Local point spread function Hessian approximation

Nick Alger[1]

Joint work with:
Noémi Petra,[2] Tucker Hartland,[2] Omar Ghattas[1]

[1]Oden Institute
The University of Texas at Austin
[2]Applied Mathematics, School of Natural Sciences
University of California, Merced

CCGO meeting

Outline

- Motivation: **Antartic ice sheet**
- New Hessian approximation: **big idea**
- New Hessian approximation: **technical details**
- Preliminary **numerical results** (heat inverse problem)

Outline

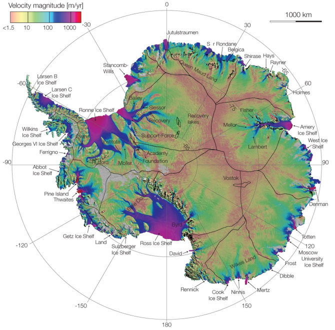
- Motivation: **Antartic ice sheet**

- New Hessian approximation: **big idea**

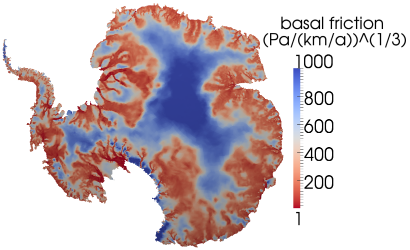
- New Hessian approximation: **technical details**

- Preliminary **numerical results** (heat inverse problem)

Antartic ice sheet



Observed surface flow velocity from InSAR (Rignot et. al, 2011)



Antarctic ice sheet inversion for the basal friction parameter field from InSAR surface velocities

Details in: T. Isaac, N. Petra, G. Stadler, and O. Ghattas. *Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet*, Journal of Computational Physics, 296, 348-368 (2015).

Ice sheet dynamics: forward and inverse

Balance of linear momentum, mass, and energy

$$-\boldsymbol{\nabla} \cdot [\eta(\theta, \boldsymbol{u})\, \dot{\boldsymbol{\varepsilon}} - \boldsymbol{I}p] = \rho \boldsymbol{g}, \qquad [\dot{\boldsymbol{\varepsilon}} = \tfrac{1}{2}(\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}\boldsymbol{u}^T)]$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0,$$

$$\rho c \left(\frac{\partial \theta}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla}\theta \right) - \boldsymbol{\nabla} \cdot (K \boldsymbol{\nabla}\theta) = 2\,\eta\, \mathrm{tr}(\dot{\boldsymbol{\varepsilon}}^2)$$

We have: Satellite observations of surface velocity

We want: The sliding/friction coefficient β in Robin boundary condition

$$\mathbf{T}(\boldsymbol{\sigma}\mathbf{n}) + \beta(x)\mathbf{T}\mathbf{u} = 0$$

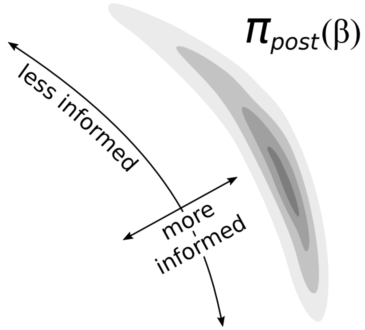
($\mathbf{T}$ is tangential component)

Bayesian approach

Inverse problem: given noisy data $\boldsymbol{d}$ and a model $\boldsymbol{f}$, infer parameters $\boldsymbol{\beta}$ that characterize the model, i.e.,

$$\boldsymbol{f}(\boldsymbol{\beta}) + \boldsymbol{e} = \boldsymbol{d}$$

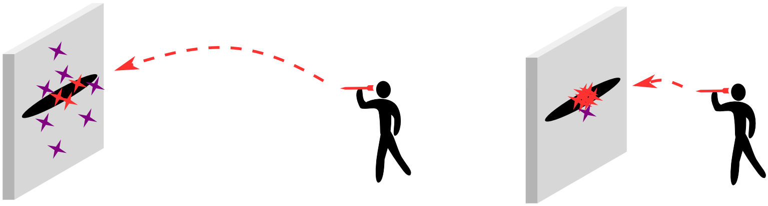
Interpret $\boldsymbol{\beta}$, $\boldsymbol{d}$ as random variables; solution of inverse problem is the "posterior" probability density function $\pi_{\text{post}}(\boldsymbol{\beta})$ found via Bayes' theorem.



Ill-conditioning and sampling

Objective: characterize posterior distribution by drawing samples with Markov chain Monte Carlo.

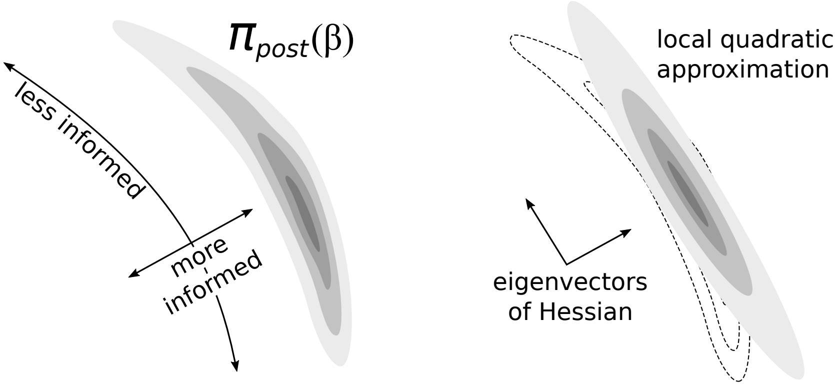
Dilemma: If the directional scalings of the proposal distribution are inconsistent with the directional scalings of the posterior, then sampling will be slow.



Hessian: local Gaussian approximation

Local Gaussian approximation proposal:

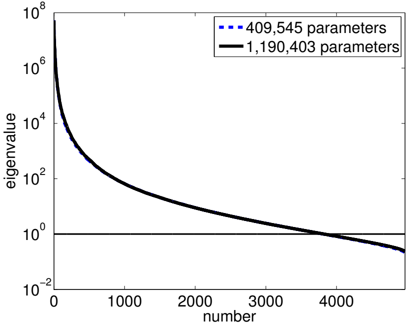
$$\pi_{\mathsf{prop}}(\beta) := \frac{\det \boldsymbol{H}^{1/2}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{\beta}_k + \boldsymbol{H}^{-1}\boldsymbol{g}\right)^T \boldsymbol{H}\left(\boldsymbol{y} - \boldsymbol{\beta}_k + \boldsymbol{H}^{-1}\boldsymbol{g}\right)\right)$$



Low-rank approximation/Woodbury formula:

$$\Gamma_{\text{prop}} = H^{-1} = \left(F^T \Gamma_{\text{noise}}^{-1} F + \Gamma_{\text{prior}}^{-1}\right)^{-1} \approx \Gamma_{\text{prior}}^{1/2} (V_r \Lambda_r V_r^T + I)^{-1} \Gamma_{\text{prior}}^{1/2}$$

where V_r and Λ_r are the eigenvectors/values of $F^T \Gamma_{\text{noise}}^{-1} F v_i = \lambda_i \Gamma_{\text{prior}}^{-1} v_i$



Details in: T. Isaac, N. Petra, G. Stadler, and O. Ghattas. *Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet*, Journal of Computational Physics, 296, 348-368 (2015).

MCMC sampling: stochastic Newton

Performance results / Convergence diagnostics

	MPSRF	IAT	ESS	MSJ	ARR	#Stokes	time (s)
SN	1.348	600	875	64	2	**8400**	420

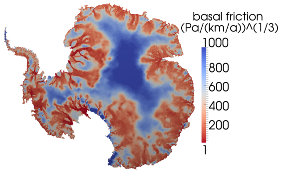
- **MPSRF:** multivariate potential scale reduction factor
- **IAT:** integrated autocorrelation time
- **ESS:** effecitive sample size
- **MSJ:** mean squared jump distance
- **ARR:** average rejection rate
- **#Stokes:** # of Stokes solves per independent sample
- **time:** time per independent sample

- **Statistics:** 21 parallel chains (each 25k); # samples: 525k; dof: 139; rank Hessian: 15

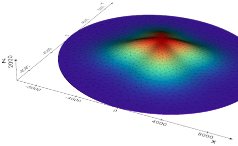
Too many PDE solves!!

Details in: N. Petra, J. Martin, G. Stadler, O. Ghattas. *A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet inverse problems*, SIAM Journal on Scientific Computing, 2014

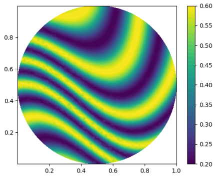
Model problems



(a) Antartica (Stokes)



(b) Ice mountain (Stokes)

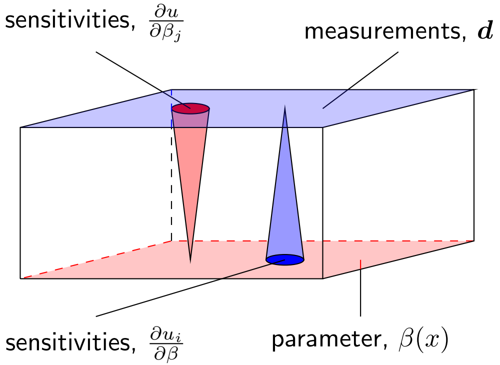


(c) Heat swirl (heat)

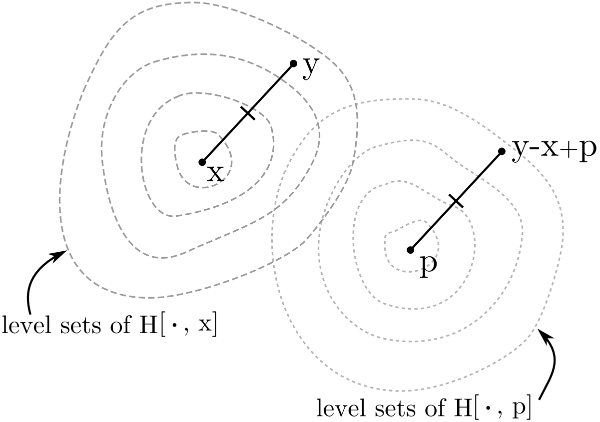
Outline

- Motivation: **Antartic ice sheet**
- New Hessian approximation: **big idea**
- New Hessian approximation: **technical details**
- Preliminary **numerical results** (heat inverse problem)

Local sensitivities



Local translation invariance



Hessian impulse responses

Hessian impulse response

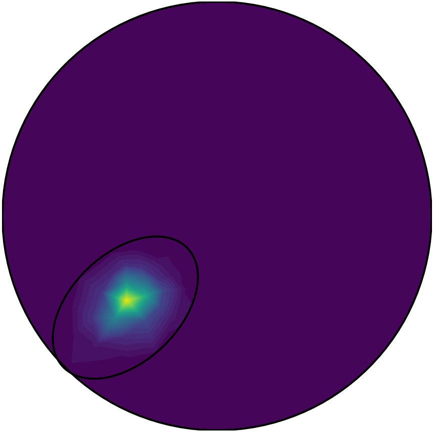


*image shown from ice mountain

Hessian impulse responses

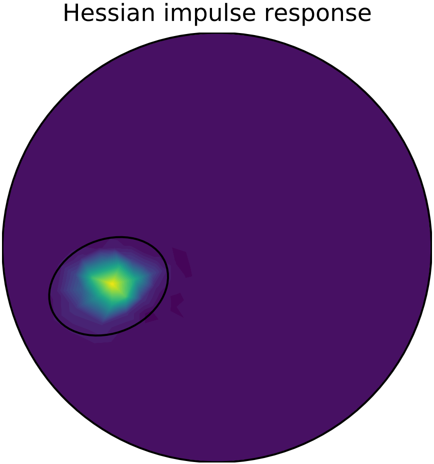


*image shown from ice mountain

Hessian impulse responses

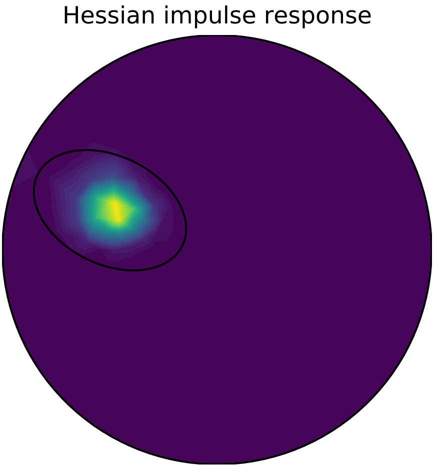


*image shown from ice mountain

Hessian impulse responses

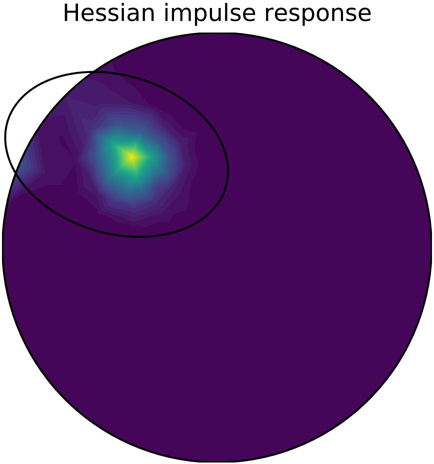


*image shown from ice mountain

Hessian impulse responses

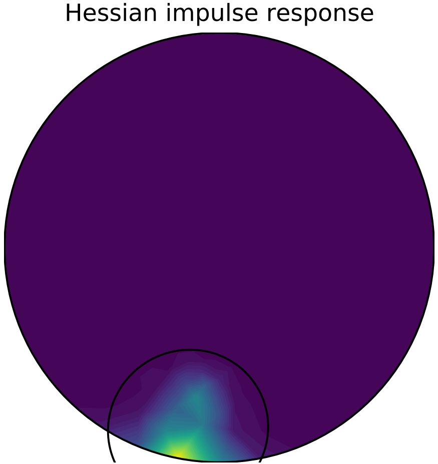


*image shown from ice mountain

Hessian impulse responses

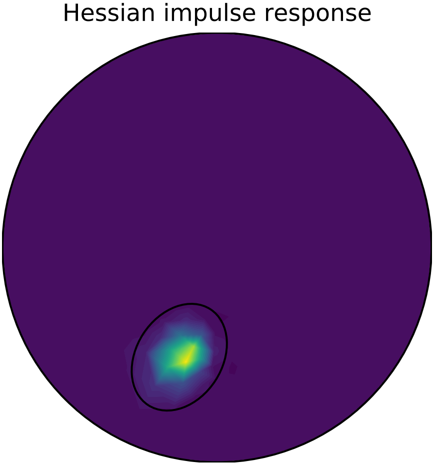


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

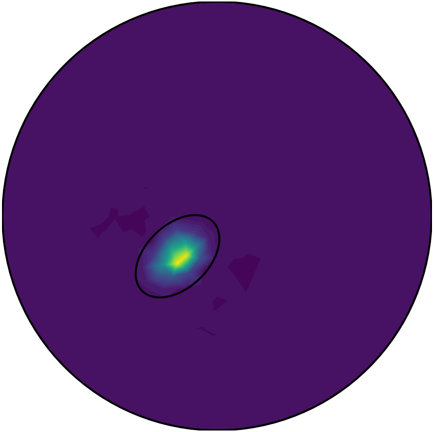


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

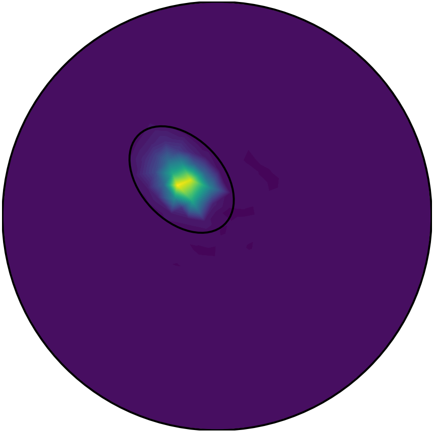


*image shown from ice mountain

Hessian impulse responses

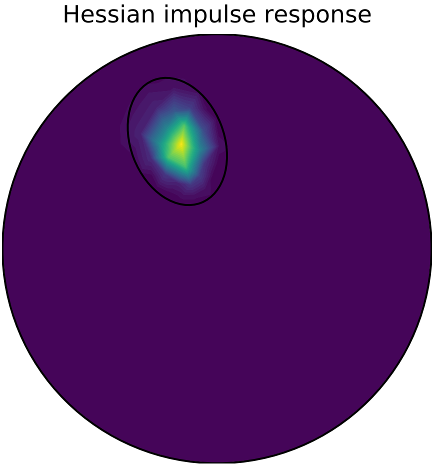


*image shown from ice mountain

Hessian impulse responses

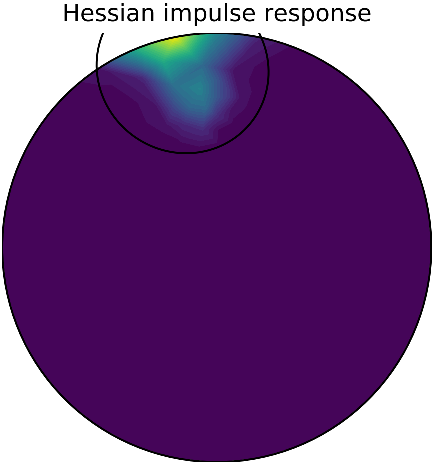


*image shown from ice mountain

Hessian impulse responses

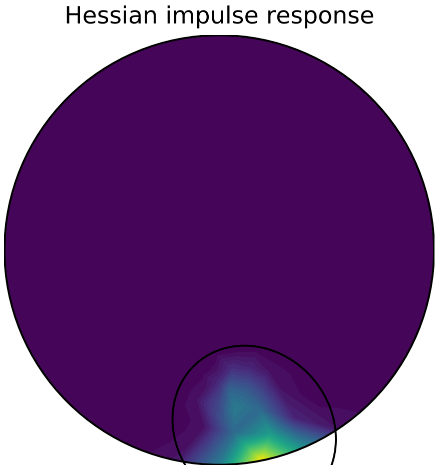


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

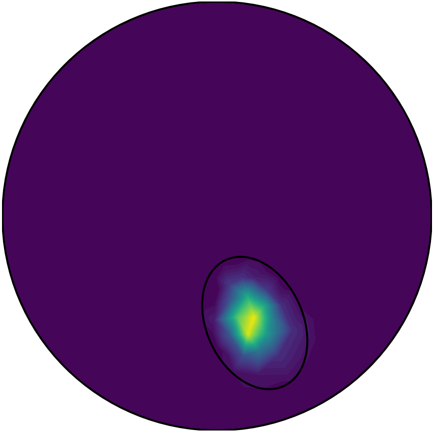


*image shown from ice mountain

Hessian impulse responses

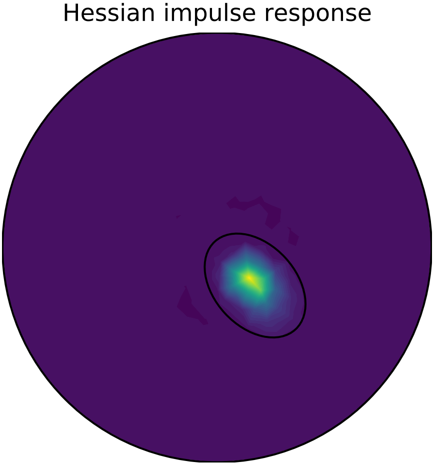


*image shown from ice mountain

Hessian impulse responses

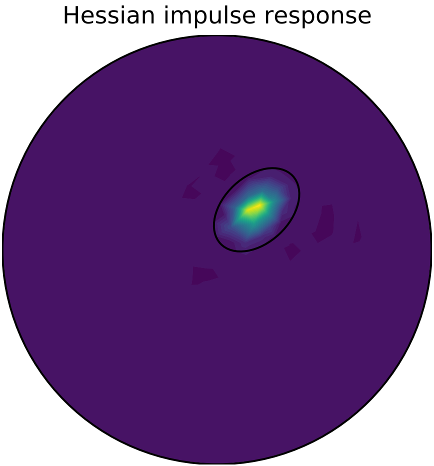


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

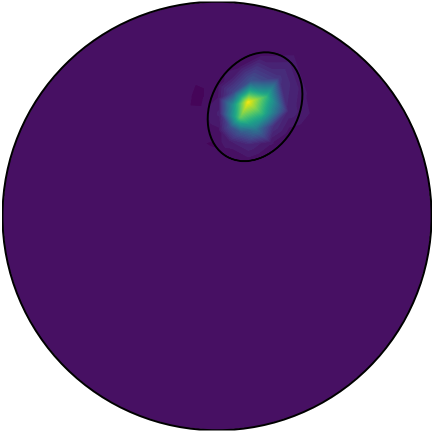


*image shown from ice mountain

Hessian impulse responses

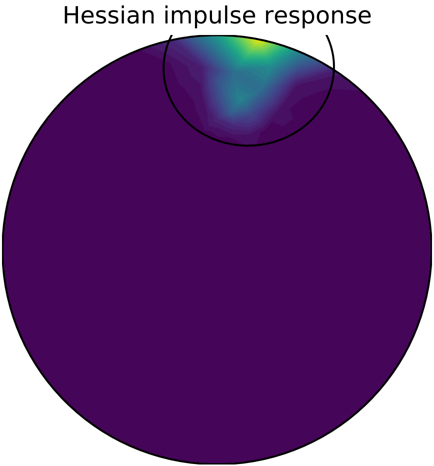


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

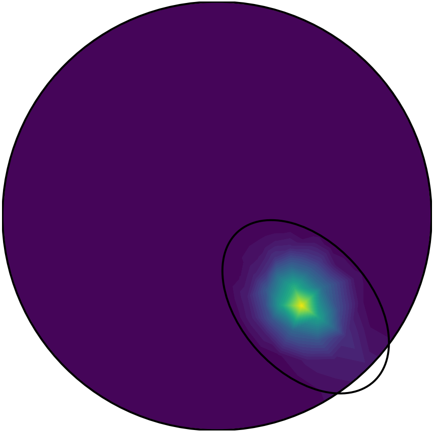


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

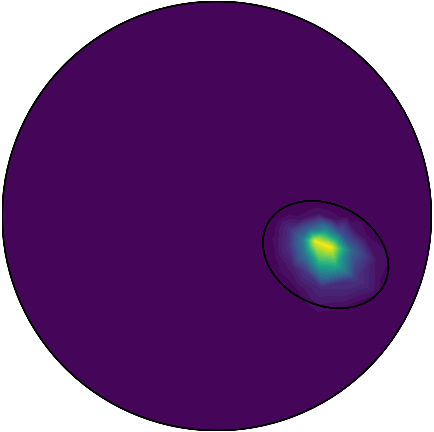


*image shown from ice mountain

Hessian impulse responses

Hessian impulse response

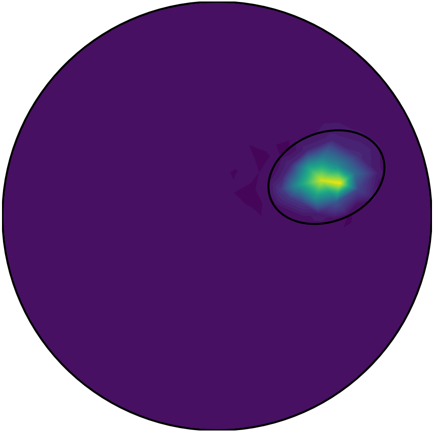


*image shown from ice mountain

Hessian impulse responses

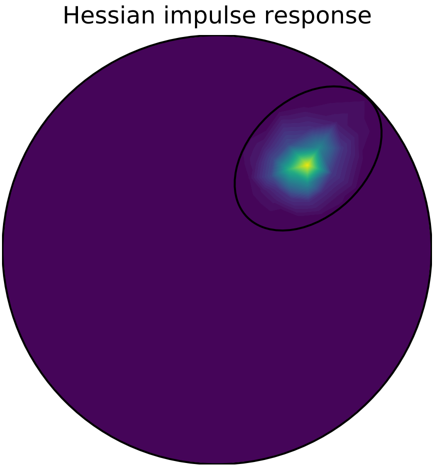
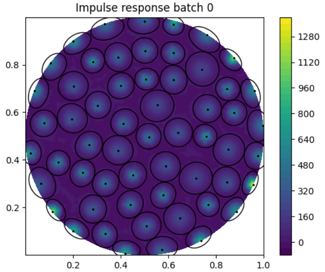
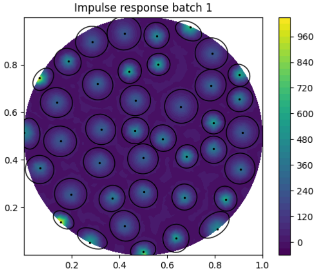


*image shown from ice mountain

Hessian approximation method: big idea

- **Step 1:** Compute "batches" of impulse responses by applying Hessian to Dirac combs
- **Step 2:** Interpolate known impulse responses to approximate unknown impulse responses
- **Step 3:** Convert to $\mathcal{H}$-matrix to do linear algebra





*images shown from heat swirl

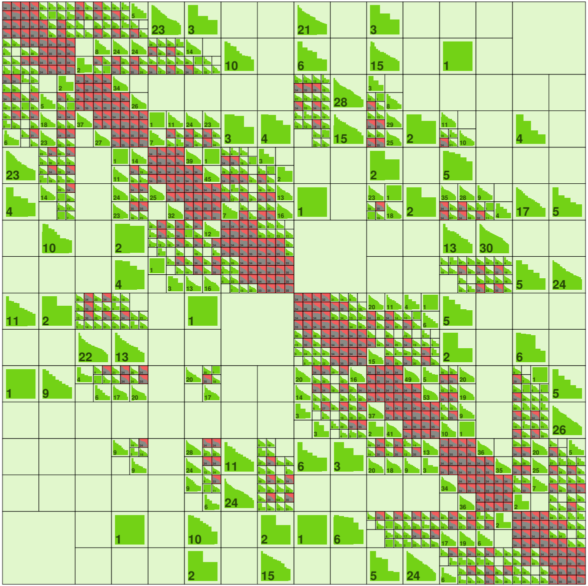
Outline

- Motivation: **Antartic ice sheet**
- New Hessian approximation: **big idea**
- New Hessian approximation: **technical details**
- Preliminary **numerical results** (heat inverse problem)

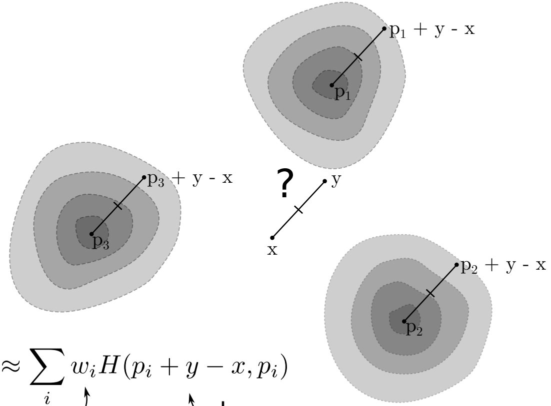
Technical details

- How do we choose the impulse response points?
- How do we interpolate the impulse responses?
  - What about boundary issues?
- What are $\mathcal{H}$-matrices and how do we use them?

Hierarchical matrices ($\mathcal{H}$-matrices)



Radial basis function interpolation



$$H(y,x) \approx \sum_i w_i H(p_i + y - x, p_i)$$

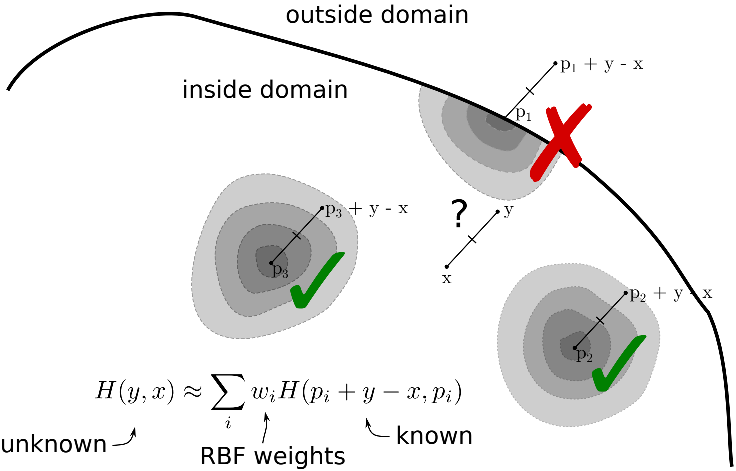
unknown

RBF weights

known

- Interpolate impulse responses using polyharmonic spline radial basis functions.
- Use only k-nearest neighbors (must solve $k \times k$ linear system)

Boundary considerations (1)



$$H(y,x) \approx \sum_i w_i H(p_i + y - x, p_i)$$

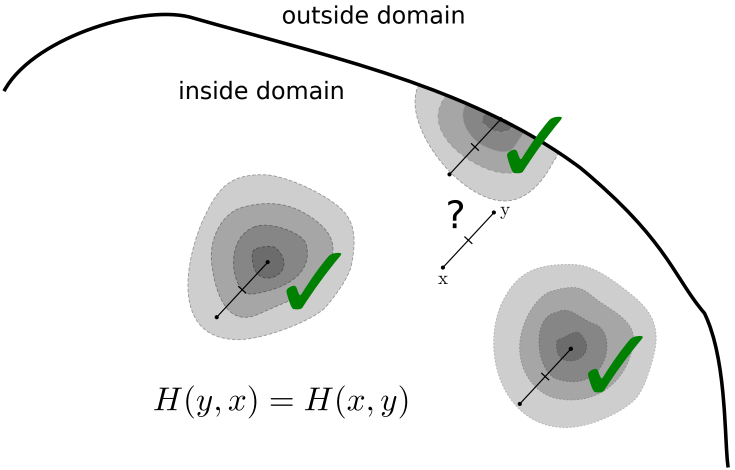
unknown

RBF weights

known

- If $p_i + y - x$ is outside the domain, don't use ith impulse response for $H(y,x)$

Boundary considerations (2)

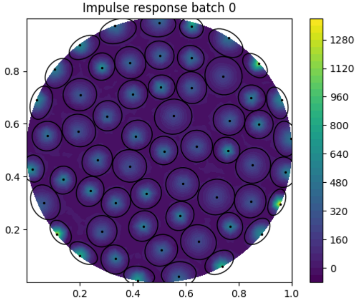


$$H(y, x) = H(x, y)$$

- Take advantage of symmetry

How to choose impulse response points?

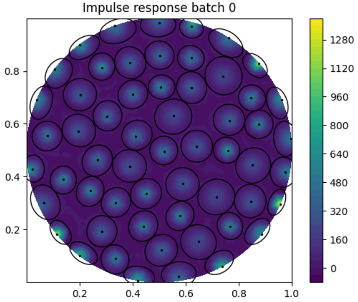
One hessian matrix-vector product $\to$ many impulse responses



- **Goal:** choose as many points as possible, such that the impulse response supports don't overlap
- **Dilemma:** How can we know the impulse response supports before we compute them?

How to choose impulse response points?

One hessian matrix-vector product $\rightarrow$ many impulse responses



- **Goal:** choose as many points as possible, such that the impulse response supports don't overlap
- **Dilemma:** How can we know the impulse response supports before we compute them?

Matrix analogy: getting all row sums

Matrix: let $A \in \mathbb{R}^{N \times N}$. Then

$$A^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \text{sum of } A \text{ col } 1 \\ \text{sum of } A \text{ col } 2 \\ \vdots \\ \text{sum of } A \text{ col } N \end{bmatrix}$$

Apply matrix to vector of ones $\rightarrow$ get row sums for all rows

Operator: let $C(x) = 1$ be the constant function. Then

$$(H^T C)^*(y) = \int_\Omega (H\delta_y)(x) dx$$

Apply Hessian to constant function $\rightarrow$ get volumes of every impulse response

Mean and standard deviations of impulse responses

- Let C, L^i, and Q^{ij} be the following functions:

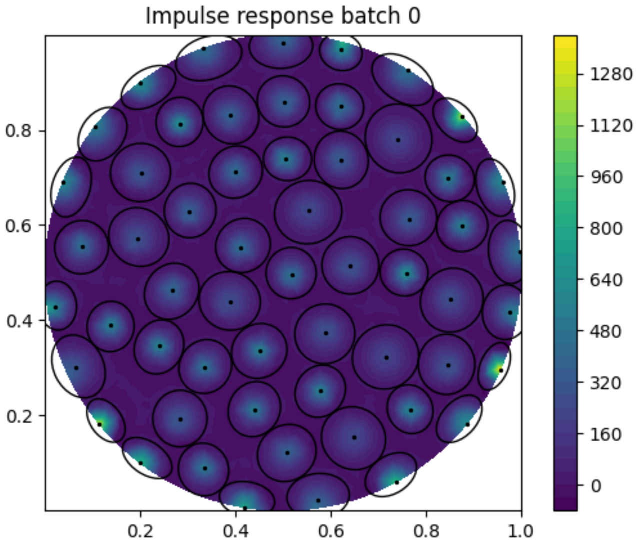
$$C(x) := 1, \qquad L^i(x) := x^i, \qquad Q^{ij}(x) = x^i x^j$$

- Then

$$\begin{aligned} V &= \left(H^T C\right)^* \\ \mu^i &= \left(H^T L^i\right)^* / V \\ \Sigma^{ij} &= \left(H^T Q^{ij}\right)^* / V - \mu^i \cdot \mu^j \end{aligned}$$

- Apply Hessian to constant, linear, and quadratic functions $\rightarrow$ get estimates of support for every impulse response

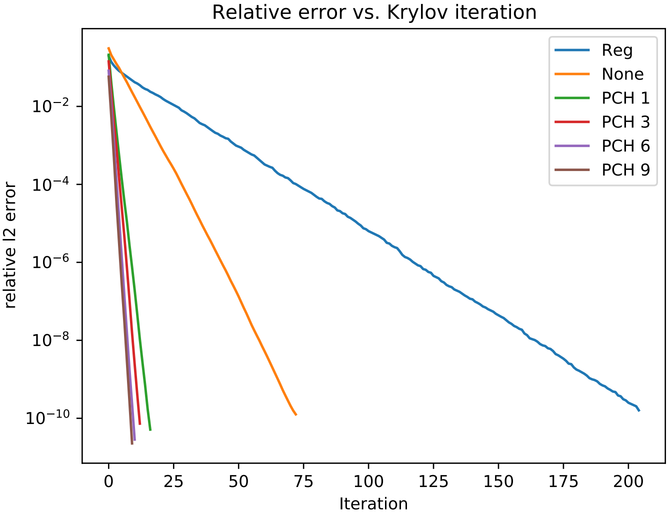
Impulse response support ellipsoids



Outline

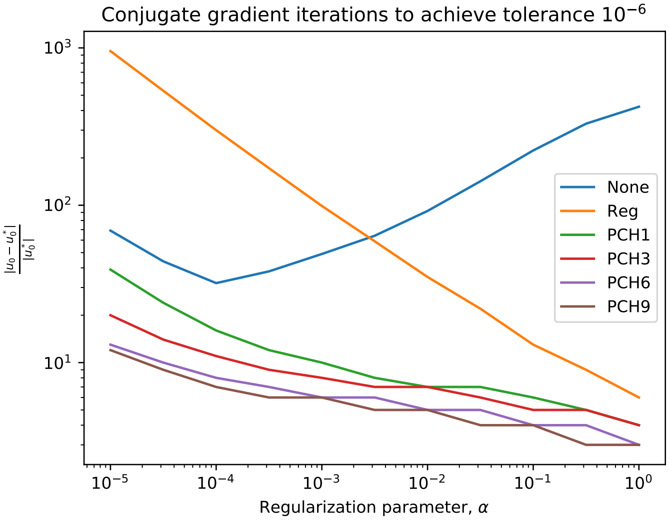
- Motivation: **Antartic ice sheet**
- New Hessian approximation: **big idea**
- New Hessian approximation: **technical details**
- Preliminary **numerical results** (heat inverse problem)

Heat swirl: CG Hessian solve with different preconditioners



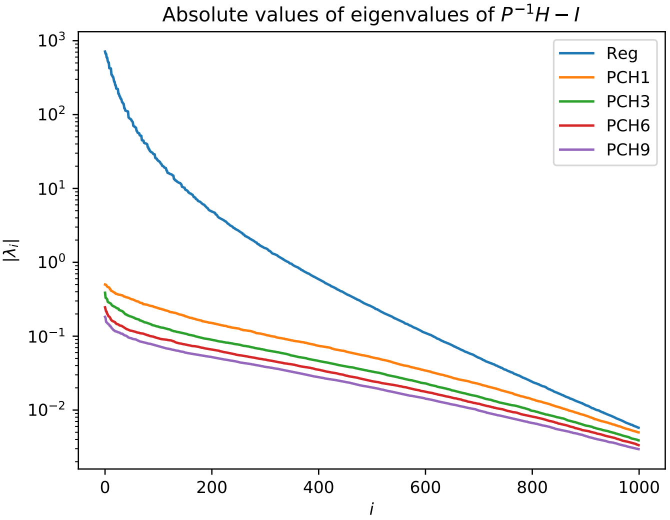
Solving $Hp = -g$ with preconditioned conjutate gradient

Heat swirl: Krylov iter vs. regularization parameter

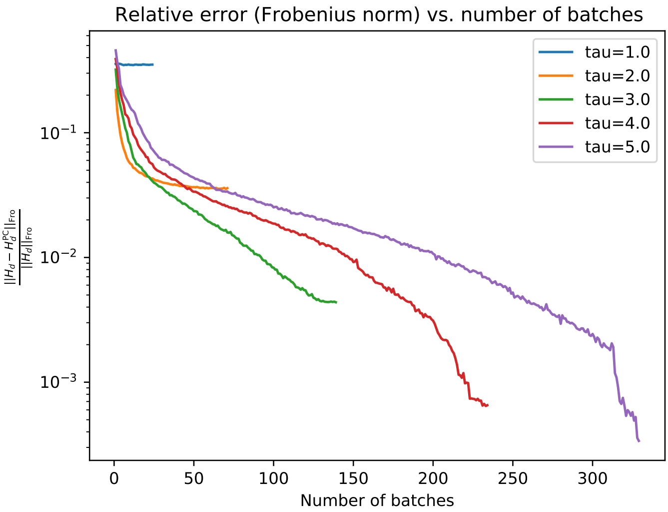


Solving $Hp = -g$ with preconditioned conjutate gradient

Heat swirl: preconditioned spectrum

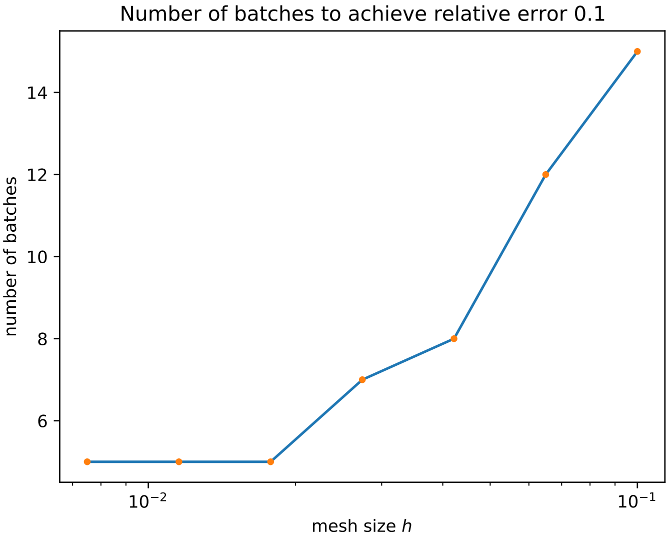


Heat swirl: error vs. num batches



tau: number of standard deviations used for ellipsoids

Heat swirl: error vs. mesh size



Summary

- Hessian approximations or preconditioners are essential for Bayesian sampling in inverse problems governed by partial differential equations.
- Low-rank approximations of the Hessian become prohibitive as the data becomes more informative (as is the case for ice sheet inverse problems).
- Local point spread function interpolation combined with Hierarchical matrix representations promise a more efficient Hessian approximation.