Bayesian inverse problems with application to the flow of the Greenland ice sheet

SAND2021-7138C

Tucker Hartland[1,3]

joint work with:

Georg Stadler[2], Mauro Perego[3], Kim Liégeois[3], Noémi Petra[1]

[1]Department of Applied Mathematics, University of California, Merced

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

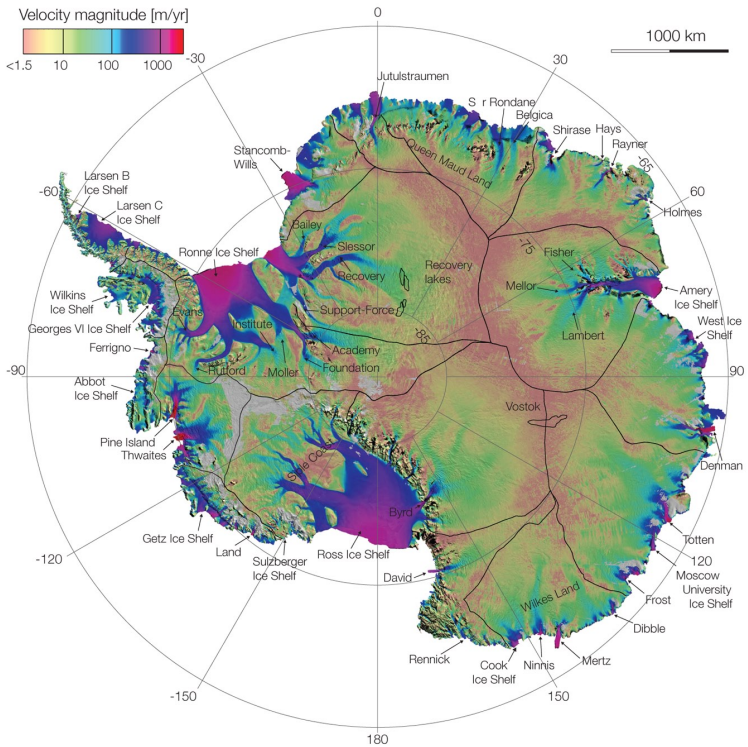
[3]Center for Computing Research, Sandia National Laboratories

MS101 Inverse Problems Governed by Viscous Flow
SIAM Conference on Mathematical and Computational Issues in the Geosciences
June 24, 2021

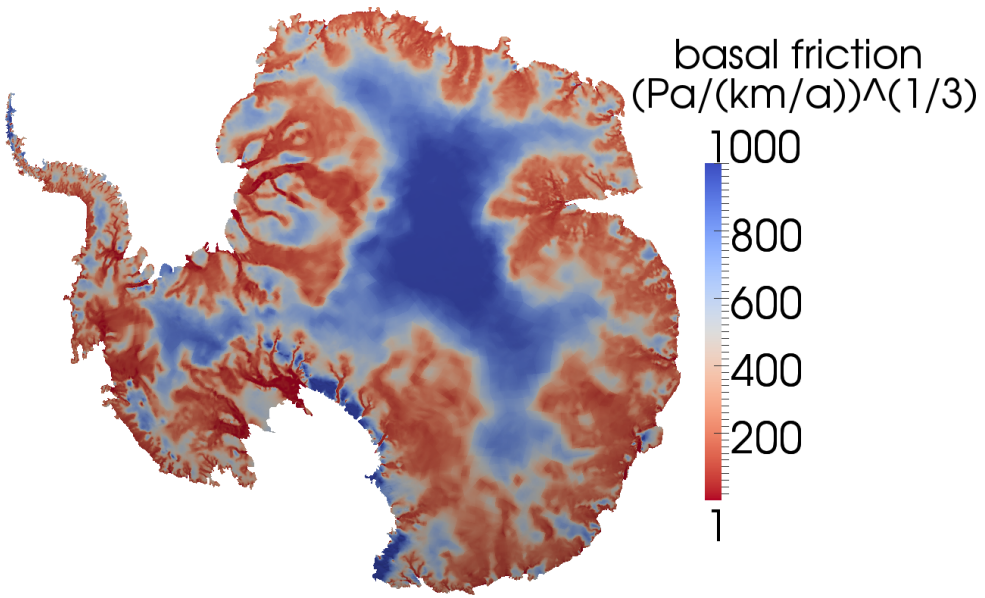
Research supported by NSF-Grant CAREER-1654311 and DOE SciDAC project ProSPect.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Motivation: Bayesian inversion for continental scale ice sheet models



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

Balance of mass and linear momentum

$$-\boldsymbol{\nabla}\cdot[\eta(\boldsymbol{u})\,\dot{\boldsymbol{\varepsilon}}-\boldsymbol{I}p]=\rho\,\boldsymbol{g},\quad[\dot{\boldsymbol{\varepsilon}}=\tfrac{1}{2}(\boldsymbol{\nabla u}+\boldsymbol{\nabla u}^{\top})]$$
$$\boldsymbol{\nabla}\cdot\boldsymbol{u}=0$$

Mathematical challenges:

- highly ill-conditioned linear systems
- complex, high aspect ratio geometry
- strong nonlinearities
- multiphysics couplings
- uncertain basal boundary conditions, topography, heat flux
- observational data: InSAR, laser altimetry, GRACE satellite, ice cores, radar

Inference of basal bdry. cond.

- critical for climate simulations
- inference of effective sliding/friction coefficient $\beta(\boldsymbol{x})$ in Robin boundary condition

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

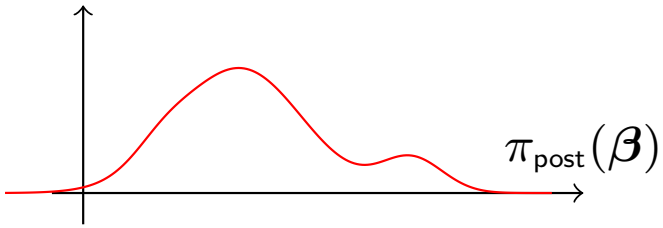
($\mathbf{T}$ is tangential component) from surface velocity observations

Bayesian approach to inverse problems

Inverse problem: given (possibly noisy) data $\boldsymbol{d}$ and (a possibly uncertain) 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 $\pi_{\text{post}}(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$:



Remarks:

- Bayesian framework quantifies uncertainty in the inverse solution, given uncertainty in the prior, the data, and the model.
- Prior incorporates known information and in infinite dimensions chosen to act as a regularization.
- Bayesian solution is probability density in as many dimensions as the number of parameters.

Exploring the posterior

If the prior is Gaussian with mean $\boldsymbol{m}_{\text{pr}}$ and covariance $\boldsymbol{\Gamma}_{\text{prior}}$, and the measurement noise is additive and described by a Gaussian distribution, i.e., $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{noise}})$, then we obtain for the posterior pdf:

$$\pi_{\text{post}}(\boldsymbol{\beta}) \propto \exp\left(-\tfrac{1}{2} \parallel \boldsymbol{f}(\boldsymbol{\beta}) - \boldsymbol{d} \parallel^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \tfrac{1}{2} \parallel \boldsymbol{\beta} - \boldsymbol{\beta}_{\text{pr}} \parallel^2_{\boldsymbol{\Gamma}^{-1}_{\text{prior}}}\right)$$

The "maximum a posteriori" (MAP) point is

$$\begin{aligned}\boldsymbol{\beta}_{\text{MAP}} &\overset{\text{def}}{=} \arg\ \max_{\boldsymbol{\beta}}\ \pi_{\text{post}}(\boldsymbol{\beta}) \\ &= \arg\ \min_{\boldsymbol{\beta}}\ \frac{1}{2}\left\|\boldsymbol{f}(\boldsymbol{\beta}) - \boldsymbol{d}\right\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} + \frac{1}{2}\left\|\boldsymbol{\beta} - \boldsymbol{\beta}_{\text{pr}}\right\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{prior}}}\end{aligned}$$

$\Rightarrow$ deterministic inverse problem with appropriate weighted norms!

Details in: J. Kaipio and E. Somersalo, Statistical and Computational Inverse Problems, (2005).

Scalability of an ice sheet inverse solver

Inexact Newton-CG

#sdof	#pdof	#N	#CG	avgCG	#Stokes
95,796	10,371	42	2718	65	7031
233,834	25,295	39	2342	60	6440
848,850	91,787	39	2577	66	6856
3,372,707	364,649	39	2211	57	6193
22,570,303	1,456,225	40	1923	48	5376

- **#sdof**: number of degrees of freedom for the state variables;
- **#pdof**: number of degrees of freedom for the inversion parameter field;
- **#N**: number of Newton iterations;
- **#CG**, **avgCG**: total and average (per Newton iteration) number of CG iterations;
- **#Stokes**: total number of linear(ized) Stokes solves (from forward, adjoint, and incremental forward and adjoint problems)

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).

Scalability of an ice sheet inverse solver

Inexact Newton-CG

#sdof	#pdof	#N	#CG	avgCG	#Stokes
95,796	10,371	42	2718	65	**7031**
233,834	25,295	39	2342	60	**6440**
848,850	91,787	39	2577	66	**6856**
3,372,707	364,649	39	2211	57	**6193**
22,570,303	1,456,225	40	1923	48	**5376**

- **#sdof**: number of degrees of freedom for the state variables;
- **#pdof**: number of degrees of freedom for the inversion parameter field;
- **#N**: number of Newton iterations;
- **#CG**, **avgCG**: total and average (per Newton iteration) number of CG iterations;
- **#Stokes**: total number of linear(ized) Stokes solves (from forward, adjoint, and incremental forward and adjoint problems)

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

Proposal density: $$\frac{\det \boldsymbol{H}^{1/2}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{\beta}_k+\boldsymbol{H}^{-1}\boldsymbol{g}\right)^{\top}\boldsymbol{H}\left(\boldsymbol{y}-\boldsymbol{\beta}_k+\boldsymbol{H}^{-1}\boldsymbol{g}\right)\right)$$

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.

The Hessian (of the negative log posterior) plays a critical role in inverse problems

- Its spectral properties characterize the degree of ill-posedness.
- The Hessian drives Newton-type optimization algorithms for solving the inverse problem.
- The inverse of the Hessian locally characterizes the uncertainty in the solution of the inverse problem (under the Gaussian assumption, it is precisely the posterior covariance matrix).
- Goal: rapidly perform linear algebraic operations, i.e., manipulation of the Hessian (and its square root and inverse) actions needed by sampling or CG solvers, hence seek approaches to approximate the Hessian(-applies).
- These approximations can then be used as pre-conditioners, and to build MCMC proposals based on local Gaussian approximations.

Exploiting problem structure

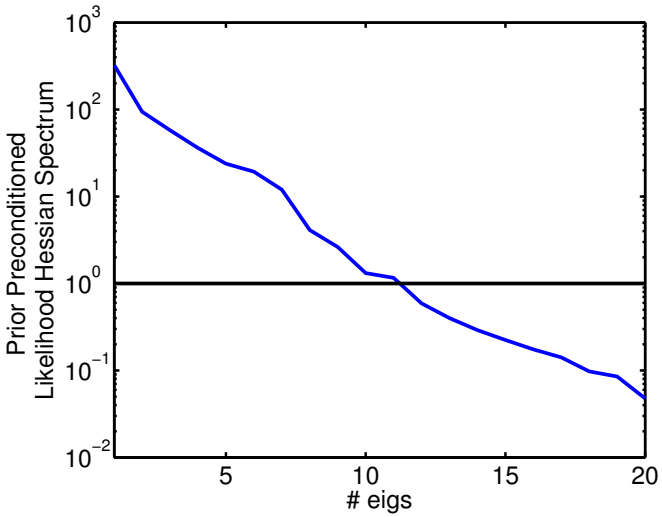
Low-rank properties of the data-misfit Hessian

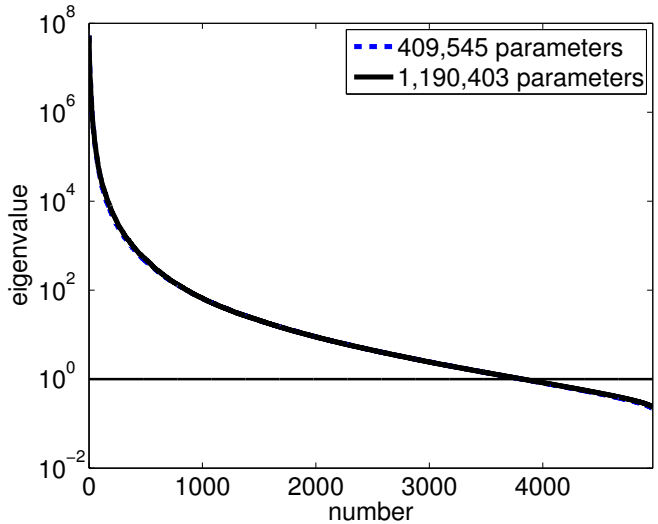
- Invoke low-rank (LR) approximation given by the inverse of the Hessian:

$$\boldsymbol{\Gamma}_{\text{post}} = \boldsymbol{H}^{-1} = \left(\boldsymbol{F}^\top \boldsymbol{\Gamma}_{\text{noise}}^{-1} \boldsymbol{F} + \boldsymbol{\Gamma}_{\text{prior}}^{-1}\right)^{-1} \approx \boldsymbol{\Gamma}_{\text{prior}}^{1/2} (\boldsymbol{V}_r \boldsymbol{\Lambda}_r \boldsymbol{V}_r^\top + \boldsymbol{I})^{-1} \boldsymbol{\Gamma}_{\text{prior}}^{1/2}$$

  where $\boldsymbol{V}_r$ and $\boldsymbol{\Lambda}_r$ are the eigenvectors/values of $\boldsymbol{F}^\top \boldsymbol{\Gamma}_{\text{noise}}^{-1} \boldsymbol{F} \boldsymbol{v}_i = \lambda_i \boldsymbol{\Gamma}_{\text{prior}}^{-1} \boldsymbol{v}_i$

- Spectrum of the prior-preconditioned likelihood Hessian for the Arolla glacier (139 parameters, left) and Antarctica (1.19M parameters, right)

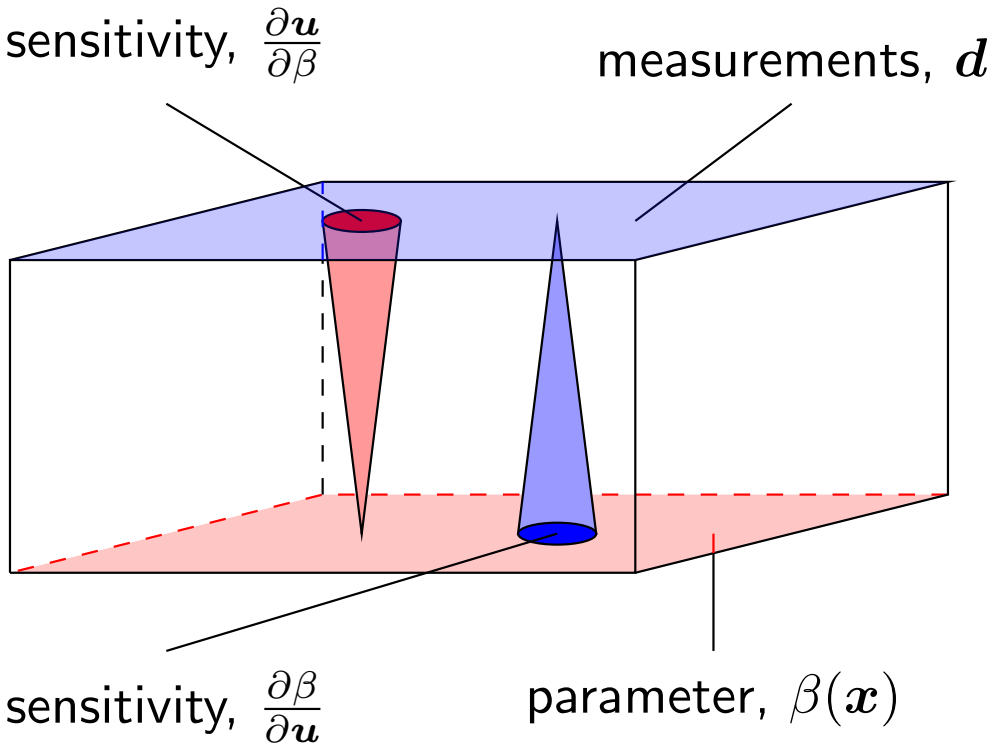




Details in: T. Isaac, N. Petra, G. Stadler, and O. Ghattas. *Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet*, Journal of Computational Physics, 296, 348-368 (2015).

Exploiting problem structure

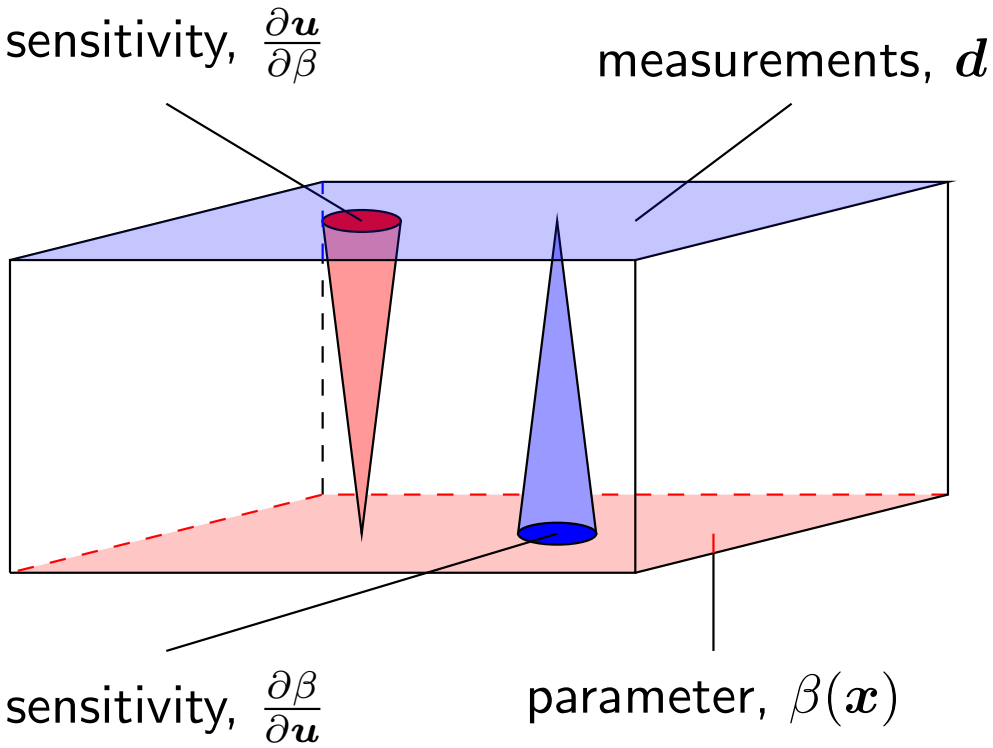
Localized sensitivities of the parameter-to-observable mapping



- ⋆ **Observation:** a variation of β at $\boldsymbol{x}_j$ will impact $\boldsymbol{u}$ in a neighborhood of $\boldsymbol{x}_j$.

Exploiting problem structure

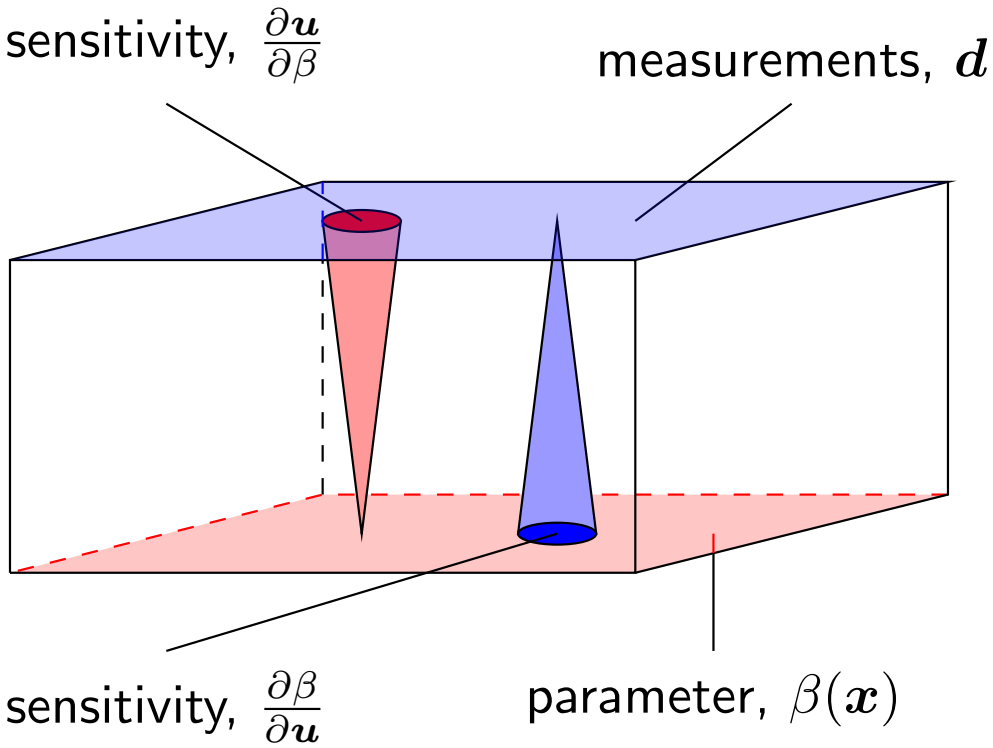
Localized sensitivities of the parameter-to-observable mapping



- ⋆ **Observation:** a variation of β at $\boldsymbol{x}_j$ will impact $\boldsymbol{u}$ in a neighborhood of $\boldsymbol{x}_j$.
- ⋆ **Hypothesis:** the Hessian has numerically low-rank off-diagonal blocks.

Exploiting problem structure

Localized sensitivities of the parameter-to-observable mapping



- ⋆ **Observation:** a variation of β at $\boldsymbol{x}_j$ will impact $\boldsymbol{u}$ in a neighborhood of $\boldsymbol{x}_j$.
- ⋆ **Hypothesis:** the Hessian has numerically low-rank off-diagonal blocks.
- ⋆ **Computational strategy:** generate Hessian approximants by compressing to HODLR format.

$\mathcal{H}$-matrices

Hierarchical off-diagonal low-rank (HODLR) matrices

$$\boldsymbol{H}_{\text{HODLR}} = \begin{bmatrix} \boldsymbol{A}_1^{(1)} & \boldsymbol{B}_1^{(1)} \\ \boldsymbol{B}_2^{(1)} & \boldsymbol{A}_2^{(1)} \end{bmatrix}$$

$$= \left[\begin{array}{cc:c} \boldsymbol{A}_1^{(2)} & \boldsymbol{B}_1^{(2)} & \boldsymbol{B}_1^{(1)} \\ \boldsymbol{B}_2^{(2)} & \boldsymbol{A}_2^{(2)} & \\ \hdashline \boldsymbol{B}_2^{(1)} & & \boldsymbol{A}_3^{(2)} \quad \boldsymbol{B}_3^{(2)} \\ & & \boldsymbol{B}_4^{(2)} \quad \boldsymbol{A}_4^{(2)} \end{array}\right]$$

$$= \left[\begin{array}{cc:c:cc} \boldsymbol{A}_1^{(3)} & \boldsymbol{B}_1^{(3)} & \boldsymbol{B}_1^{(2)} & & \\ \boldsymbol{B}_2^{(3)} & \boldsymbol{A}_2^{(3)} & & & \boldsymbol{B}_1^{(1)} \\ \hdashline \boldsymbol{B}_2^{(2)} & & \boldsymbol{A}_3^{(3)} \quad \boldsymbol{B}_3^{(3)} & & \\ & & \boldsymbol{B}_4^{(3)} \quad \boldsymbol{A}_4^{(3)} & & \\ \hdashline \boldsymbol{B}_2^{(1)} & & & \boldsymbol{A}_5^{(3)} \quad \boldsymbol{B}_5^{(3)} & \boldsymbol{B}_3^{(2)} \\ & & & \boldsymbol{B}_6^{(3)} \quad \boldsymbol{A}_6^{(3)} & \\ \hdashline & & \boldsymbol{B}_4^{(2)} & & \boldsymbol{A}_7^{(3)} \quad \boldsymbol{B}_7^{(3)} \\ & & & & \boldsymbol{B}_8^{(3)} \quad \boldsymbol{A}_8^{(3)} \end{array}\right]$$

- Low-rank

Details in: J. Ballani, D. Kressner. *Matrices with Hierarchical Low-Rank Structures*, Exploiting Hidden Structure in Matrix Computations: Algorithms and Applications, (2016).

Fast linear algebraic manipulation of HODLR matrices

- HODLR approximation generation of matrix-free operators by selective random sampling;

  P. Martinsson. *Compressing Rank-Structured Matrices via Randomized Sampling*, SIAM Journal on Scientific Computing, (2016).

- Absolute and relative tolerance stopping criteria;

  Y. Xi, J. Xia, R. Chan. *A Fast Randomized Eigensolver with Structured LDL Factorization Update*, SIAM Journal on Matrix Analysis and Applications, (2014).

- $\mathcal{O}(N \log^2 N)$ direct solves;

  S. Ambikasaran, E. Darve. *An $\mathcal{O}(N \log N)$ Fast Direct Solver for Partial Hierarchically Semi-Separable Matrices*, Springer Journal of Scientific Computing, (2013).

- $\mathcal{O}(N \log^2 N)$ symmetric factorization;

  S. Ambikasaran, M. O'Neil, KR Singh. *Fast Symmetric Factorization of Hierarchical Matrices with Applications*, arXiv, (2014).

- $\mathcal{O}(N \log N)$ HODLR mat-vec action;

  P. Martinsson. *Compressing Rank-Structured Matrices via Randomized Sampling*, SIAM Journal on Scientific Computing, (2016).

Application to the Greenland ice sheet

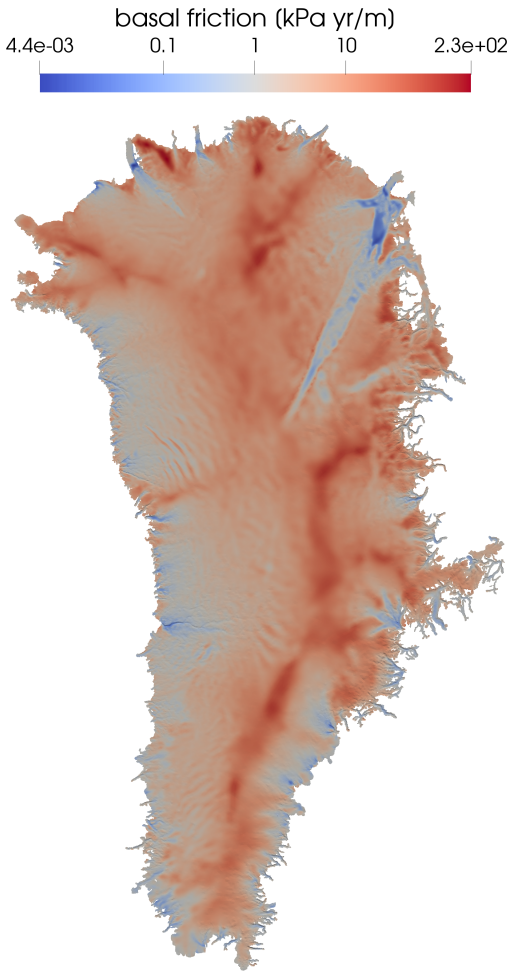
Utilizing the Albany multiphysics code*

- Exploit the localized sensitivity structure of the Greenland ice sheet flow;
- Preliminary results suggest off-diagonal low-rank structure**;

Details in:

*I.K. Tezaur, M. Perego, et al. *Albany/FELIX: a parallel, scalable and robust, finite element, first-order Stokes approximation ice sheet solver built for advanced analysis*, Geoscientific Model Development 2015

**T. Hartland, G. Stadler, N. Petra, M. Perego and K. Liégeois. *Hierarchical Off-Diagonal Approximation of Hessians in Inverse Problems*, In preparation.



Basal friction field*

Application to the Greenland ice sheet

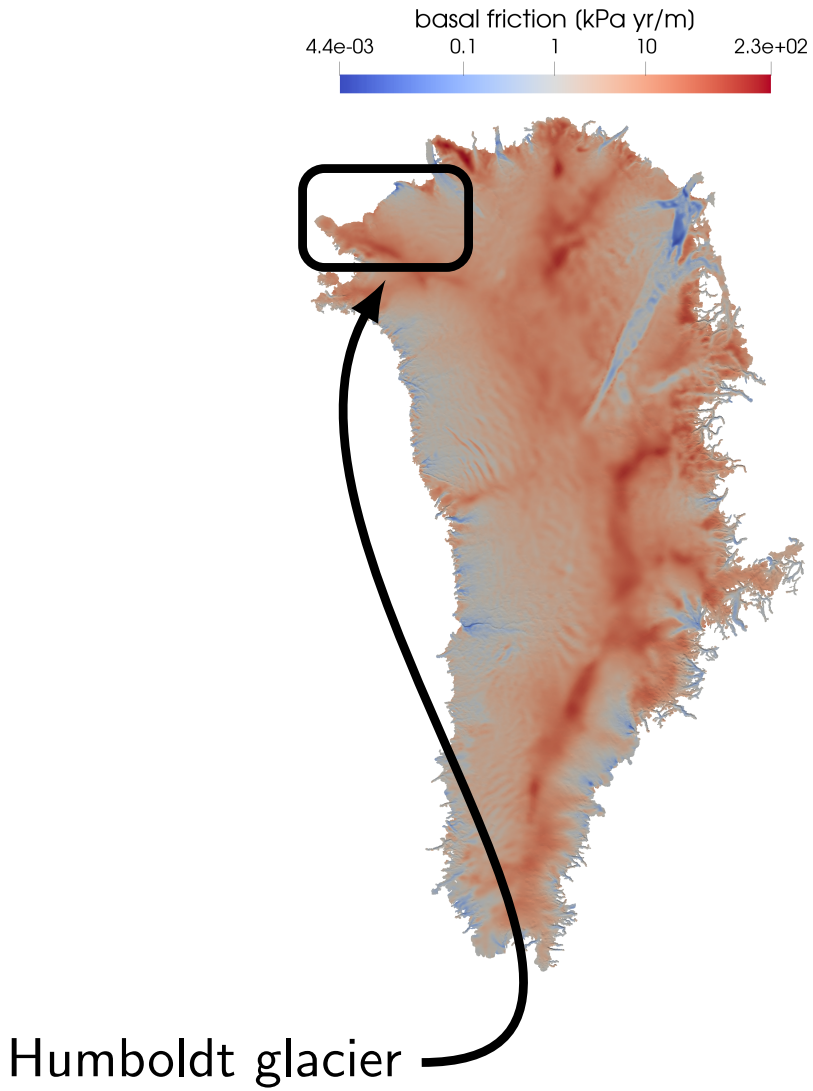
Utilizing the Albany multiphysics code*

- Exploit the localized sensitivity structure of the Greenland ice sheet flow;
- Preliminary results suggest off-diagonal low-rank structure**;

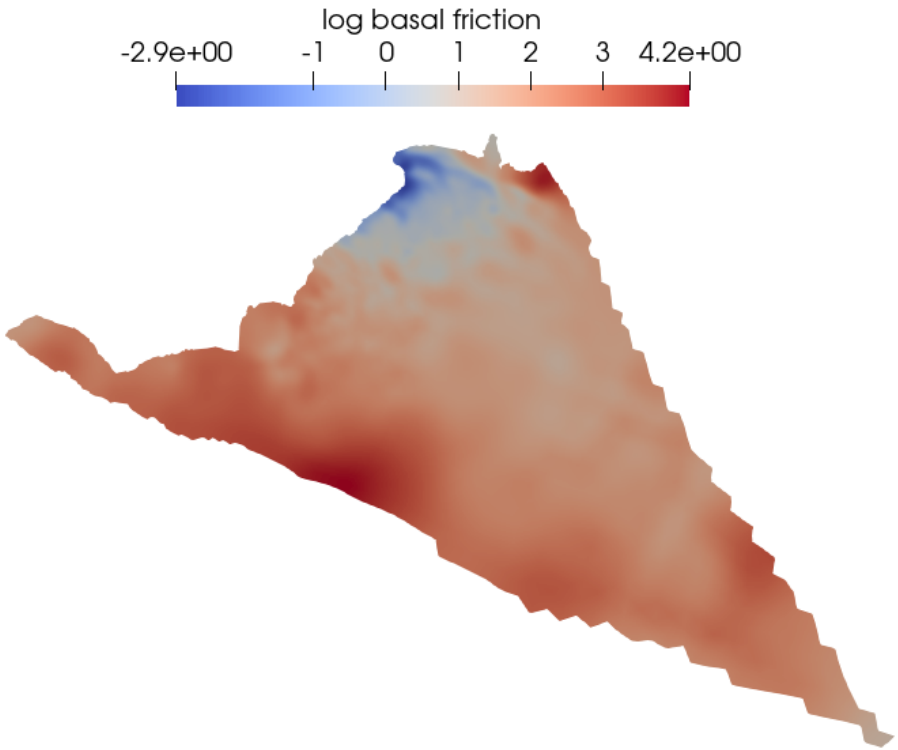
Details in:

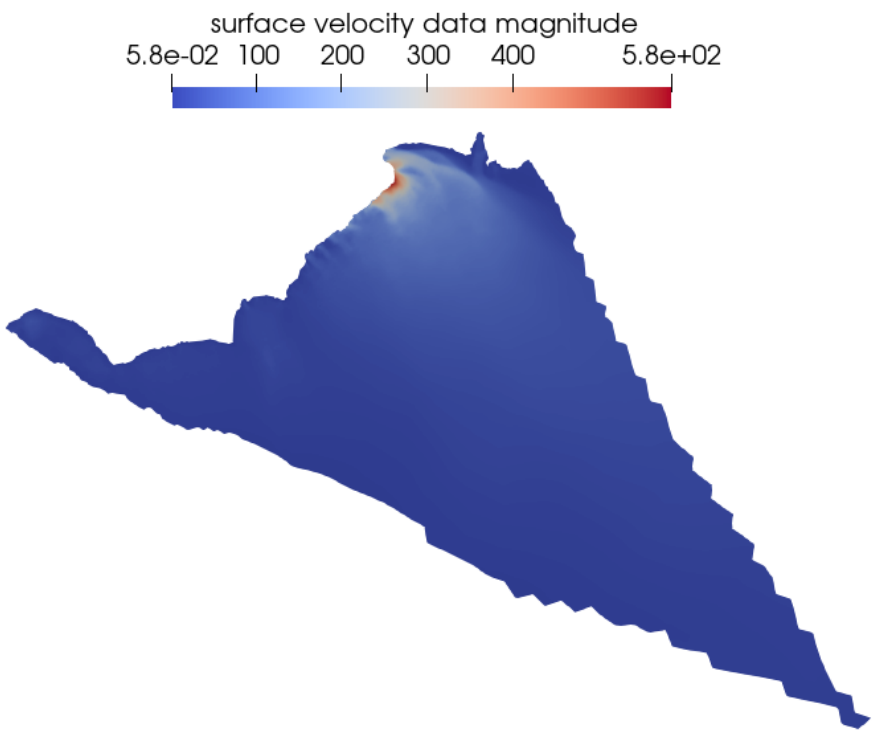
*I.K. Tezaur, M. Perego, et al. *Albany/FELIX: a parallel, scalable and robust, finite element, first-order Stokes approximation ice sheet solver built for advanced analysis*, Geoscientific Model Development 2015

**T. Hartland, G. Stadler, N. Petra, M. Perego and K. Liégeois. *Hierarchical Off-Diagonal Approximation of Hessians in Inverse Problems*, In preparation.



Humboldt glacier problem setup





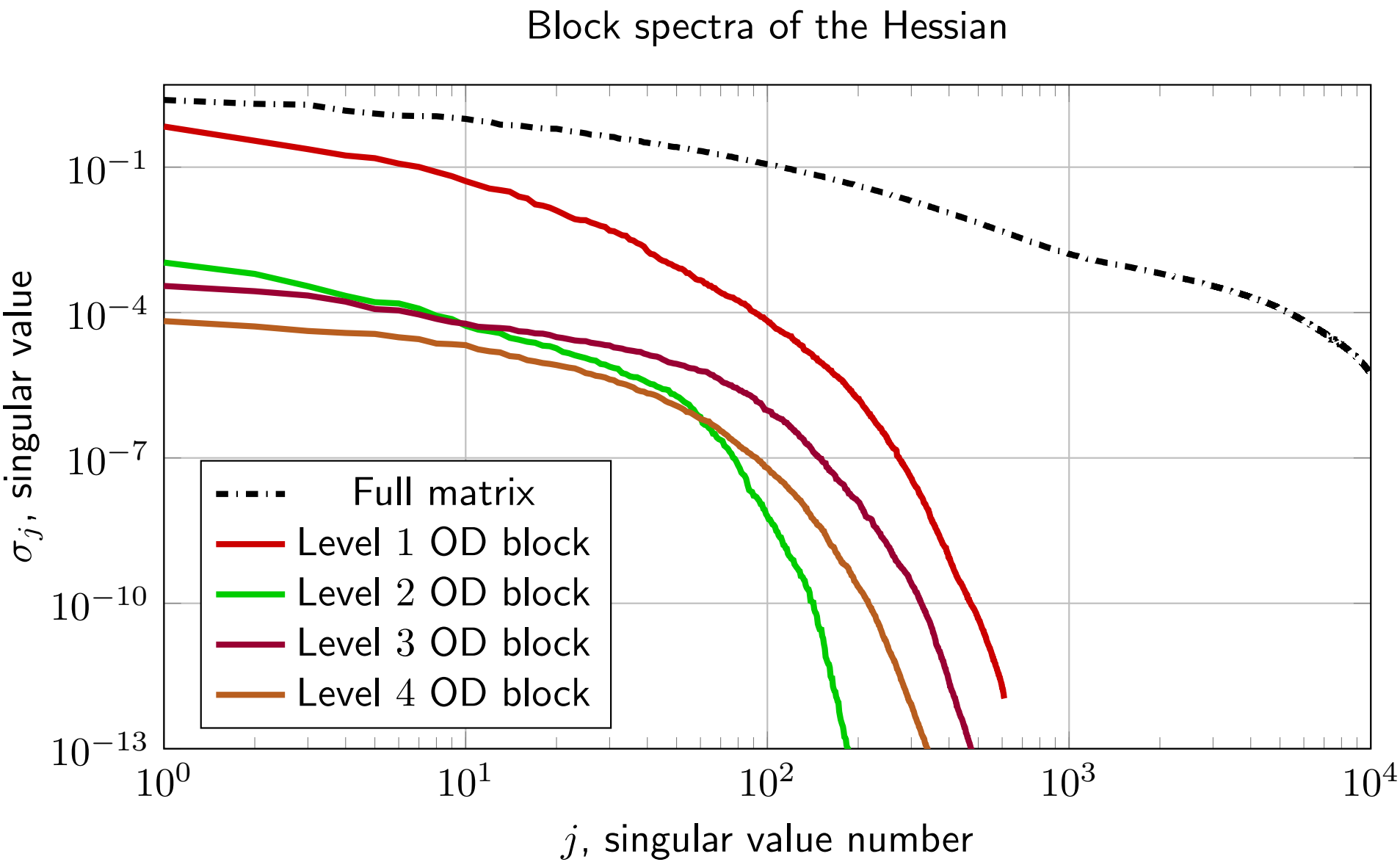
- three-dimensional first-order Stokes model
- # of discretized state variables: $\mathbf{255\,376}$
- # of discretized parameter variables: $\mathbf{11\,608}$
- # of processors: $\mathbf{64}$

`Albany`

- Hessian action obtained by automatic differentiation (`SACADO`)
- Nonlinear forward solves (`NOX`)
- Multigrid preconditioners (`MueLu`)
- ...

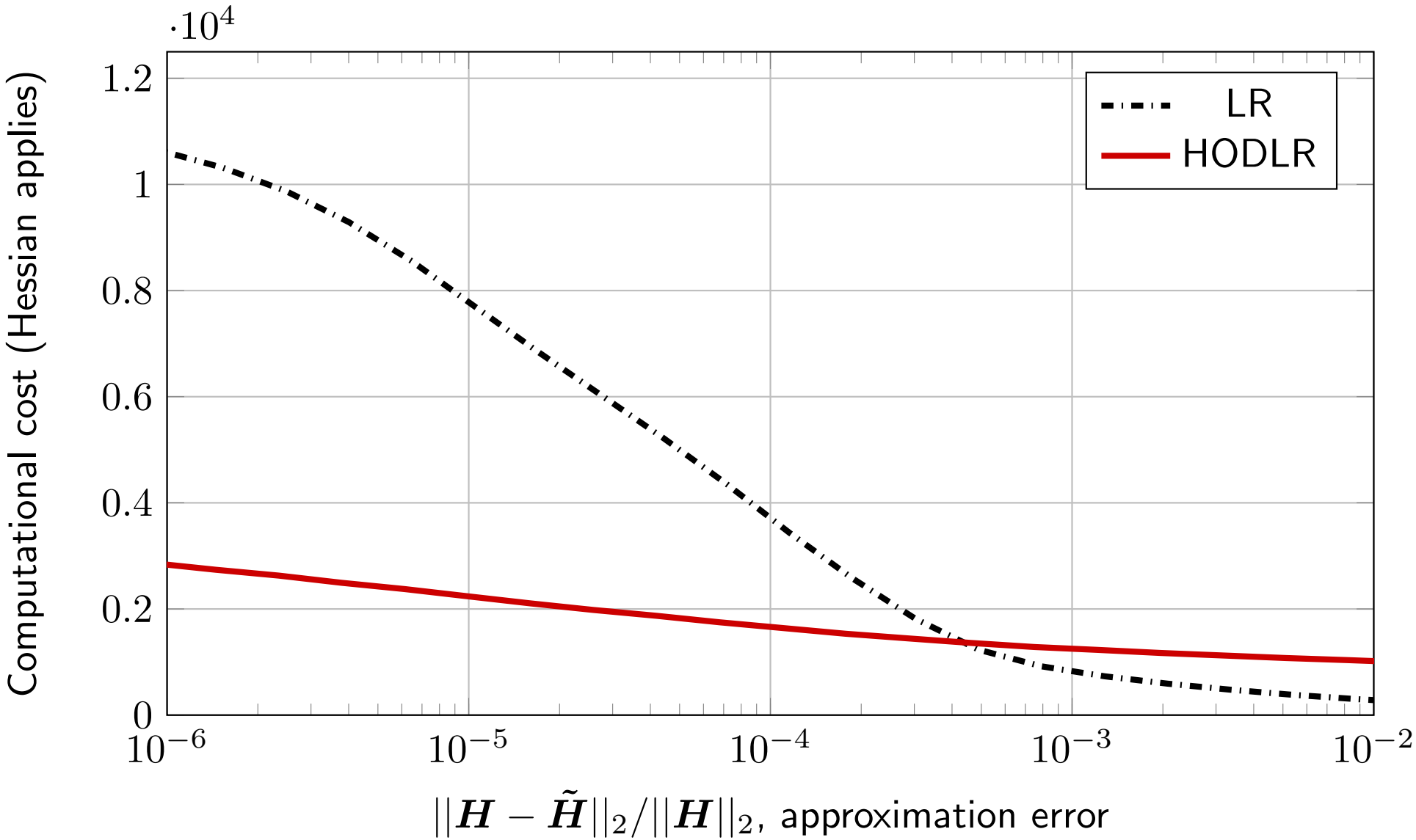
Rank-structure of the Hessian misfit

Singular values of off-diagonal (OD) blocks decay rapidly



Rank-structured approximation costs

Comparison of the HODLR and low-rank (LR) formats



Summary

- Hessians in inverse problems governed by partial differential equations are essential for efficient, dimension-independent methods for both
  - Newton solution of deterministic inverse problems (i.e., for computing the MAP point);
  - Markov chain Monte Carlo sampling to characterize the posterior.
- To overcome the computational challenges of exploring the posterior in a Bayesian inference context, it is imperative to further exploit problem structure, e.g., local sensitivity of the data with respect to parameters, *local translation invariance and approximate local support*.

  Details in: N. Alger, T. Hartland, N. Petra and O. Ghattas. *Fast Matrix-Free Approximation of Smoothly Varying Blur Operators, with Application to Hessians in PDE-Constrained Inverse Problems with Highly Informative Data*, in preparation.
- Hierarchical matrix representations provide an efficient means of generating Hessian approximants.