

Exploiting Low-Dimensional Structure in Bayesian Inverse Problems Governed by Ice Sheet Flow Models¹

Noémi Petra

Department of Applied Mathematics
University of California, Merced

Joint work with:

Nick Alger (UT Austin), Tucker Hartland (LLNL), Omar Ghattas (UT Austin)

October 6 – 10, 2025

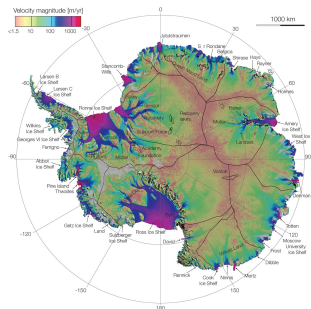
Workshop on “Data Assimilation and Inverse Problems for Digital Twins”
Institute for Mathematical and Statistical Innovation (IMSI)

¹Research supported partially by NSF-DMS-CAREER-1654311 and NSF-OAC-1550547.

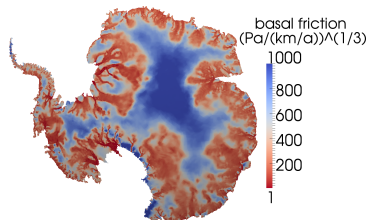
- Motivation
 - Why do we need predictive models with quantified uncertainties to accurately anticipate future sea level rise?
 - Why do we need more efficient approximations of high rank Hessians?
- Bayesian inverse problems governed by PDEs (preliminaries)
- Exploiting low-dimensional structure in Bayesian inverse problems governed by PDEs
 - (Global) low rank approximation
 - Hierarchical off-diagonal low rank (HODLR) approximation
 - Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation
- Numerical examples
 - Inversion for the basal friction coefficient in an ice sheet flow problem
 - Spatially varying blurring problem (if time permits)
- Conclusions and outlook

- 1 Motivation
- 2 Bayesian inverse problems governed by PDEs (preliminaries)
- 3 Exploiting low-dimensional structure in Bayesian inverse problems governed by PDEs
 - (Global) low rank approximation
 - Hierarchical off-diagonal low rank (HODLR) approximation
 - Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation
- 4 Numerical examples
 - Inversion for the basal friction coefficient in an ice sheet flow problem
 - Spatially varying blurring problem (if time permits)
- 5 Conclusions and outlook

Need predictive models with quantified uncertainties to accurately anticipate future sea level rise.



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



Antarctic ice sheet inversion for the basal friction parameter field using InSAR surface velocity measurements

- Several port cities will be at risk from coastal flooding in the future.
- Ice flowing from ice sheets to ocean is primary contributor to sea level rise.

Details in: Intergovernmental Panel on Climate Change (IPCC), "Climate Change 2021 - The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change". Cambridge University Press, 2023.

- 1 Motivation
- 2 Bayesian inverse problems governed by PDEs (preliminaries)
- 3 Exploiting low-dimensional structure in Bayesian inverse problems governed by PDEs
 - (Global) low rank approximation
 - Hierarchical off-diagonal low rank (HODLR) approximation
 - Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation
- 4 Numerical examples
 - Inversion for the basal friction coefficient in an ice sheet flow problem
 - Spatially varying blurring problem (if time permits)
- 5 Conclusions and outlook

The inverse problem

Use available observations/data $\mathbf{d}$ to infer the values of the unknown parameter field $\mathbf{m}$ that characterize a physical process modeled by PDEs, i.e.,

$$\mathbf{d} = \mathcal{F}(\mathbf{m}) + \boldsymbol{\eta}.$$

- The map $\mathcal{F} : \mathcal{M} \rightarrow \mathbb{R}^q$ is the so-called *parameter-to-observable* map.
- Evaluations of $\mathcal{F}$ involve the solution of the PDE (e.g., nonlinear Stokes in the context of ice sheet problems) given $\mathbf{m}$, followed by the application of an observation operator $\mathcal{B} : \mathcal{V} \rightarrow \mathbb{R}^q$ to extract the observations from the state.
- $\boldsymbol{\eta}$ accounts for noisy measurements and model errors and is modeled as $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}})$, i.e., a centered Gaussian at $\mathbf{0}$ with covariance $\boldsymbol{\Gamma}_{\text{noise}}$.

Bayesian formulation of the inverse problem

Describes probability of all models that are consistent with the observations/data and any prior knowledge about the parameters:

$$d\mu_{\text{post}} \propto \exp \left\{ -\frac{1}{2} \|\mathcal{F}(\boldsymbol{m}) - \boldsymbol{d}\|_{\boldsymbol{\Gamma}_{\text{noise}}^{-1}}^2 - \frac{1}{2} \|\boldsymbol{m} - \boldsymbol{m}_{\text{pr}}\|_{\boldsymbol{C}_{\text{prior}}^{-1}}^2 \right\}.$$

- The first term in the exponential is the negative log-likelihood (representing the probability that a given set of parameters might give rise to the observed data).
- The second term represents the negative log-prior (e.g., Gaussian prior, i.e., $\boldsymbol{m} \sim \mathcal{N}(\boldsymbol{m}_{\text{pr}}, \boldsymbol{C}_{\text{prior}})$).

Goals:

- Characterize the posterior statistically (MAP point, mean, covariance, etc.)
 - for functions $\boldsymbol{m}$ (large vectors after discretization), and
 - for expensive $\mathcal{F}(\cdot)$.
- Exploit connection to PDE-constrained optimization.

Inverse problems governed by PDEs

- The *maximum a posteriori* (MAP) point m_{MAP} is defined as the parameter field that maximizes the posterior distribution:

$$\begin{aligned} m_{\text{MAP}} &:= \operatorname{argmin}_{\mathbf{m} \in \mathcal{M}} (-\log d\mu_{\text{post}}(\mathbf{m})) \\ &= \operatorname{argmin}_{\mathbf{m} \in \mathcal{M}} \frac{1}{2} \|\mathcal{F}(\mathbf{m}) - \mathbf{d}\|_{\Gamma_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|\mathbf{m} - m_{\text{pr}}\|_{\mathcal{C}_{\text{prior}}^{-1}}^2, \\ &= \operatorname{argmin}_{\mathbf{m} \in \mathcal{M}} \frac{1}{2} \|\mathcal{B}(\mathbf{u}) - \mathbf{d}\|_{\Gamma_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|\mathbf{m} - m_{\text{pr}}\|_{\mathcal{C}_{\text{prior}}^{-1}}^2, \end{aligned}$$

where for given $\mathbf{m}$, $\mathbf{u}$ solves the PDE (e.g., Stokes).

Inverse problems governed by PDEs

- The *maximum a posteriori* (MAP) point m_{MAP} is defined as the parameter field that maximizes the posterior distribution:

$$\begin{aligned} m_{\text{MAP}} &:= \operatorname{argmin}_{\mathbf{m} \in \mathcal{M}} (-\log d\mu_{\text{post}}(\mathbf{m})) \\ &= \operatorname{argmin}_{\mathbf{m} \in \mathcal{M}} \frac{1}{2} \|\mathcal{F}(\mathbf{m}) - \mathbf{d}\|_{\mathbf{\Gamma}_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|\mathbf{m} - m_{\text{pr}}\|_{\mathcal{C}_{\text{prior}}^{-1}}^2, \\ &= \operatorname{argmin}_{\mathbf{m} \in \mathcal{M}} \frac{1}{2} \|\mathcal{B}(\mathbf{u}) - \mathbf{d}\|_{\mathbf{\Gamma}_{\text{noise}}^{-1}}^2 + \frac{1}{2} \|\mathbf{m} - m_{\text{pr}}\|_{\mathcal{C}_{\text{prior}}^{-1}}^2, \end{aligned}$$

where for given $\mathbf{m}$, $\mathbf{u}$ solves the PDE (e.g., Stokes).

- When $\mathcal{F}$ is linear, due to the particular choice of prior and noise model, the posterior measure is Gaussian, $\mathcal{N}(m_{\text{MAP}}, \mathcal{C}_{\text{post}})$

$$m_{\text{MAP}} = \mathcal{C}_{\text{post}}(\mathcal{F}^* \mathbf{\Gamma}_{\text{noise}}^{-1} \mathbf{d} + \mathcal{C}_{\text{prior}}^{-1} m_{\text{pr}}), \quad \mathcal{C}_{\text{post}} = \mathcal{H}^{-1} = (\mathcal{F}^* \mathbf{\Gamma}_{\text{noise}}^{-1} \mathcal{F} + \mathcal{C}_{\text{prior}}^{-1})^{-1},$$

where $\mathcal{F}^* : \mathbb{R}^q \rightarrow \mathcal{M}$ is the adjoint of $\mathcal{F}$, and $\mathcal{H}$ is the Hessian (second derivative) of the negative-log posterior.

- **Note:** In the general case of nonlinear parameter-to-observable map $\mathcal{F}$ the posterior distribution is not Gaussian.

Challenges in solving inverse problems (under uncertainty)

Severe mathematical and computational challenges place significant barriers on solving complex inverse problems (governed by PDEs) under uncertainty.

- **Expensive-to-evaluate/solve forward problems:**

- Challenges: **large state dimension**, complex and high-aspect ratio (thin) geometry, **nonlinear** and anisotropic constitutive laws, ill-conditioned linear and nonlinear algebraic systems that arise upon discretization as a result of heterogeneous and widely varying parameters, etc.
- All scalable methods will benefit from some combination of tools, e.g., ROM, multi-fidelity, data-driven models/corrections.

- **Large-scale optimization problem governed by the forward problem:**

- Challenges: **large parameter dimension**, bound constraints, etc.
- **Newton-type algorithms** are essential with efficient **preconditioning**.

- **Uncertainty quantification:**

- The solution of the Bayesian inverse problem takes the form of a very **high-dimensional posterior** probability density function.
- Models often times have **additional uncertainties** and/or are approximations; need to take into account **model uncertainty/error/discrepancy** (how do we represent/model this, and handle it in a scalable way?).

What do we mean by scalable?

- **Computational cost** of solving inverse problems governed by PDEs can be measured in number of (linearized) forward (and adjoint) solves.
- **Algorithmic scalability:** the computational cost is independent of the **state variable dimension**, the **parameter dimension**, and the **data dimension**.
- **Hardware scalability:** the compute time is decreased upon more hardware resources.
- The algorithm should be robust with respect to over discretization of state and parameter and overprovisioning of data.
- The algorithm should exploit the problem structure (e.g., global/hierarchical low-rankness, locality, smoothness, low-dimensionality) in Bayesian inverse problems governed by PDEs.

Scalability of an ice sheet inverse solver

Inexact Newton-CG (results obtained using the ymir Stokes solver)

#s dof	#p dof	#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

- **#s dof**: number of degrees of freedom for the state variables;
- **#p dof**: 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 (results obtained using the ymir Stokes solver)

#s dof	#p dof	#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

- **#s dof**: number of degrees of freedom for the state variables;
- **#p dof**: 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 (SN)

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**: effective sample size
- **MSJ**: mean squared jump distance
- **Statistics**: 21 parallel chains (each 25k); # samples: 525k; dof: 139; rank Hessian: 15
- **ARR**: average rejection rate
- **#Stokes**: # of Stokes solves per independent sample
- **time**: time per independent sample

$$\text{Proposal density: } \frac{\det \mathbf{H}^{1/2}}{(2\pi)^{n/2}} \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{m}_k + \mathbf{H}^{-1} \mathbf{g})^T \mathbf{H} (\mathbf{y} - \mathbf{m}_k + \mathbf{H}^{-1} \mathbf{g}) \right)$$

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
- K. Kim, U. Villa, M. Parno, Y. Marzouk, O. Ghattas, N. Petra, hippylib-MUQ, TOMS, 2023

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.

- 1 Motivation
- 2 Bayesian inverse problems governed by PDEs (preliminaries)
- 3 Exploiting low-dimensional structure in Bayesian inverse problems governed by PDEs
 - (Global) low rank approximation
 - Hierarchical off-diagonal low rank (HODLR) approximation
 - Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation
- 4 Numerical examples
 - Inversion for the basal friction coefficient in an ice sheet flow problem
 - Spatially varying blurring problem (if time permits)
- 5 Conclusions and outlook

Low rank approximation of the Hessian of the likelihood

Under the assumption of Gaussian noise and linearized parameter-to-observable map

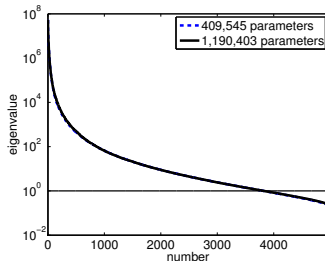
- The Hessian (in discrete form) takes the form:

$$\mathbf{H} = \underbrace{\mathbf{H}_d}_{\text{Hessian of the neg-log post}} + \underbrace{\mathbf{H}_{\text{like}}}_{\text{Hessian of the neg-log prior}} = \mathbf{\Gamma}_{\text{prior}}^{-1}$$

- Invoking low rank approximation of $\mathbf{H}_d$ combined with Sherman-Morrison-Woodbury formula to approximate the inverse Hessian (i.e., $\mathbf{\Gamma}_{\text{post}}$):

$$\mathbf{\Gamma}_{\text{post}} = \mathbf{H}^{-1} = (\mathbf{\Gamma}_{\text{prior}} \mathbf{H}_d + \mathbf{I})^{-1} \mathbf{\Gamma}_{\text{prior}} \approx \mathbf{\Gamma}_{\text{prior}} - \mathbf{W}_r \mathbf{D}_r \mathbf{W}_r^T$$

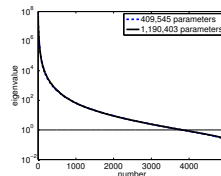
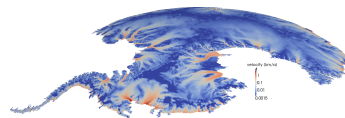
- $\mathbf{W}_r$: contains the largest r eigenvectors of $\mathbf{\Gamma}_{\text{prior}} \mathbf{H}_d$
- $\mathbf{D}_r = \text{diag}(\lambda_i / (\lambda_i + 1)) \in \mathbb{R}^{r \times r}$, where λ_i are eigenvalues of $\mathbf{\Gamma}_{\text{prior}} \mathbf{H}_d$
- Figure: spectrum of $\mathbf{\Gamma}_{\text{prior}} \mathbf{H}_d$ for Antarctica (1.19M parameters).



Details in: T. Isaac, N. Petra, G. Stadler, and O. Ghattas. JCP, 2015.

Motivation to go beyond “*global* low rank approximation”

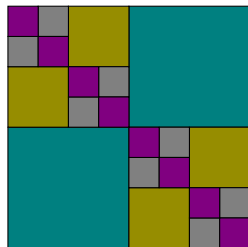
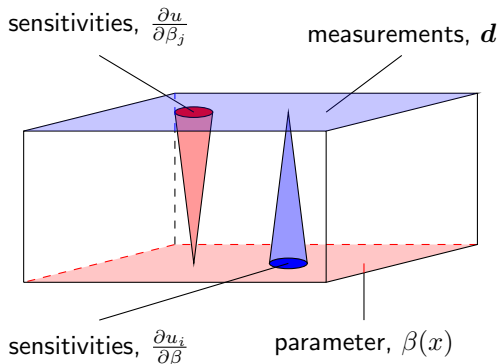
- Low rank Hessian approximation-based methods require computing twice as many linearized forward or adjoint partial differential equation (PDE) solves as the numerical rank of the Hessian.
- These methods are inefficient when the numerical rank of the Hessian is large, as is the case in continental scale ice sheet inverse problems.



Our goal is to use additional structure combined with hierarchical matrix compression to reduce the computational cost of solving the (Bayesian) ice sheet inverse problem.

Hierarchical off-diagonal low rank (HODLR) approximation

Exploit the narrow sensitivities



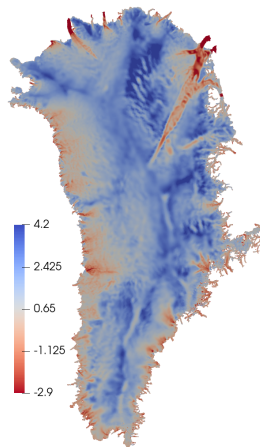
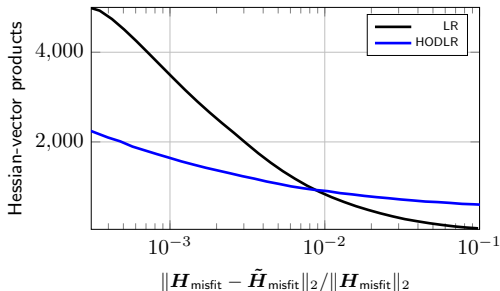
Details in:

Per-Gunnar Martinsson, *Fast direct solvers for elliptic PDEs*, SIAM, 2020.

J. Ballani and D. Kressner, *Matrices with hierarchical low rank structures*, Springer, 2016.

Rank-structure approximation costs

Numerical results obtained using Albany/Sandia



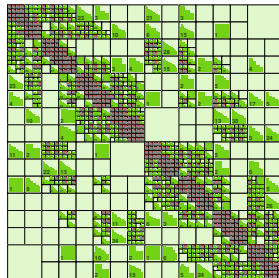
Details in: T. Hartland, G. Stadler, M. Perego, K. Liegeois, N. Petra. "Hierarchical off-diagonal low-rank approximation of Hessians in inverse problems, with application to ice sheet model initialization", *Inverse Problems*, 39 (8), 2023.

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

- Matrix is reordered and subdivided recursively into blocks.
- Many off-diagonal blocks are low rank.
- The matrix as a whole may be high rank.
- Work and memory required for an $N \times N$ $\mathcal{H}$ -matrix with rank k blocks:

$$O(k^a N (\log N)^b)$$

- $a, b \in \{0, 1, 2, 3\}$ depends on the type of $\mathcal{H}$ -matrix used and the operation being performed
- typical operations: matrix-vector products, matrix-matrix addition, matrix-matrix multiplication, matrix factorization, matrix inversion



Details in:

- Hackbusch, *Hierarchical matrices: algorithms and analysis*, Springer, 2015.
- L. Grasedyck and W. Hackbusch, Construction and arithmetics of H-matrices, *Computing*, 70, 2003.

Hierarchical matrix versus matrix free

- Classical methods for building $\mathcal{H}$ -matrix require matrix entries $\mathbf{H}_{ij}$.
- New algebraic methods based on “peeling process” can build $\mathcal{H}$ -matrix from matrix-vector products.
- **Problem:** peeling process better than low rank, but **still expensive** (e.g., ❶).
- Here we build the $\mathcal{H}$ -matrix faster by taking advantage of the problem structure (e.g., ❷), i.e.,
 - **Local sensitivities**
 - **Local mean-displacement invariance**
 - **Non-negative impulse responses.**

Details in:

- ❶ T. Hartland, G. Stadler, M. Perego, K. Liegeois, N. Petra. “Hierarchical off-diagonal low-rank approximation of Hessians in inverse problems, with application to ice sheet model initialization”, *Inverse Problems*, 39 (8), 2023.
- ❷ N. Alger, T. Hartland, N. Petra, O. Ghattas. “Point spread function approximation of high rank Hessians with locally supported non-negative integral kernels”, *SIAM Journal on Scientific Computing* (SISC), 46 (3), 2024 (<http://arxiv.org/abs/2307.03349>).

Fast high rank Hessian approximation for Bayesian inverse problems

- **Stage 1: Point spread function (PSF) approximation of (high rank) Hessians**
 - **Apply** the Hessian to a delta distribution (i.e., point source, impulse) to compute **impulse responses** of the Hessian operator at scattered points.
 - **Interpolate** these impulse responses to **approximate the entries** of the integral kernel
- **Stage 2: $\mathcal{H}$ -matrix compression:**
 - Convert the PSF Hessian approximation to **hierarchical matrix** format.
 - Invert the compressed matrix using fast hierarchical matrix arithmetic.
- **Stage 3: Hessian-approximation via $\mathcal{H}$ -matrix compression:**
 - Use this approximation to **precondition linear systems** involving the Hessian (e.g., solve the Newton system to compute the MAP point) and/or **draw samples from the posterior**.

Details in: N. Alger, T. Hartland, N. Petra, O. Ghattas. "Point spread function approximation of high rank Hessians with locally supported non-negative integral kernels", SIAM Journal on Scientific Computing (SISC), 46 (3), 2024 (<http://arxiv.org/abs/2307.03349>).

The computational cost of the PSF-based method

- **Computing impulse response moments and batches:**

$$1 + d + d(d + 1)/2 + n_b \quad \text{operator applications}$$

- d : spatial dimension; n_b : number of impulse response batches
- In a typical application one might have $d = 2$ and $n_b = 5$, in which case a modest 11 operator applications are required.

- **Building the $\mathcal{H}$ -matrix:**

$$O\left((k_h N \log N) (k_n \log N + k_n^3)\right) \quad \text{elementary operations}$$

- k_h : $\mathcal{H}$ -matrix rank; N : parameter size; k_n : number of nearest neighbors

- **Performing linear algebra operations with the $\mathcal{H}$ -matrix:**

- One matrix-matrix addition to add the $\mathcal{H}$ -matrix approximation of the data misfit term to the regularization term.
- One matrix-matrix addition to enforce symmetry.
- Few matrix-matrix additions and matrix factorizations to handle negative eigenvalues.
- These $\mathcal{H}$ -matrix linear algebra operations (as implemented in the HLIBPro library) were less costly than the PDE solves.

- 1 Motivation
- 2 Bayesian inverse problems governed by PDEs (preliminaries)
- 3 Exploiting low-dimensional structure in Bayesian inverse problems governed by PDEs
 - (Global) low rank approximation
 - Hierarchical off-diagonal low rank (HODLR) approximation
 - Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation
- 4 Numerical examples
 - Inversion for the basal friction coefficient in an ice sheet flow problem
 - Spatially varying blurring problem (if time permits)
- 5 Conclusions and outlook

Example: Ice sheet inverse problem

Nonlinear Stokes ice sheet model (for viscous, shear-thinning, incompressible fluid)

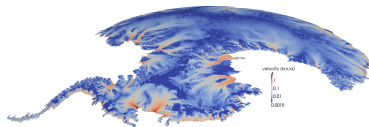
Invoking the balance of mass and linear momentum:

$$-\nabla \cdot [2\gamma(\mathbf{u}, n) \dot{\epsilon}_{\mathbf{u}} - \mathbf{I}p] = \rho \mathbf{g} \quad \text{in } \Omega$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega$$

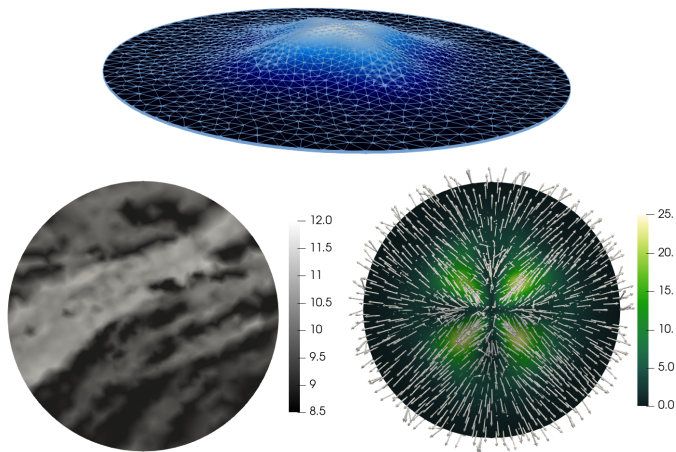
$$\boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n} = \mathbf{0} \quad \text{on } \Gamma_t$$

$$\mathbf{u} \cdot \mathbf{n} = 0, \mathbf{T} \boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n} + \exp(\beta) \mathbf{T} \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_b$$



- $\mathbf{u}$ ice flow velocity, p pressure
- $\boldsymbol{\sigma}_{\mathbf{u}} = -\mathbf{I}p + 2\gamma(\mathbf{u}, n) \dot{\epsilon}_{\mathbf{u}}$ stress tensor
- $\dot{\epsilon}_{\mathbf{u}} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$ strain rate tensor
- $\gamma(\mathbf{u}, n) = \frac{1}{2} A^{-\frac{1}{n}} \dot{\epsilon}_{\text{II}}^{\frac{1-n}{2n}}$ effective viscosity
- $\dot{\epsilon}_{\text{II}} = \frac{1}{2} \text{tr}(\dot{\epsilon}_{\mathbf{u}}^2)$ second invariant of the strain rate tensor
- ρ density, $\mathbf{g}$ gravity
- $\mathbf{n}$ unit normal vector
- β log basal sliding coefficient
- $\mathbf{T} = \mathbf{I} - \mathbf{n} \otimes \mathbf{n}$ tangential operator
- Γ_t and Γ_b top and base boundaries

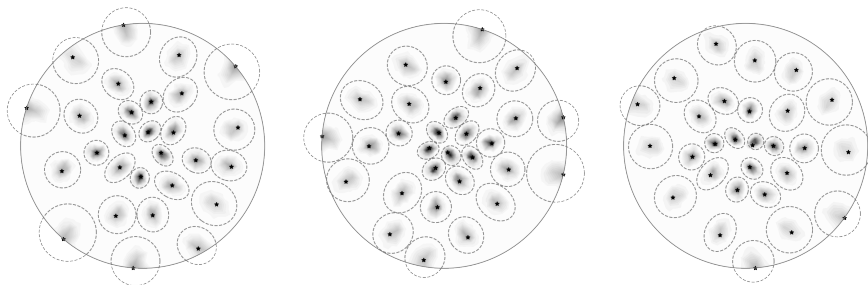
Example: Ice sheet inverse problem



The ice sheet geometry and discretization (top), the “true” parameter (bottom left) and the corresponding velocity field (bottom right).

Example: Ice sheet inverse problem

Impulse responses



- Black stars are point source locations.
- Shading shows the magnitude of the normalized impulse responses (darker means larger function values).
- Dashed gray ellipses are estimated impulse response support ellipsoids based on the moment method.

Example: Ice sheet inverse problem

Convergence history for solving the Stokes inverse problem using inexact Newton PCG

Iter	PSF (5)			REG			NONE		
	#CG	#Stokes	$\ g\ $	#CG	#Stokes	$\ g\ $	#CG	#Stokes	$\ g\ $
0	1	4	1.9e+7	3	8	1.9e+7	1	4	1.9e+7
1	2	6	6.1e+6	8	18	8.4e+6	2	6	6.1e+6
2	4	10	2.6e+6	16	34	4.1e+6	4	10	2.6e+6
3	2	6+22	6.9e+5	34	70	1.8e+6	14	30	6.9e+5
4	3	8	4.4e+4	52	106	5.6e+5	29	60	1.3e+5
5	5	12	2.2e+3	79	160	9.4e+4	38	78	1.0e+4
6	0	2	1.1e+1	102	206	6.5e+3	58	118	1.8e+2
7	—	—	—	151	304	1.2e+2	0	2	5.5e-1
8	—	—	—	0	2	2.9e-1	—	—	—
Total	17	70	—	445	908	—	146	308	—

- #CG: the number of PCG iterations used to solve the Newton system.
- #Stokes: the total number of Stokes PDE solves performed in each Newton iteration.
- $\|g\|$: the l^2 norm of the gradient.

Example: Ice sheet inverse problem

Convergence history for solving the Stokes inverse problem using inexact Newton PCG

Iter	PSF (5)			REG			NONE		
	#CG	#Stokes	$\ g\ $	#CG	#Stokes	$\ g\ $	#CG	#Stokes	$\ g\ $
0	1	4	1.9e+7	3	8	1.9e+7	1	4	1.9e+7
1	2	6	6.1e+6	8	18	8.4e+6	2	6	6.1e+6
2	4	10	2.6e+6	16	34	4.1e+6	4	10	2.6e+6
3	2	6+22	6.9e+5	34	70	1.8e+6	14	30	6.9e+5
4	3	8	4.4e+4	52	106	5.6e+5	29	60	1.3e+5
5	5	12	2.2e+3	79	160	9.4e+4	38	78	1.0e+4
6	0	2	1.1e+1	102	206	6.5e+3	58	118	1.8e+2
7	—	—	—	151	304	1.2e+2	0	2	5.5e-1
8	—	—	—	0	2	2.9e-1	—	—	—
Total	17	70	—	445	908	—	146	308	—

- #CG: the number of PCG iterations used to solve the Newton system.
- #Stokes: the total number of Stokes PDE solves performed in each Newton iteration.
- $\|g\|$: the l^2 norm of the gradient.

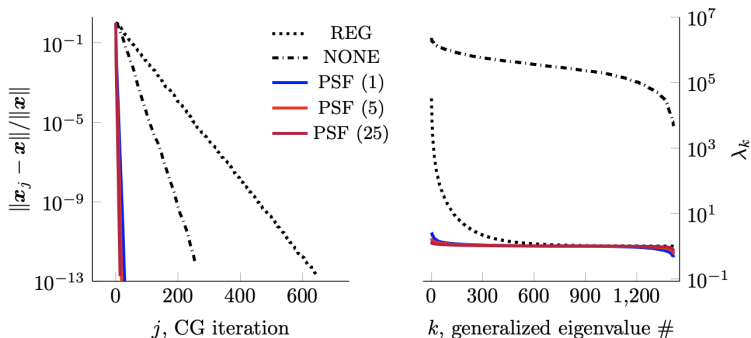
Example: Ice sheet inverse problem



The log basal sliding parameter computed by solving the ice sheet inverse problem with noise levels: 25% (left), 5% (middle), and 1% (right).

Example: Ice sheet inverse problem

Convergence history and eigenvalue clustering



Left: convergence history for solving $\mathbf{H}\mathbf{x} = \mathbf{b}$ using PCG, where $\mathbf{b}$ has i.i.d. random entries drawn from the standard Gaussian distribution and $\mathbf{H}$ is evaluated at the solution of the inverse problem. Right: eigenvalues of the generalized eigenvalue problem $\mathbf{H}\mathbf{u}_k = \lambda_k \tilde{\mathbf{H}}\mathbf{u}_k$.

- 1 Motivation
- 2 Bayesian inverse problems governed by PDEs (preliminaries)
- 3 Exploiting low-dimensional structure in Bayesian inverse problems governed by PDEs
 - (Global) low rank approximation
 - Hierarchical off-diagonal low rank (HODLR) approximation
 - Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation
- 4 Numerical examples
 - Inversion for the basal friction coefficient in an ice sheet flow problem
 - Spatially varying blurring problem (if time permits)
- 5 Conclusions and outlook

- Hessian approximations or preconditioners are essential for Bayesian inverse problems governed by partial differential equations.
- This is particularly important when the model must be continually updated with new information, as with digital twins.
- 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.

Extending to large-scale with simpler Gaussian PSF

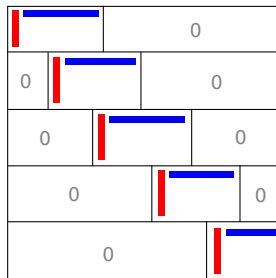
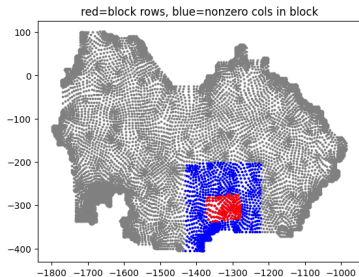
Joint: N. Alger (UT Austin), T. Hartland (LLNL), T. Isaac (NVIDIA), O. Ghattas (UT Austin)

- Don't bother computing impulse responses. Use moments only.
- Use Gaussian approximation of kernel:

$$(Hu)(y) \approx \int_x \Phi(y, x) u(x) dx$$

$$\Phi(y, x) := \frac{V(x)}{2\pi|\Sigma(x)|^{1/2}} \exp\left(-\frac{1}{2}(y - \mu(x))^T \Sigma(x)^{-1}(y - \mu(x))\right)$$

- Use block row low rank (BRLR) structure instead of H-matrix:



Joint parameter and state dimension reduction for Bayesian ice sheet inverse problems

Joint work with: P. Mrada, K. Kim, T. Cui, B. Peherstorfer and S. Minkoff

- 1 **Parameter dimension reduction:** identify a likelihood-informed parameter subspace that captures parameter directions where the change from prior to posterior is most significant and reduce the parameter dimension:

$$\pi_{\text{post}}(\mathbf{m}|\mathbf{d}) \approx \underbrace{\pi_{\text{like}}(\mathbf{d} | \Phi \mathbf{m}_r) \pi_{\text{prior}}(\mathbf{m}_r)}_{\text{parameter-reduced posterior } \tilde{\pi}_{\text{post}}(\mathbf{m}_r|\mathbf{d})} \underbrace{\pi_{\text{prior}}(\mathbf{m}_{\perp})}_{\text{complement prior}}$$

- 2 **State dimension reduction via POD and DEIM:** accelerate the **nonlinear forward model** evaluations by identifying a low-dimensional subspace of the state and construct a reduced version of the forward model:
Key idea: POD-DEIM reduced Newton system:

$$\mathbf{V}^{\top} (\mathbf{K}_{\text{lin}} + \mathbf{W}(\mathbf{S}^{\top} \mathbf{W})^{-1} \mathbf{S}^{\top} \mathbf{K}_{\text{nl}}) \mathbf{V} \tilde{\mathbf{u}} = -\mathbf{V}^{\top} (r_{\text{lin}} + \mathbf{W}(\mathbf{S}^{\top} \mathbf{W})^{-1} \mathbf{S}^{\top} r_{\text{nl}})$$

Details in:

T. Cui, Y. Marzouk and K. Willcox, *Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction*, JCP, 2016.

S. Chaturantabut and D.C. Sorensen, *Nonlinear Model Reduction via Discrete Empirical Interpolation* SIAM J. Sci. Comput., 2010

Accounting for the model error

using the Bayesian Approximation Error (BAE) approach

- **Key idea:** explore the full posterior using a reduced order model and **take into account the model error**:

$$\mathbf{d} = \mathcal{F}(\mathbf{m}) + \boldsymbol{\eta} = (\tilde{\mathcal{F}}(\tilde{\mathbf{m}}) + \boldsymbol{\varepsilon}) + \boldsymbol{\eta} = \tilde{\mathcal{F}}(\tilde{\mathbf{m}}) + \boldsymbol{\nu}$$

- $\mathcal{F}(\mathbf{m})$: the accurate (*high fidelity*) parameter-to-observable map
- $\tilde{\mathcal{F}}(\tilde{\mathbf{m}})$: the approximative (*low fidelity*) parameter-to-observable map

$$\boldsymbol{\eta} = \mathbf{d} - \mathcal{F}(\mathbf{m}) \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_{\text{noise}}) \quad (\text{data noise})$$

$$\boldsymbol{\varepsilon} = \mathcal{F}(\mathbf{m}) - \tilde{\mathcal{F}}(\tilde{\mathbf{m}}) \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\bar{\boldsymbol{\varepsilon}}, \mathbf{\Gamma}_{\varepsilon}) \quad (\text{approximation/model error})$$

$$\boldsymbol{\nu} = \boldsymbol{\varepsilon} + \boldsymbol{\eta} \quad \boldsymbol{\nu} \sim \mathcal{N}(\bar{\boldsymbol{\nu}}, \mathbf{\Gamma}_{\nu}) \quad (\text{total error})$$

- $\bar{\boldsymbol{\nu}} = \bar{\boldsymbol{\varepsilon}} + \bar{\boldsymbol{\eta}}$: the mean; $\mathbf{\Gamma}_{\nu} = \mathbf{\Gamma}_{\varepsilon} + \mathbf{\Gamma}_{\text{noise}}$: the *total error* covariance

Details in:

- Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Springer, 2005
- Jari Kaipio and Ville Kolehmainen, *Approximate marginalization over modeling errors and uncertainties in inverse problems*, Bayesian Theory and Applications, 2013
- Ruanui Nicholson, Noemi Petra, and Jari Kaipio, "Estimation of the Robin coefficient field in a Poisson problem with uncertain conductivity field", Inverse Problems, 2018.

Accounting for the model error

using the Bayesian Approximation Error (BAE) approach

- **Key idea:** explore the full posterior using a reduced order model and **take into account the model error**:

$$\mathbf{d} = \mathcal{F}(\mathbf{m}) + \boldsymbol{\eta} = (\tilde{\mathcal{F}}(\tilde{\mathbf{m}}) + \boldsymbol{\varepsilon}) + \boldsymbol{\eta} = \tilde{\mathcal{F}}(\tilde{\mathbf{m}}) + \boldsymbol{\nu}$$

- $\mathcal{F}(\mathbf{m})$: the accurate (*high fidelity*) parameter-to-observable map
- $\tilde{\mathcal{F}}(\tilde{\mathbf{m}})$: the approximative (*low fidelity*) parameter-to-observable map

$$\boldsymbol{\eta} = \mathbf{d} - \mathcal{F}(\mathbf{m}) \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_{\text{noise}}) \quad (\text{data noise})$$

$$\boldsymbol{\varepsilon} = \mathcal{F}(\mathbf{m}) - \tilde{\mathcal{F}}(\tilde{\mathbf{m}}) \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\bar{\boldsymbol{\varepsilon}}, \mathbf{\Gamma}_{\boldsymbol{\varepsilon}}) \quad (\text{approximation/model error})$$

$$\boldsymbol{\nu} = \boldsymbol{\varepsilon} + \boldsymbol{\eta} \quad \boldsymbol{\nu} \sim \mathcal{N}(\bar{\boldsymbol{\nu}}, \mathbf{\Gamma}_{\boldsymbol{\nu}}) \quad (\text{total error})$$

- $\bar{\boldsymbol{\nu}} = \bar{\boldsymbol{\varepsilon}} + \bar{\boldsymbol{\eta}}$: the mean; $\mathbf{\Gamma}_{\boldsymbol{\nu}} = \mathbf{\Gamma}_{\boldsymbol{\varepsilon}} + \mathbf{\Gamma}_{\text{noise}}$: the *total error* covariance

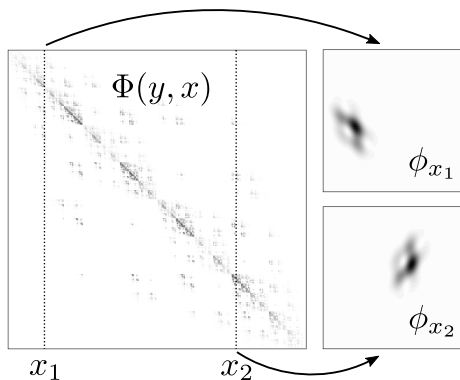
Details in:

- Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Springer, 2005
- Jari Kaipio and Ville Kolehmainen, *Approximate marginalization over modeling errors and uncertainties in inverse problems*, Bayesian Theory and Applications, 2013
- Ruanui Nicholson, Noemi Petra, and Jari Kaipio, "Estimation of the Robin coefficient field in a Poisson problem with uncertain conductivity field", *Inverse Problems*, 2018.

- N. Alger, T. Hartland, N. Petra, O. Ghattas. “Point spread function approximation of high rank Hessians with locally supported non-negative integral kernels”, *SIAM Journal on Scientific Computing (SISC)*, 46 (3), 2024 (<http://arxiv.org/abs/2307.03349>).
- T. Hartland, G. Stadler, M. Perego, K. Liegeois, N. Petra. “Hierarchical off-diagonal low-rank approximation of Hessians in inverse problems, with application to ice sheet model initialization”, **Inverse Problems**, 39 (8), 2023.
- 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 (JCP)**, 296, 348-368 (2015).
- U. Villa, N. Petra, and O. Ghattas, “hIPPYlib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs: Part I: Deterministic Inversion and Linearized Bayesian Inferences”, **ACM Transactions on Mathematical Software (TOMS)**, 47 (2), (2021).

Stage1: Hessian approximation via PSF

Example: Spatially varying blurring problem



Left: Matrix created by evaluating the integral kernel at mesh points. Right: Impulse responses ϕ_{x_i} associated with points x_i .

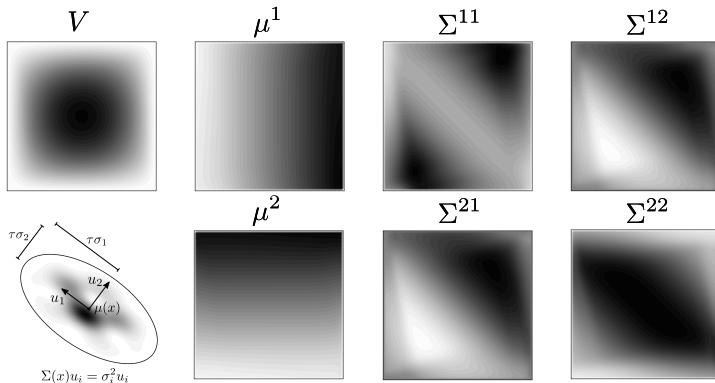
$$\Phi(y, x) = (1 - af(y, x))g(x) \exp\left(-\frac{1}{2}(h(y, x)^T C^{-1} h(y, x))\right)$$

Impulse responses can be thought of intuitively as “columns” of the integral kernel.

Stage1: Hessian approximation via PSF

Example: Spatially varying blurring problem

To choose batches of impulse responses, we form **ellipsoid estimates** for the supports of all ϕ_x via a “moment method”.



Impulse response moments and the ellipsoid support for an impulse response.

$\mu(x)$: location that ϕ_x is centered at. $\Sigma(x)$: a matrix with eigenvectors and eigenvalues that characterize the width of the support of ϕ_x . $V(x)$: scaling factor.

Stage1: Hessian approximation via PSF

Example: Spatially varying blurring problem

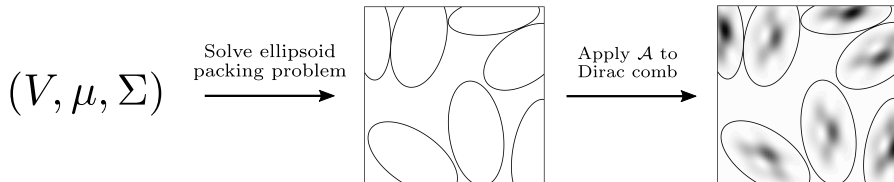


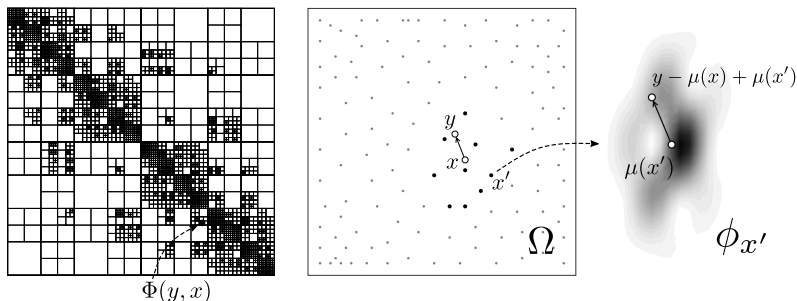
Illustration of the process to compute one impulse response batch.

- 1 Impulse response moments are first used to form ellipsoid shaped estimates of the supports of impulse responses.
- 2 Then, an ellipsoid packing problem is solved to choose batches of non-overlapping support ellipsoids.
- 3 The blurring operator (here denoted by $\mathcal{A}$) is applied to a Dirac comb (a weighted sum of point sources) associated with the points x_i , which correspond to the ellipsoids.

Batches of impulse responses may be thought of intuitively as sets of “columns” of the kernel.

Stage 2: PSF Hessian approximation to $\mathcal{H}$ -matrix

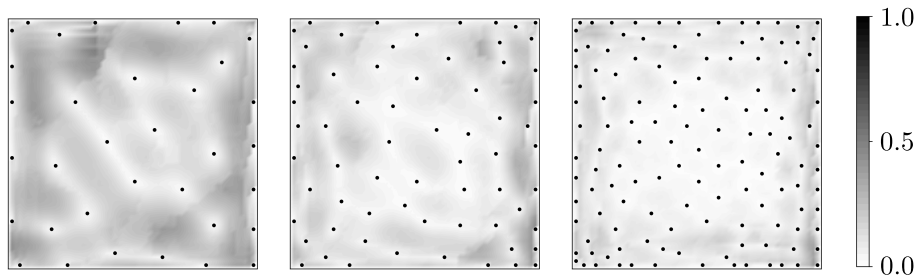
Kernel entry approximation via radial basis function interpolation



- **Left:** $\mathcal{H}$ -matrix structure. Computing an entry of this matrix requires evaluating the integral kernel, $\Phi(y, x)$, at a pair of points (y, x) .
- **Center:** Kernel evaluation points x and y (black circles), sample points for the approximation (light gray and black dots), and the k_n sample points, x' , that are nearest to x (black dots).
- **Right:** Known impulse responses at x' . Using radial basis function interpolation, the desired kernel entry is approximated as a weighted linear combination of translated and scaled versions of impulse responses at the points x' .

Stage 2: PSF Hessian approximation to $\mathcal{H}$ -matrix

Quality of the approximation

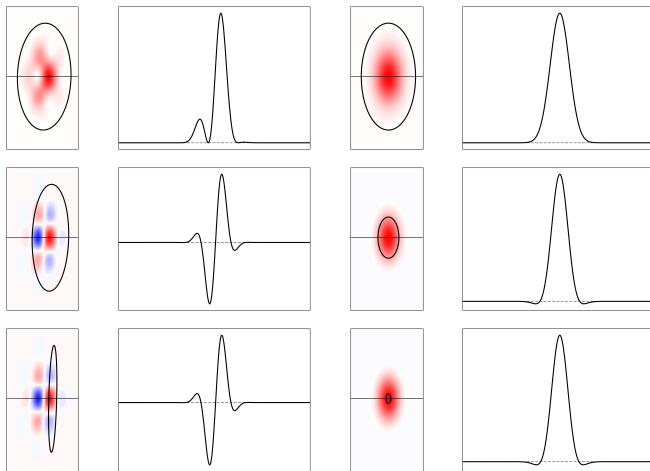


Relative error, $\|\Phi(\cdot, x) - \tilde{\Phi}(\cdot, x)\| / \|\Phi(\cdot, x)\|$, in the approximation of the “column” of the integral kernel associated with x , using 5 (left), 10 (center) and 20 (right) impulse response batches.

Adding more batches yields a more accurate approximation
but this leads to more operator applies.

Ellipsoid estimates for the supports of impulse responses

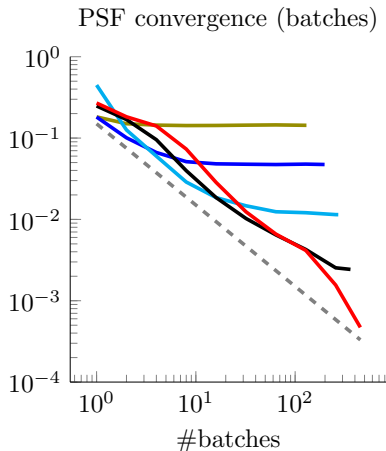
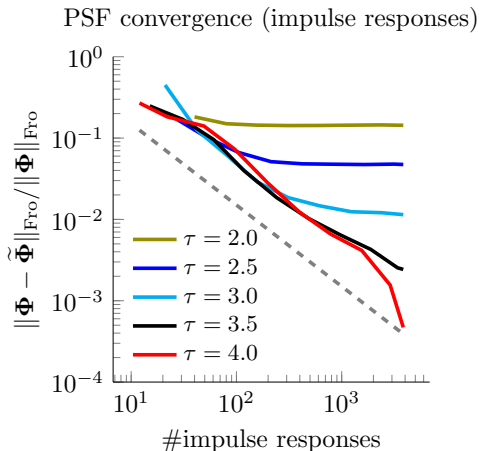
Blur kernel (left two columns) and a Ricker wavelet-type kernel (right two columns)



Spatially variant blur kernel (left two columns) and a Ricker wavelet-type kernel (right two columns) for progressively more significant negative numbers in the integral kernel (top to bottom).

Relative error in the PSF approximation of the kernel

Example: Spatially varying blurring problem



Left: Relative error vs. the total number of impulse responses used in the approximation. Right: Relative error vs. the number of impulse response batches. The dashed gray lines show linear convergence rates.

Comparison of computational cost

Example: Spatially varying blurring problem

	Error	#applies PSF	#applies HODLR	#applies RSVD
$L = 1$	20%	11	592	354
	10%	15	772	520
	5%	23	924	674
$L = 1/2$	20%	8	852	1316
	10%	10	1144	1916
	5%	12	1404	2456
$L = 1/3$	20%	7	932	2624
	10%	8	1264	3734
	5%	8	1520	4660

- L : scales the width of the impulse responses (influences the rank)
- **Error**: is the relative error in the approximation of the kernel measured in the Frobenius norm.
- **#applies**: number of operator applies required to achieve the given error tolerances, using the **PSF**, **HODLR** (hierarchical off diagonal low rank), and **GLR** (global low rank) methods.

Relevant literature

- **(Global) low rank approximation**

- 1 Flath, Wilcox, Akçelik, Hill, Van Bloemen Waanders, and Ghattas, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*, SISC, 2011.
- 2 Halko, Martinsson, and Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM review, 2011.
- 3 Ghattas and Willcox, *Learning physics-based models from data: perspectives from inverse problems and model reduction*, Acta Numerica, 2021.

- **Hierarchical ($\mathcal{H}$ -)matrix approximations**

- 1 Hackbusch, *Hierarchical matrices: algorithms and analysis*, Springer, 2015.
- 2 Martinsson, *Fast direct solvers for elliptic PDEs*, SIAM, 2020.
- 3 Lin, Lu, and Ying, *Fast construction of hierarchical matrix representation from matrix-vector multiplication*, JCP, 2011.
- 4 Ambartsumyan, Boukaram, Bui-Thanh, Ghattas, Keyes, Stadler, Turkiyyah, and Zampini, *Hierarchical matrix approximations of Hessians arising in inverse problems governed by PDEs*, SISC, 2020.
- 5 Hartland, Stadler, Perego, Liegeois, and Petra, *Hierarchical off-diagonal low-rank approximation of Hessians in inverse problems, with application to ice sheet model initialization*, Inverse Problems, 2023.
- 6 Chen, Anitescu, *Scalable physics-based maximum likelihood estimation using hierarchical matrices*, JUQ, 2023.

- **Point spread function approximation combined with hierarchical matrix ($\mathcal{H}$ -matrix) approximation**

- 1 Gentile, Courbin, and Meylan, *Interpolating point spread function anisotropy*, 2013.
- 2 Escande and Weiss, *Approximation of integral operators using product-convolution expansions*, Journal of Mathematical Imaging and Vision, 2017.
- 3 Alger, Rao, Myers, Bui-Thanh, and Ghattas, *Scalable matrix-free adaptive product-convolution approximation for locally translation-invariant operators*, SISC, 2019.