of plate decoupling for a cross section of Earth

Vary data/noise standard deviation

$\sigma_A = 4.0$, $\sigma_B = 1.0$, $\sigma_C = 0.5$ mm/yr

Figure 3.

Results for series B.1. Velocities at the surface from model output after completion of inference, using decreasing standard deviations of the data error $\sigma^2 \{0.5, 1.0, 4.0\}$ mm/yr (bottom to top). Effective viscosity around plate margins (center and right columns). The jumps in effective viscosity which occur at depth correspond to the 410 km and 660 km depths.

For the first set of inverse problems, we vary the standard deviation of the data misfit term, $\sigma^2 \{4, 2, 1, 0.5, 0.25\}$ mm/yr. With a large value of the data error, the estimate for the plate motion data for the larger plates, especially for the large, fast-moving Pacific Plate, generally fits the data well for any of the assumed data errors. But the small back-arc basin for the Mariana subduction zone is not well fit (Fig. 3B,C). When the data error is large, the surface plate motions do not display divergence above the Mariana slab. As the data is reduced, the fit of the velocity of this small Mariana plate improves, especially between data values of 1 and 0.5 mm/yr. During this trend toward resolving the back-arc motion better, there is a substantial transition in the recovered parameters (Fig. 4): The global yield stress drops from about 100 MPa to about 45 MPa, which leads to much more (e.g., broader scale) yielding within the hinge zones of the three slabs. When the yield stress decreases, there is a jump in the weak zone factor for the Mariana from $10^5$ to $10^4$ (Fig. 4C). Essentially, the divergence above the slab, referred to as trench roll-back, causes the Mariana slab to roll back. The slab is able to roll back only if it can easily bend in the hinge zone and a lower effective viscosity in the hinge zone is required (Alisic et al., 2012); consequently, fitting the roll-back in the kinematic data well leads to a global reduction in the yield stress and hence the quite evident sharp reduction of yield stress as data decreases. While it is interesting that the model is capable of fitting the back-arc motion, the results for $\sigma = 0.5$ and 0.25 need to be taken with caution, because the inferred parameters might be a result of underestimating observation and model errors. Hence, the shifted and contracted posteriors for small values of the data arise from setting the noise possibly artificially low.

Furthermore, we observe the similarities of priors and posteriors for $\sigma^2 \{1, 2, 4\}$ and for yield stress and weak zone factor parameters.
Inference of plate decoupling for a cross section of Earth

Sensitivities of the response of inferred plate coupling factors $w_{\text{min}}$ to a (prescribed) accuracy between data and model outputs.

Prior and posterior distributions for $w_{\text{min}}$

- Ryuku weak zone
- Mariana weak zone
- Chile weak zone
Inference for a cross section with plate-dependent weights

Plate size-dependent standard deviation

2D marginals of the (approx.) posterior

Forward sensitivities of stresses (QOIs)
Towards full-sphere inference | current & future work

Data misfit at initial guess of plate coupling parameters $w_{\text{min}}$ at subduction faults and spatially-constant prefactors

Plate motion from MORVEL65 (black arrows) and from model outputs (red arrows). (Credit: J. Hu)
Towards full-sphere inference | current & future work

Data misfit toward optimality point (iteration 9)

Plate motion from MORVEL65 (black arrows) and from model outputs (red arrows). (Credit: J. Hu)
Towards full-sphere inference | current & future work

Approximately inferred plate coupling factors $w_{\text{min}}$

Strong variation of weak zone factors $w_{\text{min}}$ (colors in $\log_{10}$-scale). (Credit: J. Hu)
References I


