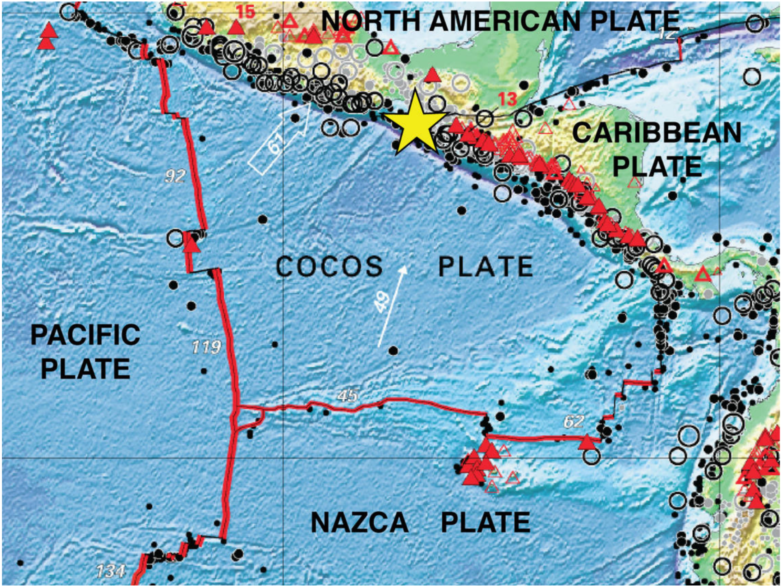
Computing Eigenvalue Sensitivities for Sensitivity Analysis of the Information Gain in Bayesian Linear Inverse Problems

Abhijit Chowdhary and Alen Alexanderian

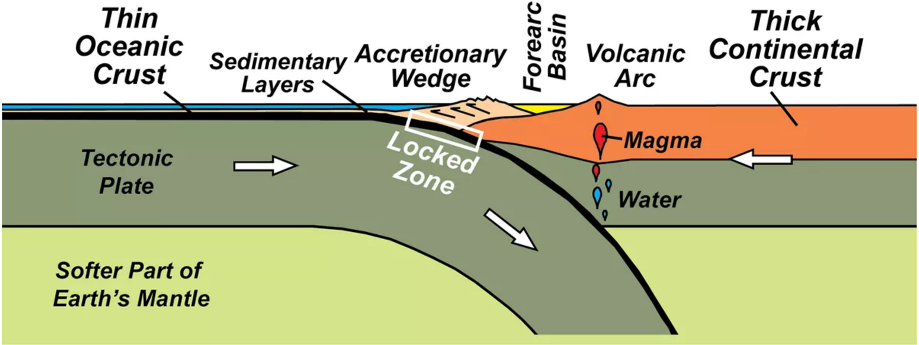
Department of Mathematics
North Carolina State University

September 30, 2022

Work done through NSF-DMS-2111044



Tom Simkin et al. *This dynamic planet: World map of volcanoes, earthquakes, impact craters and plate tectonics*. Tech. rep. 2006. DOI: 10.3133/i2800



Robert J Lillie. *Oregon's Island In The Sky: Geology Road Guide to Marys Peak*. English. OCLC: 979996650. 2017. ISBN: 9781540611963

Goals

Understand the subduction zone from collected observations.

Goals

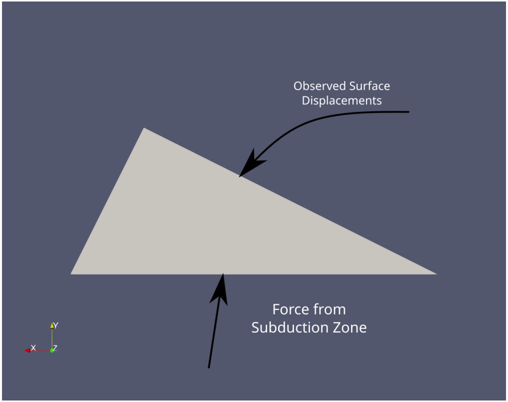
Understand the subduction zone from collected observations.

Do so while quantifying measurement uncertainties.

Table of Contents

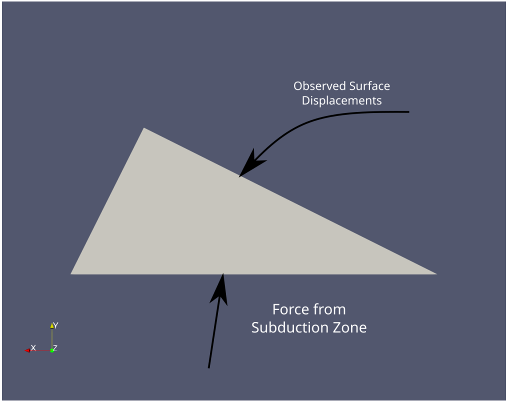
Table of Contents

Model Assumptions



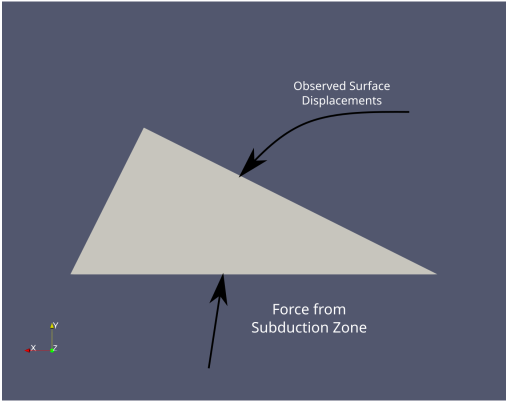
- **Governing PDE** (forward model): Linear elasticity

Model Assumptions



- **Governing PDE** (forward model): Linear elasticity
- **Uncertain parameter**: Displacement on fault plane

Model Assumptions



- **Governing PDE** (forward model): Linear elasticity
- **Uncertain parameter**: Displacement on fault plane
- **Inverse Problem**: Given measurements of surface deformation $\mathbf{u}^{\text{obs}}$ reconstruct fault plane displacement.

Forward Model

$$-\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{u}) = \mathbf{0} \text{ in } \Omega,$$

where:

- $\boldsymbol{\sigma}(\boldsymbol{u}) = \mathbb{C}\boldsymbol{\varepsilon}(\boldsymbol{u})$ with
  - $\mathbb{C}[\boldsymbol{\varepsilon}] = 2\mu\boldsymbol{\varepsilon} + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon})\mathbf{I}$ the fourth-order linear elasticity tensor:
  - $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}\left[\boldsymbol{\nabla}\boldsymbol{u} + (\boldsymbol{\nabla}\boldsymbol{u})^{\mathrm{T}}\right]$ the strain tensor.
- μ and λ are the Láme constants.

Kimberly Alison McCormack. *Earthquakes, groundwater and surface deformation : Exploring the poroelastic response to megathrust earthquakes*. Aug. 2018. URL: hdl.handle.net/2152/68892

Forward Model

$$
\begin{aligned}
-\boldsymbol{\nabla}\left[\mu(\boldsymbol{\nabla}\boldsymbol{u}+(\boldsymbol{\nabla}\boldsymbol{u})^{\mathrm{T}})+\lambda\boldsymbol{\nabla}\cdot\boldsymbol{u}\mathbf{I}\right] &= \mathbf{0} \quad \text{in } \Omega,\\
\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \mathbf{0} \quad \text{on } \Gamma_t\\
\boldsymbol{u}+\beta\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{h} \quad \text{on } \Gamma_s\\
\boldsymbol{u}\cdot\boldsymbol{n} &= 0 \quad \text{on } \Gamma_b\\
\delta\mathbf{T}(\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n})+\mathbf{T}\boldsymbol{u} &= \boldsymbol{m} \quad \text{on } \Gamma_b
\end{aligned}
$$

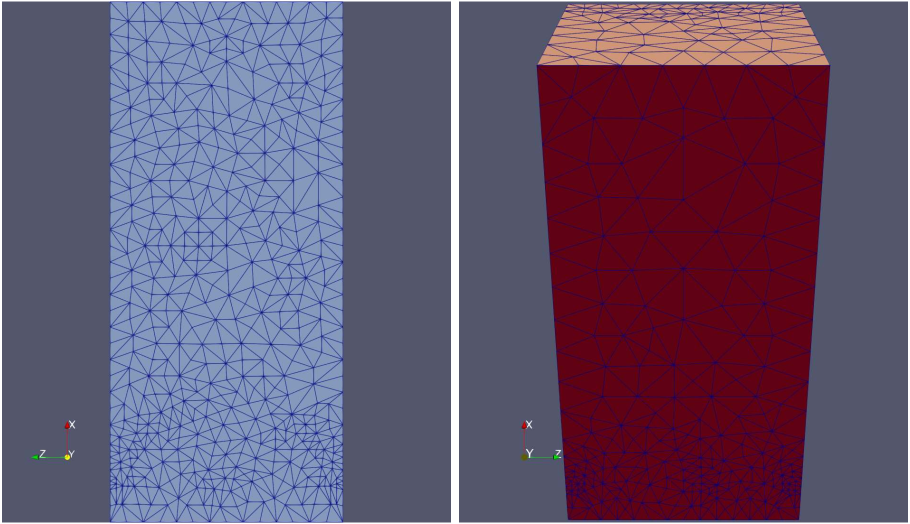
- $\mathbf{T}$ is the tangential operator $\mathbf{T}\boldsymbol{u}=(\mathbf{I}-\boldsymbol{n}\otimes\boldsymbol{n})\boldsymbol{u}=\boldsymbol{u}-(\boldsymbol{n}^{\mathrm{T}}\boldsymbol{u})\boldsymbol{n}$.

Forward Model

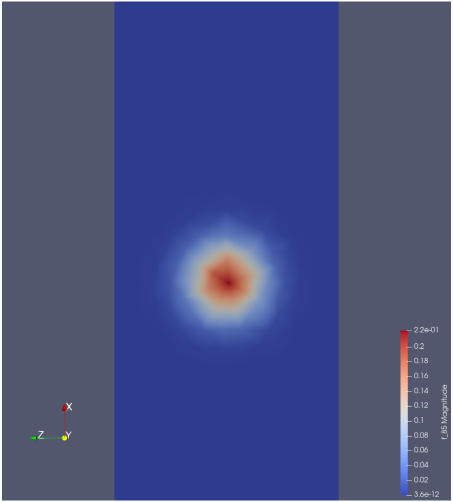
$$
\begin{aligned}
-\boldsymbol{\nabla}\left[\mu(\boldsymbol{\nabla}\boldsymbol{u} + (\boldsymbol{\nabla}\boldsymbol{u})^{\mathrm{T}}) + \lambda\boldsymbol{\nabla}\cdot\boldsymbol{u}\mathbf{I}\right] &= \mathbf{0} \quad \text{in } \Omega, \\
\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \mathbf{0} \quad \text{on } \Gamma_t \\
\boldsymbol{u} + \beta\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{h} \quad \text{on } \Gamma_s \\
\boldsymbol{u}\cdot\boldsymbol{n} &= 0 \quad \text{on } \Gamma_b \\
\delta\mathbf{T}(\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n}) + \mathbf{T}\boldsymbol{u} &= \boldsymbol{m} \quad \text{on } \Gamma_b
\end{aligned}
$$

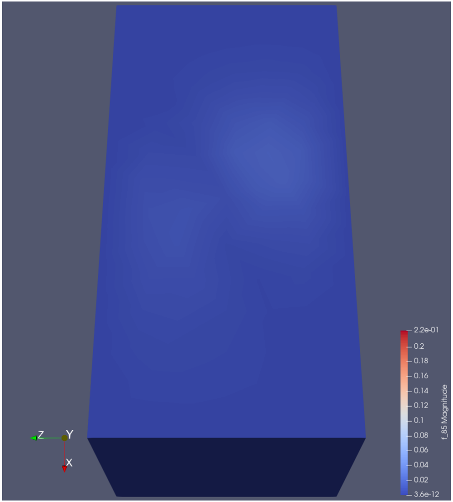
- $\mathbf{T}$ is the tangential operator $\mathbf{T}\boldsymbol{u} = (\mathbf{I} - \boldsymbol{n}\otimes\boldsymbol{n})\boldsymbol{u} = \boldsymbol{u} - (\boldsymbol{n}^{\mathrm{T}}\boldsymbol{u})\boldsymbol{n}$.
- $\boldsymbol{m}$ is the displacement on the fault plane that is being inverted for.

Mesh

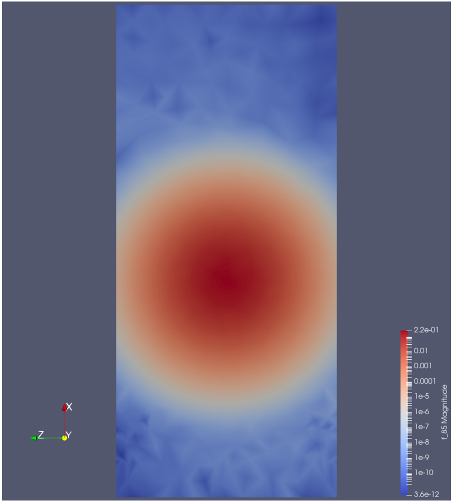


Forward Solution





Forward Solution



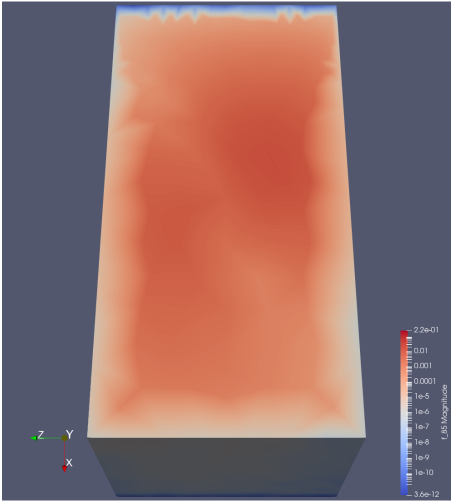


Table of Contents

The Deterministic Inverse Problem

To reconstruct the fault displacement we construct the PDE-constrained optimization problem:

$$\mathcal{J}(\boldsymbol{m}) = \frac{1}{2}\|\boldsymbol{\mathcal{B}}\boldsymbol{u}(\boldsymbol{m}) - \mathbf{u}^{\mathrm{obs}}\|^2 + \frac{1}{2}\|\mathcal{R}\boldsymbol{m}\|^2$$

where $\boldsymbol{u}$ is given by the solution of the linear elasticity equation.

- $\boldsymbol{u}(\boldsymbol{m})$ is given by the forward model.
- $\boldsymbol{\mathcal{B}} : (L^2(\Omega))^3 \to \mathbb{R}^N$ is an observation operator.
- $\mathbf{u}^{\mathrm{obs}} \in \mathbb{R}^N$ where N is the number of data points.
- $\mathcal{R}$ is some regularization operator.

[1] In literature, called the *parameter-to-observable operator*.

The Deterministic Inverse Problem

To reconstruct the fault displacement we construct the PDE-constrained optimization problem:

$$\mathcal{J}(\boldsymbol{m}) = \frac{1}{2}\|\boldsymbol{\mathcal{B}}\boldsymbol{u}(\boldsymbol{m}) - \mathbf{u}^{\text{obs}}\|^2 + \frac{1}{2}\|\mathcal{R}\boldsymbol{m}\|^2$$

where $\boldsymbol{u}$ is given by the solution of the linear elasticity equation.

- $\boldsymbol{u}(\boldsymbol{m})$ is given by the forward model.
- $\boldsymbol{\mathcal{B}} : (L^2(\Omega))^3 \to \mathbb{R}^N$ is an observation operator.
- $\mathbf{u}^{\text{obs}} \in \mathbb{R}^N$ where N is the number of data points.
- $\mathcal{R}$ is some regularization operator.

If we let $\boldsymbol{\mathcal{S}}$ be the forward model operator, i.e. $\boldsymbol{u} = \boldsymbol{\mathcal{S}}\boldsymbol{m}$, and let[1] $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{S}}$, then:

$$\mathcal{J}(\boldsymbol{m}) = \frac{1}{2}\|\boldsymbol{\mathcal{F}}\boldsymbol{m} - \mathbf{u}^{\text{obs}}\|^2 + \frac{1}{2}\|\mathcal{R}\boldsymbol{m}\|^2$$

[1] In literature, called the *parameter-to-observable operator*.

Bayesian Inversion in Infinite Dimensions

Theorem (Bayes Theorem in Infinite Dimensions)

$$\frac{\mathrm{d}\mu_{\text{post}}^{\mathbf{u}^{\text{obs}}}}{\mathrm{d}\mu_{\text{pr}}} \propto \pi_{\text{like}}(\mathbf{u}^{\text{obs}}|\boldsymbol{m})$$

- Gaussian Prior $\boldsymbol{m} \sim \mathcal{N}(\boldsymbol{m}_{\text{pr}}, \boldsymbol{\mathcal{C}}_0)$.
- Additive Gaussian noise

$$\mathbf{u}^{\text{obs}} = \boldsymbol{\mathcal{F}}\boldsymbol{m} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$$

- Gaussian posterior:

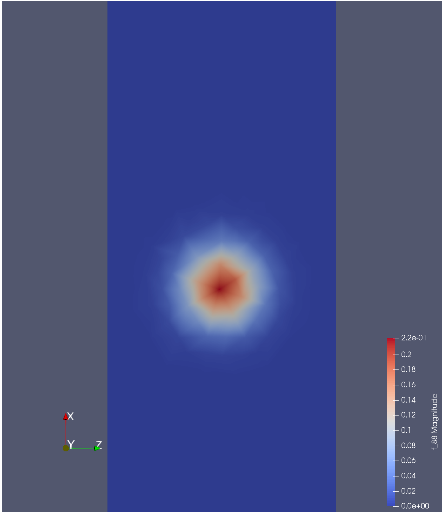
$$\mu_{\text{post}}^{\mathbf{u}^{\text{obs}}} = \mathcal{N}(\boldsymbol{m}_{\text{post}}, \boldsymbol{\mathcal{C}}_{\text{post}})$$

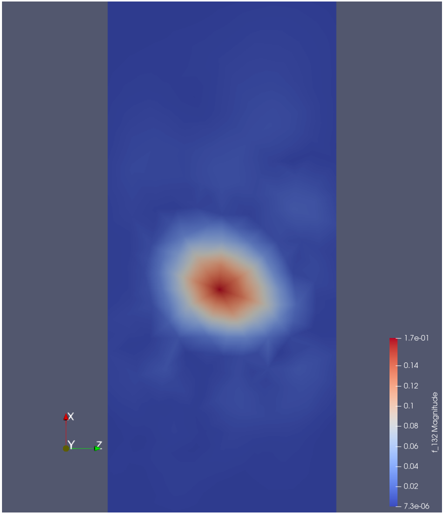
where:

$$\boldsymbol{m}_{\text{post}} = \boldsymbol{\mathcal{C}}_{\text{post}}\left(\boldsymbol{\mathcal{F}}^*\boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathbf{u}^{\text{obs}} + \boldsymbol{\mathcal{C}}_{\text{pr}}^{-1}\boldsymbol{m}_{\text{pr}}\right)$$

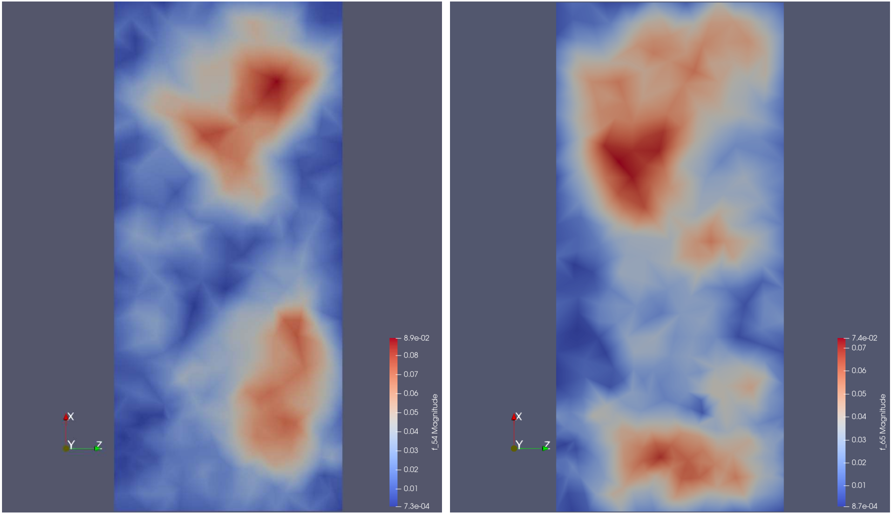
$$\boldsymbol{\mathcal{C}}_{\text{post}} = \left(\boldsymbol{\mathcal{F}}^*\boldsymbol{\Gamma}_{\text{noise}}^{-1}\boldsymbol{\mathcal{F}} + \boldsymbol{\mathcal{C}}_{\text{pr}}^{-1}\right)^{-1}$$

MAP Estimation vs. True Parameter

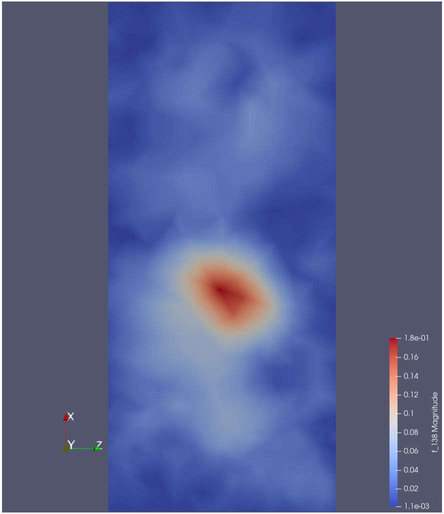




Prior Samples



Posterior Samples



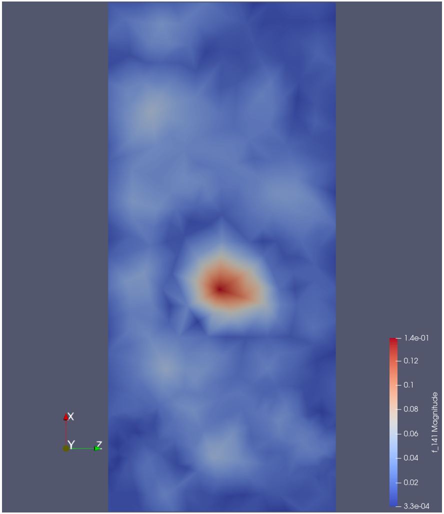


Table of Contents

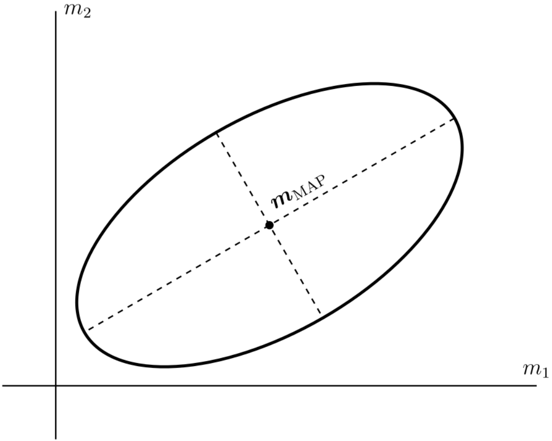


Figure: The confidence region corresponding to a linear inverse problem with a bivariate Gaussian posterior.

Data-Misfit Hessian

Definition (Data-Misfit Hessian)

The action of the data-misfit Hessian is given by:

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{F}}^* \boldsymbol{\Gamma}_{\text{noise}}^{-1} \boldsymbol{\mathcal{F}}$$

Data-Misfit Hessian

Definition (Data-Misfit Hessian)

The action of the data-misfit Hessian is given by:

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{F}}^* \boldsymbol{\Gamma}_{\mathrm{noise}}^{-1} \boldsymbol{\mathcal{F}}$$

Definition (Prior-Preconditioned Data-Misfit Hessian)

$$\tilde{\boldsymbol{\mathcal{H}}} = \boldsymbol{\mathcal{C}}_{\mathrm{prior}}^{1/2} \boldsymbol{\mathcal{H}} \boldsymbol{\mathcal{C}}_{\mathrm{prior}}^{1/2}$$

Let $\{(\lambda_i, \psi_i)\}_{i=1}^k$ be the dominant eigenpairs of $\tilde{\boldsymbol{\mathcal{H}}}$, then:

$$\tilde{\boldsymbol{\mathcal{H}}} v \approx \sum_{i=1}^{k} \lambda_i \langle \psi_i, v \rangle \psi_i$$

is a low rank approximation of $\tilde{\boldsymbol{\mathcal{H}}}$.

A Measure of Posterior Uncertainty

Definition (Kullback-Leibler Divergence)

The KL-Divergence from the posterior to the prior quantifies the information gain and in the context of Bayesian linear inverse problems has the form:

$$\Phi_{\mathrm{KL}} = \frac{1}{2}\left[\log\det\left(\tilde{\boldsymbol{\mathcal{H}}} + \boldsymbol{\mathcal{I}}\right) - \mathrm{trace}\left(\tilde{\boldsymbol{\mathcal{H}}}(\boldsymbol{\mathcal{I}} + \tilde{\boldsymbol{\mathcal{H}}})^{-1}\right) + \left\|\boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}\right]$$
$$\approx \frac{1}{2}\sum_{j=1}^{k}\left[\log\left(\lambda_j + 1\right) - \frac{\lambda_j}{1+\lambda_j}\right] + \frac{1}{2}\left\|\boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}$$

A Measure of Posterior Uncertainty

Definition (Kullback-Leibler Divergence)

The KL-Divergence from the posterior to the prior quantifies the information gain and in the context of Bayesian linear inverse problems has the form:

$$\Phi_{\mathrm{KL}} = \frac{1}{2}\left[\log\det\left(\tilde{\boldsymbol{\mathcal{H}}} + \boldsymbol{\mathcal{I}}\right) - \mathrm{trace}\left(\tilde{\boldsymbol{\mathcal{H}}}(\boldsymbol{\mathcal{I}} + \tilde{\boldsymbol{\mathcal{H}}})^{-1}\right) + \left\|\boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}\right]$$

$$\approx \frac{1}{2}\sum_{j=1}^{k}\left[\log\left(\lambda_j + 1\right) - \frac{\lambda_j}{1+\lambda_j}\right] + \frac{1}{2}\left\|\boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}$$

To assess the sensitivity of a Bayesian inverse problem with respect to modeling uncertaintities, we compute partial derivative of Φ_{KL} with respect to said model parameters.

A Measure of Posterior Uncertainty

Definition (Kullback-Leibler Divergence)

The KL-Divergence from the posterior to the prior quantifies the information gain and in the context of Bayesian linear inverse problems has the form:

$$\Phi_{\mathrm{KL}} = \frac{1}{2}\left[\log\det\left(\tilde{\mathcal{H}} + \mathcal{I}\right) - \mathrm{trace}\left(\tilde{\mathcal{H}}(\mathcal{I} + \tilde{\mathcal{H}})^{-1}\right) + \left\|\boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}\right]$$
$$\approx \frac{1}{2}\sum_{j=1}^{k}\left[\log\left(\lambda_j + 1\right) - \frac{\lambda_j}{1+\lambda_j}\right] + \frac{1}{2}\left\|\boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}}\right\|^2_{\mathcal{C}_{\mathrm{pr}}^{-1}}$$

To assess the sensitivity of a Bayesian inverse problem with respect to modeling uncertaintities, we compute partial derivative of Φ_{KL} with respect to said model parameters.

Q: How do we differentiate this expression?

Q: How do we differentiate this expression?

$$\Phi_{\mathrm{KL}} =\approx \sum_{j=1}^{k} \frac{1}{2} \left[\underbrace{\log\left(\lambda_j + 1\right) - \frac{\lambda_j}{1+\lambda_j}}_{\text{Eigenvalue Sensitivities}} + \underbrace{\left\| \boldsymbol{m}_{\mathrm{MAP}} - \boldsymbol{m}_{\mathrm{pr}} \right\|^2_{\boldsymbol{\Gamma}^{-1}_{\mathrm{noise}}}}_{\text{HDSA}} \right]$$

Isaac Sunseri et al. "Hyper-differential sensitivity analysis for inverse problems constrained by partial differential equations". In: *Inverse Problems* 36.12 (Dec. 2020), p. 125001. DOI: 10.1088/1361-6420/abaf63

Table of Contents

Simple Finite Dimensional Example

Consider an implicitly defined matrix $\mathbf{H}$ whose action is defined by:

$$\mathbf{H}\mathbf{v} := \mathbf{C}^{\mathrm{T}}\mathbf{p} \tag{1a}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0} \tag{1b}$$

$$\mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0} \tag{1c}$$

Furthermore, consider $\tilde{\mathbf{H}} = \mathbf{\Gamma}^{1/2}\mathbf{H}\mathbf{\Gamma}^{1/2}$ with $\mathbf{\Gamma}$ symmetric positive definite.

Simple Finite Dimensional Example

Consider an implicitly defined matrix $\mathbf{H}$ whose action is defined by:

$$\mathbf{H}\mathbf{v} := \mathbf{C}^{\mathrm{T}}\mathbf{p} \tag{1a}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0} \tag{1b}$$

$$\mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0} \tag{1c}$$

Furthermore, consider $\tilde{\mathbf{H}} = \mathbf{\Gamma}^{1/2}\mathbf{H}\mathbf{\Gamma}^{1/2}$ with $\mathbf{\Gamma}$ symmetric positive definite.

To illustrate eigenvalue sensitivity analysis, we consider sensitivity of the largest eigenvalue of $\tilde{\mathbf{H}}$ with respect to θ.

Simple Finite Dimensional Example

Consider an implicitly defined matrix $\mathbf{H}$ whose action is defined by:

$$\mathbf{H}\mathbf{v} := \mathbf{C}^{\mathrm{T}}\mathbf{p} \tag{1a}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0} \tag{1b}$$

$$\mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0} \tag{1c}$$

Furthermore, consider $\tilde{\mathbf{H}} = \mathbf{\Gamma}^{1/2}\mathbf{H}\mathbf{\Gamma}^{1/2}$ with $\mathbf{\Gamma}$ symmetric positive definite.

To illustrate eigenvalue sensitivity analysis, we consider sensitivity of the largest eigenvalue of $\tilde{\mathbf{H}}$ with respect to θ.

Note:

Let $(\lambda, \mathbf{w})$ be the dominant eigenpair of $\tilde{\mathbf{H}}$, then $\mathbf{H}\mathbf{v} = \lambda\mathbf{\Gamma}^{-1}\mathbf{v}$ with $\mathbf{v} = \mathbf{\Gamma}^{1/2}\mathbf{w}$.

Derivation

We can define λ as a function of uncertain model parameters implicitly:

$$\lambda(\theta) = \mathbf{v}^{\mathrm{T}} \mathbf{C}^{\mathrm{T}} \mathbf{p}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$$

$$\mathbf{A}(\theta)^{\top}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0}$$

$$\mathbf{v}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{v} - 1 = 0$$

Derivation

We can define λ as a function of uncertain model parameters implicitly:

$$\lambda(\theta) = \mathbf{v}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{p}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$$

$$\mathbf{A}(\theta)^{\top}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0}$$

$$\mathbf{v}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{v} - 1 = 0$$

Definition

To facilitate derivative computation, we consider the "meta-Lagrangian":

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p} \rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} \rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} \rangle + \lambda^* \left(1 - \mathbf{v}^{\top}\boldsymbol{\Gamma}^{-1}\mathbf{v}\right)$$

Solving for Lagrange Multipliers

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p} \rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} \rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} \rangle + \lambda^* \left(1 - \mathbf{v}^\top \mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

$$\begin{aligned}
\mathcal{L}_{\mathbf{u}} &= \mathbf{A}^\top \mathbf{p}^* + \mathbf{B}^\top \mathbf{B}\mathbf{u}^* = 0 \\
\mathcal{L}_{\mathbf{p}} &= \mathbf{A}\mathbf{u}^* + \mathbf{C}\mathbf{v} = 0 \\
\mathcal{L}_{\mathbf{v}} &= 2\mathbf{C}^\top \mathbf{p} - 2\lambda^* \mathbf{\Gamma}^{-1}\mathbf{v} = 0
\end{aligned}$$

Solving for Lagrange Multipliers

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p} \rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} \rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} \rangle + \lambda^* \left(1 - \mathbf{v}^\top \mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

$$\mathcal{L}_{\mathbf{u}} = \mathbf{A}^\top \mathbf{p}^* + \mathbf{B}^\top \mathbf{B}\mathbf{u}^* = 0$$

$$\mathcal{L}_{\mathbf{p}} = \mathbf{A}\mathbf{u}^* + \mathbf{C}\mathbf{v} = 0$$

$$\mathcal{L}_{\mathbf{v}} = 2\mathbf{C}^\top \mathbf{p} - 2\lambda^* \mathbf{\Gamma}^{-1}\mathbf{v} = 0$$

That is, $\mathbf{u}^* = \mathbf{u}$ and $\mathbf{p}^* = \mathbf{p}$.

$$\lambda(\theta) = \mathbf{v}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{p}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$$

$$\mathbf{A}(\theta)^\top \mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0}$$

$$\mathbf{v}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{v} - 1 = 0$$

Solving for Lagrange Multipliers

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p} \rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} \rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} \rangle + \lambda^* \left(1 - \mathbf{v}^\top \mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

$$\mathcal{L}_{\mathbf{u}} = \mathbf{A}^\top \mathbf{p}^* + \mathbf{B}^\top \mathbf{B}\mathbf{u}^* = 0$$
$$\mathcal{L}_{\mathbf{p}} = \mathbf{A}\mathbf{u}^* + \mathbf{C}\mathbf{v} = 0$$
$$\mathcal{L}_{\mathbf{v}} = 2\mathbf{C}^\top \mathbf{p} - 2\lambda^* \mathbf{\Gamma}^{-1}\mathbf{v} = 0$$

$$\lambda(\theta) = \mathbf{v}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{p}$$

where

$$\mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$$
$$\mathbf{A}(\theta)^\top \mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0}$$
$$\mathbf{v}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{v} - 1 = 0$$

That is, $\mathbf{u}^* = \mathbf{u}$ and $\mathbf{p}^* = \mathbf{p}$. Regarding λ^*:

$$0 = 2\mathbf{v}^{\mathrm{T}}\mathbf{C}^\top \mathbf{p} - 2\lambda^* \mathbf{v}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{v} = 2\lambda - 2\lambda^* \implies \lambda = \lambda^*$$

Eigenvalue Sensitivity

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p} \rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} \rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} \rangle + \lambda^* \left(1 - \mathbf{v}^\top \mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

Finally, $\mathcal{L}_{\theta_j}$ simplifies to

$$\mathcal{L}_{\theta_j} = \langle \mathbf{p}^*, [\partial_j \mathbf{A}]\mathbf{u} \rangle + \langle \mathbf{u}^*, [\partial_j \mathbf{A}]^\top \mathbf{p} \rangle = 2\langle \mathbf{p}, [\partial_j \mathbf{A}]\mathbf{u} \rangle$$

Eigenvalue Sensitivity

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p}\rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v}\rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u}\rangle + \lambda^* \left(1 - \mathbf{v}^{\top}\mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

Finally, $\mathcal{L}_{\theta_j}$ simplifies to

$$\mathcal{L}_{\theta_j} = \langle \mathbf{p}^*, [\partial_j \mathbf{A}]\mathbf{u}\rangle + \langle \mathbf{u}^*, [\partial_j \mathbf{A}]^{\top}\mathbf{p}\rangle = 2\langle \mathbf{p}, [\partial_j \mathbf{A}]\mathbf{u}\rangle$$

Therefore, to compute the sensitivity of λ with respect to θ_j, we have the following algorithm:

- solve the eigenproblem for eigenvector $\mathbf{v}$

Eigenvalue Sensitivity

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p}\rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v}\rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u}\rangle + \lambda^* \left(1 - \mathbf{v}^\top \mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

Finally, $\mathcal{L}_{\theta_j}$ simplifies to

$$\mathcal{L}_{\theta_j} = \langle \mathbf{p}^*, [\partial_j \mathbf{A}]\mathbf{u}\rangle + \langle \mathbf{u}^*, [\partial_j \mathbf{A}]^\top \mathbf{p}\rangle = 2\langle \mathbf{p}, [\partial_j \mathbf{A}]\mathbf{u}\rangle$$

Therefore, to compute the sensitivity of λ with respect to θ_j, we have the following algorithm:

- solve the eigenproblem for eigenvector $\mathbf{v}$
- solve $\mathbf{A}\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$ for $\mathbf{u}$

Eigenvalue Sensitivity

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p} \rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v} \rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} \rangle + \lambda^* \left(1 - \mathbf{v}^{\top}\mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

Finally, $\mathcal{L}_{\theta_j}$ simplifies to

$$\mathcal{L}_{\theta_j} = \langle \mathbf{p}^*, [\partial_j \mathbf{A}]\mathbf{u} \rangle + \langle \mathbf{u}^*, [\partial_j \mathbf{A}]^{\top}\mathbf{p} \rangle = 2\langle \mathbf{p}, [\partial_j \mathbf{A}]\mathbf{u} \rangle$$

Therefore, to compute the sensitivity of λ with respect to θ_j, we have the following algorithm:

- solve the eigenproblem for eigenvector $\mathbf{v}$
- solve $\mathbf{A}\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$ for $\mathbf{u}$
- solve $\mathbf{A}^{\mathrm{T}}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0}$ for $\mathbf{p}$

Eigenvalue Sensitivity

$$\mathcal{L} = \langle \mathbf{v}, \mathbf{C}^{\mathrm{T}}\mathbf{p}\rangle + \langle \mathbf{p}^*, \mathbf{A}(\theta)\mathbf{u} + \mathbf{C}\mathbf{v}\rangle + \langle \mathbf{u}^*, \mathbf{A}(\theta)^{\mathrm{T}}\mathbf{p} - \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u}\rangle + \lambda^* \left(1 - \mathbf{v}^\top \mathbf{\Gamma}^{-1}\mathbf{v}\right)$$

Finally, $\mathcal{L}_{\theta_j}$ simplifies to

$$\mathcal{L}_{\theta_j} = \langle \mathbf{p}^*, [\partial_j \mathbf{A}]\mathbf{u}\rangle + \langle \mathbf{u}^*, [\partial_j \mathbf{A}]^\top \mathbf{p}\rangle = 2\langle \mathbf{p}, [\partial_j \mathbf{A}]\mathbf{u}\rangle$$

Therefore, to compute the sensitivity of λ with respect to θ_j, we have the following algorithm:

- solve the eigenproblem for eigenvector $\mathbf{v}$
- solve $\mathbf{A}\mathbf{u} + \mathbf{C}\mathbf{v} = \mathbf{0}$ for $\mathbf{u}$
- solve $\mathbf{A}^{\mathrm{T}}\mathbf{p} + \mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{u} = \mathbf{0}$ for $\mathbf{p}$
- evaluate eigenvalue sensitivities $\partial_j \lambda(\theta) = 2\langle \mathbf{p}, [\partial_j \mathbf{A}]\mathbf{u}\rangle, \quad j \in \{1, \ldots, N\}$

Simple Numerical Example

$$\mathbf{A} = \begin{bmatrix} -20(1+a_2\theta_2) & 3(1+a_3\theta_3) & 0 \\ 1+a_1\theta_1 & -20(1+a_2\theta_2) & 3(1+a_3\theta_3) \\ 0 & 1+a_1\theta_1 & -20(1+a_2\theta_2) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 0.08 & 0.06 & 0.09 \end{bmatrix}^T$$

Also, the remaining matrices for the present example are as follows:

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In this case, the matrix $\mathbf{H}$ is a 2×2 matrix.

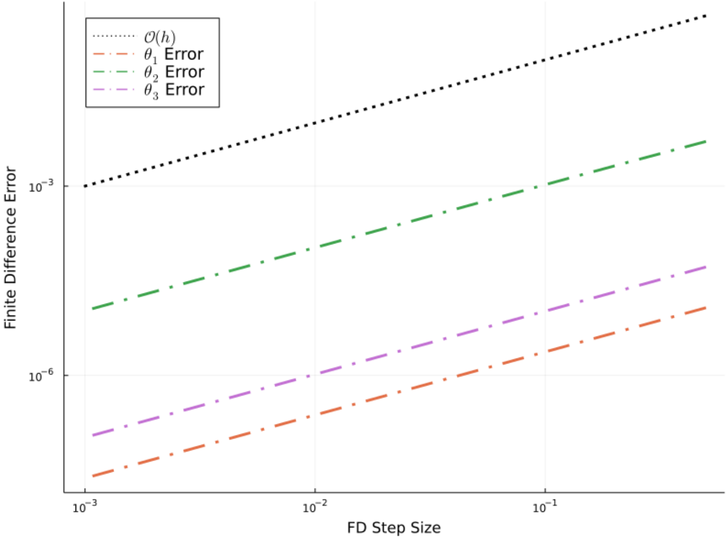


Figure: Forward finite difference error vs. exact

Table of Contents

Simple Infinite Dimensional Example

For a domain $\Omega \subset \mathbb{R}^n$ with boundary $\partial\Omega$, consider the elliptic PDE:

$$\begin{aligned} -\nabla^2 u + cu &= f, \quad \text{in } \Omega \\ \boldsymbol{\nabla} u \cdot n &= g, \quad \text{in } \partial\Omega \end{aligned}$$

Bayesian Linear Inverse Problem

Using measurements of u on Ω, we seek to estimate f.

Simple Infinite Dimensional Example

For a domain $\Omega \subset \mathbb{R}^n$ with boundary $\partial\Omega$, consider the elliptic PDE:

$$\begin{aligned} -\nabla^2 u + cu &= f, \quad \text{in } \Omega \\ \boldsymbol{\nabla} u \cdot n &= g, \quad \text{in } \partial\Omega \end{aligned}$$

Bayesian Linear Inverse Problem

Using measurements of u on Ω, we seek to estimate f.

Goal:

Study the sensitivity of the largest eigenvalue of the data-misfit Hessian for the inversion with respect to constant c.

Simple Infinite Dimensional Example

For a domain $\Omega \subset \mathbb{R}^n$ with boundary $\partial\Omega$, consider the elliptic PDE:

$$\begin{aligned} -\nabla^2 u + cu &= f, \quad \text{in } \Omega \\ \boldsymbol{\nabla} u \cdot n &= g, \quad \text{in } \partial\Omega \end{aligned}$$

Bayesian Linear Inverse Problem

Using measurements of u on Ω, we seek to estimate f.

Goal:

Study the sensitivity of the largest eigenvalue of the data-misfit Hessian for the inversion with respect to constant c.

$$\mathcal{L}(u, p, f) = \frac{1}{2}\left\|\mathcal{B}u(f) - \mathbf{u}^{\mathrm{obs}}\right\|^2_{\boldsymbol{\Gamma}^{-1}_{\mathrm{noise}}} + \frac{1}{2}\left\|f - f_{\mathrm{pr}}\right\|^2_{\mathcal{C}_0^{-1}}$$

$$\int_\Omega \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} p \,\mathrm{d}x + c\int_\Omega up\,\mathrm{d}x - \int_\Omega fp\,\mathrm{d}x - \int_{\partial\Omega} gp\,\mathrm{d}s$$

The adjoint based expression the data-misfit Hessian action

Taking variations in u, p and f twice, we find the data-misfit Hessian:

$$\mathcal{L}_f^H(\tilde{f}) = \underbrace{-\int_\Omega \tilde{f}\hat{p}\,\mathrm{d}x}_{\mathcal{H}(f)(\hat{f},\tilde{f})} + \langle \tilde{f}, \hat{f}\rangle_{\mathcal{C}_0^{-1}}$$

which, to evaluate in direction $\hat{f}$, we need to solve the incremental system:

$$\begin{aligned}
\mathcal{L}_v^H(\tilde{p}) &= \int_\Omega \nabla\hat{u}\cdot\nabla\tilde{p}\,\mathrm{d}x + c\int_\Omega \hat{u}\tilde{p}\,\mathrm{d}x - \int_\Omega \hat{f}\tilde{p}\,\mathrm{d}x &&= 0, \quad \forall\tilde{p}\in V\\
\mathcal{L}_u^H(\tilde{u}) &= \int_\Omega \nabla\tilde{u}\cdot\nabla\hat{p}\,\mathrm{d}x + c\int_\Omega \tilde{u}\hat{p}\,\mathrm{d}x + \langle\mathcal{B}\tilde{u}, \mathcal{B}\hat{u}\rangle_{\boldsymbol{\Gamma}_{\mathrm{noise}}^{-1}} &&= 0, \quad \forall\tilde{u}\in V
\end{aligned}$$

Eigenproblem Constraint

$$
\begin{aligned}
\int_\Omega \nabla \hat{u} \cdot \nabla \tilde{p}\,\mathrm{d}x + c\int_\Omega \hat{u}\tilde{p}\,\mathrm{d}x - \int_\Omega \psi \tilde{p}\,\mathrm{d}x &= 0, && \forall \tilde{p} \in V\\
\int_\Omega \nabla \tilde{u} \cdot \nabla \hat{p}\,\mathrm{d}x + c\int_\Omega \tilde{u}\hat{p}\,\mathrm{d}x + \langle \mathcal{B}\tilde{u}, \mathcal{B}\hat{u}\rangle_{\boldsymbol{\Gamma}_{\mathrm{noise}}^{-1}} &= 0, && \forall \tilde{u} \in V\\
-\int_\Omega \phi\hat{p}\,\mathrm{d}x &= \lambda\langle \phi, \psi\rangle_{\mathcal{C}_0^{-1}}, && \forall \phi \in V\\
1 - \langle \psi, \psi\rangle_{\mathcal{C}_0^{-1}} &= 0
\end{aligned}
$$

Eigenproblem Constraint

$$
\begin{aligned}
\int_\Omega \nabla \hat{u} \cdot \nabla \tilde{p} \,\mathrm{d}x + c \int_\Omega \hat{u}\tilde{p} \,\mathrm{d}x - \int_\Omega \psi \tilde{p} \,\mathrm{d}x &= 0, && \forall \tilde{p} \in V \\
\int_\Omega \nabla \tilde{u} \cdot \nabla \hat{p} \,\mathrm{d}x + c \int_\Omega \tilde{u}\hat{p} \,\mathrm{d}x + \langle \mathcal{B}\tilde{u}, \mathcal{B}\hat{u} \rangle_{\boldsymbol{\Gamma}_{\mathrm{noise}}^{-1}} &= 0, && \forall \tilde{u} \in V \\
- \int_\Omega \phi \hat{p} \,\mathrm{d}x &= \lambda \langle \phi, \psi \rangle_{\mathcal{C}_0^{-1}}, && \forall \phi \in V \\
1 - \langle \psi, \psi \rangle_{\mathcal{C}_0^{-1}} &= 0
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}^E(\hat{u}, \hat{p}, \psi, \hat{u}^*, \hat{p}^*, \lambda^*) &= \left[- \int_\Omega \psi \hat{p} \,\mathrm{d}x \right] + \lambda^* \left[1 - \langle \psi, \psi \rangle_{\mathcal{C}_0^{-1}} \right] \\
&+ \left[\int_\Omega \nabla \hat{u} \cdot \nabla \hat{p}^* \,\mathrm{d}x + c \int_\Omega \hat{u}\hat{p}^* \,\mathrm{d}x - \int_\Omega \psi \hat{p}^* \,\mathrm{d}x \right] \\
&+ \left[\int_\Omega \nabla \hat{u}^* \cdot \nabla \hat{p} \,\mathrm{d}x + c \int_\Omega \hat{u}^* \hat{p} \,\mathrm{d}x + \langle \mathcal{B}\hat{u}^*, \mathcal{B}\hat{u} \rangle_{\boldsymbol{\Gamma}_{\mathrm{noise}}^{-1}} \right]
\end{aligned}
$$

Solving for Lagrange Multipliers

$$
\begin{aligned}
\mathcal{L}^E_{\hat{p}}(\tilde{p}) &= \int_\Omega \nabla \hat{u}^* \cdot \nabla \tilde{p} \,\mathrm{d}x + c \int_\Omega \hat{u}^* \tilde{p} \,\mathrm{d}x - \int_\Omega \psi \tilde{p} \,\mathrm{d}x && = 0, \forall \tilde{p} \in V \\
\mathcal{L}^E_{\hat{u}}(\tilde{u}) &= \int_\Omega \nabla \tilde{u} \cdot \nabla \hat{p}^* \,\mathrm{d}x + c \int_\Omega \tilde{u} \hat{p}^* \,\mathrm{d}x + \langle \mathcal{B}\hat{u}^*, \mathcal{B}\tilde{u} \rangle_{\Gamma^{-1}_{\mathrm{noise}}} && = 0, \forall \tilde{u} \in V \\
\mathcal{L}^E_{\psi}(\tilde{\psi}) &= -2 \int_\Omega \tilde{\psi} \hat{p} \,\mathrm{d}x - 2\lambda^* \langle \psi, \tilde{\psi} \rangle_{\mathcal{C}_0^{-1}} && = 0, \forall \tilde{\psi} \in V
\end{aligned}
$$

Solving for Lagrange Multipliers: Relating it Back

$$\mathcal{L}^E_{\hat{p}}(\tilde{p}) = \underbrace{\int_\Omega \nabla \hat{u}^* \cdot \nabla \tilde{p} \, \mathrm{d}x + c \int_\Omega \hat{u}^* \tilde{p} \, \mathrm{d}x}_{\tilde{\mathbf{p}}^{\mathrm{T}} \mathbf{A} \hat{\mathbf{u}}^*} - \underbrace{\int_\Omega \psi \tilde{p} \, \mathrm{d}x}_{\tilde{\mathbf{p}}^{\mathrm{T}} \mathbf{C} \psi} = 0, \forall \tilde{p} \in V$$

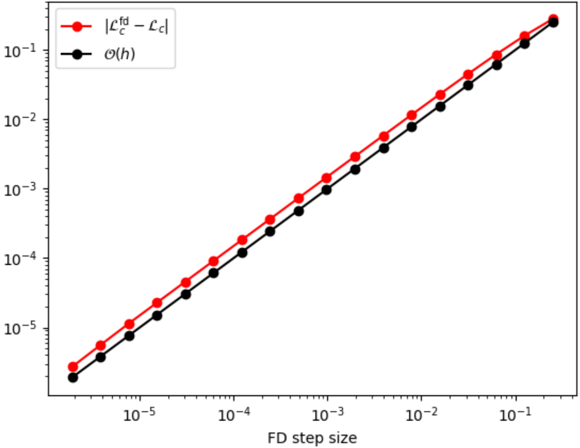
$$\mathcal{L}^E_{\hat{u}}(\tilde{u}) = \underbrace{\int_\Omega \nabla \tilde{u} \cdot \nabla \hat{p}^* \, \mathrm{d}x + c \int_\Omega \tilde{u} \hat{p}^* \, \mathrm{d}x}_{\tilde{\mathbf{u}}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \hat{\mathbf{p}}^*} + \langle \mathcal{B} \hat{u}^*, \mathcal{B} \tilde{u} \rangle_{\mathbf{\Gamma}^{-1}_{\mathrm{noise}}} = 0, \forall \tilde{u} \in V$$

$$\mathcal{L}^E_{\psi}(\tilde{\psi}) = -2 \int_\Omega \tilde{\psi} \hat{p} \, \mathrm{d}x - 2\lambda^* \langle \psi, \tilde{\psi} \rangle_{\mathcal{C}_0^{-1}} = 0, \forall \tilde{\psi} \in V$$

Like in the finite dimensional example, we find $\hat{u}^* = \hat{u}, \hat{p}^* = \hat{p}$ and $\lambda^* = \lambda$. Hence:

$$\mathcal{L}^E_c = \int_\Omega \hat{u} \hat{p}^* \, \mathrm{d}x + \int_\Omega \hat{u}^* \hat{p} \, \mathrm{d}x = 2 \int_\Omega \hat{u} \hat{p} \, \mathrm{d}x$$

Forward finite difference error vs. exact



Conclusion

In summary,

- Eigenvalue Sensitivities is a scalable and exact method of leveraging existing structures to compute derivatives of eigenvalue based expressions.
- Avoids difficulties in finite-difference methods resolving and solves half the puzzle of using the information gain as a sensitivity objective.

Conclusion

In summary,

- Eigenvalue Sensitivities is a scalable and exact method of leveraging existing structures to compute derivatives of eigenvalue based expressions.
- Avoids difficulties in finite-difference methods resolving and solves half the puzzle of using the information gain as a sensitivity objective.

Future Work

Wrap up sensitivity analysis of the Costa-Rica problem via the KL-Divergence.

- Complete HDSA portion of KL Divergence sensitivities
- Interpret Sensitivities in terms of real-world model