Structure-Exploiting Scalable Methods for Large-scale Bayesian Inverse Problems in High Dimensional Parameter Spaces

Tan Bui-Thanh

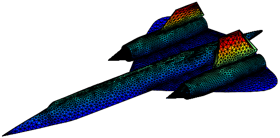
Center for Computational Geosciences and Optimization
Institute for Computational Engineering and Sciences (ICES)
The University of Texas at Austin

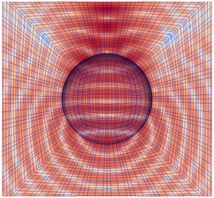
Joint work with Carsten Burstedde, Omar Ghattas, James R. Martin, Georg Stadler, and Lucas Wilcox

Multiscale Inverse Problems Workshop, University of Warwick
June 17-19

Large-scale computation under uncertainty

Inverse electromagnetic scattering



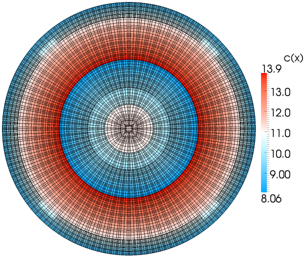


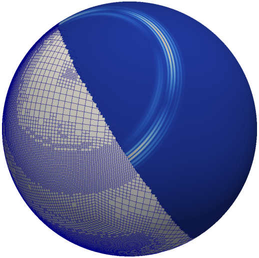
Randomness

- Random errors in measurements are unavoidable

Large-scale computation under uncertainty

Full wave form seismic inversion





Randomness

- Random errors in seismometer measurements are unavoidable

Large-scale uncertainty quantification in high dimensions

Common challenge

- Large-scale PDE forward solve (more than 10^8 DOFs)
- High dimensional parameter spaces (curse of dimensionality)
- Uncertainty Quantification (randomness)

Large-scale uncertainty quantification in high dimensions

Common challenge

- Large-scale PDE forward solve (more than 10^8 DOFs)
- High dimensional parameter spaces (curse of dimensionality)
- Uncertainty Quantification (randomness)

Solution 1: Reduce-then-sample

Exploit higher order derivatives to construct

- an accurate surrogate model

that is inexpensive to solve

Large-scale uncertainty quantification in high dimensions

Common challenge

- Large-scale PDE forward solve (more than 10^8 DOFs)
- High dimensional parameter spaces (curse of dimensionality)
- Uncertainty Quantification (randomness)

Solution 1: Reduce-then-sample

Exploit higher order derivatives to construct

- an accurate surrogate model

that is inexpensive to solve

Solution 2: Sample-then-reduce

Work directly with high-fidelity model but only explore important subspaces/directions

Outline

Reduce-then-sample

- Approach: Hessian-based adaptive Gaussian Process
- Application: Inverse Shape Electromagnetic Scattering

Outline

Reduce-then-sample

- Approach: Hessian-based adaptive Gaussian Process
- Application: Inverse Shape Electromagnetic Scattering

Sample-then-reduce

- Approach: Gaussian approximation and MCMC
- Application: Full Wave Form Seismic Wave Inversion

Reduce-then-sample

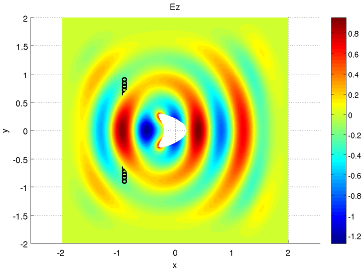
- Approach: Hessian-based adaptive Gaussian Process
- Application: Inverse Shape Electromagnetic Scattering

Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$

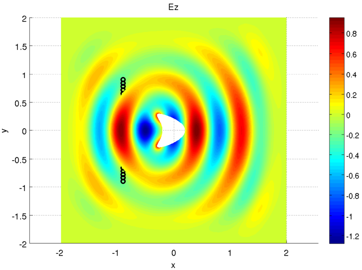


Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$



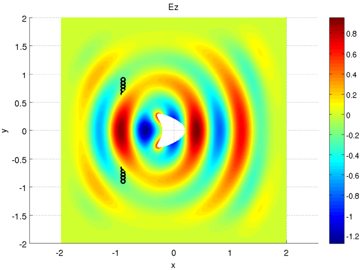
$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u}) \times \underbrace{\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})}_{\text{Computationally expensive forward model:}\mathbf{y}=\mathbf{G}(\mathbf{u})}$$

Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$



$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u}) \times \underbrace{\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})}_{\text{Computationally expensive forward model:}\mathbf{y}=\mathbf{G}(\mathbf{u})}$$

Approximate the likelihood

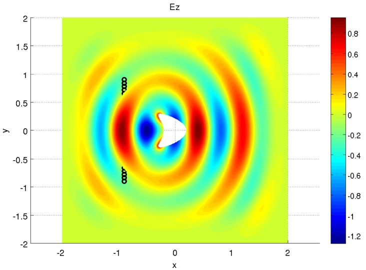
Reduced basis method, polynomial chaos, and etc

Inverse Shape Electromagnetic Scattering Problem

Maxwell Equations:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \qquad \text{(Faraday)}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \qquad \text{(Ampere)}$$



$$\pi_{\text{post}}(\mathbf{u}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\mathbf{u}) \times \underbrace{\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\mathbf{u})}_{\text{Computationally expensive forward model:}\mathbf{y}=\mathbf{G}(\mathbf{u})}$$

Approximate the likelihood

Reduced basis method, polynomial chaos, and etc

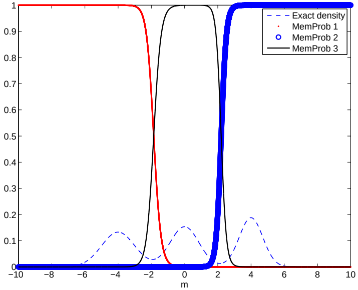
Approximate the posterior

Propose an **Hessian-based Adaptive Gaussian Process** response surface

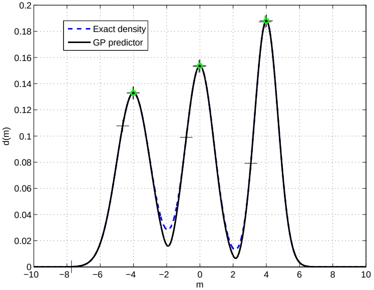
Hessian-based Adaptive Gaussian Process

Main idea ("mitigating" the curse of dimensionality)

1. Use Adaptive Sampling Algorithm to find the modes
2. Approximate the covariance matrix (Hessian inverse)
3. Partition parameter space using membership probabilities
4. Approximate the posterior with local Gaussian in subdomains
5. Glue all the local Gaussian approximations



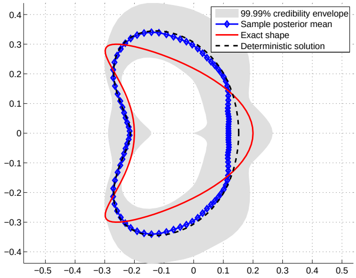
Membership probabilities



GP predictor versus exact

Inverse shape electromagnetic scattering

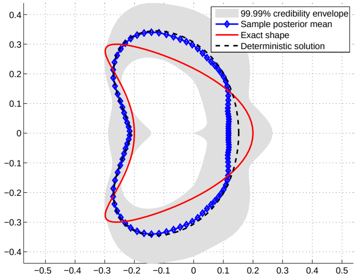
- discontinuous Galerkin discretization with 80,892 state variables
- 24 shape parameters
- 1 million MCMC simulations for the Gaussian process response surface



Details in: Bui-Thanh, T., Ghattas, O., and Higdon, D., *Adaptive Hessian-based Non-stationary Gaussian Process Response Surface Method for Probability Density Approximation with Application to Bayesian Solution of Large-scale Inverse Problems*, **SIAM Journal on Scientific Computing**, 34(6), pp. A2837–A2871, 2012.

Inverse shape electromagnetic scattering

- discontinuous Galerkin discretization with 80,892 state variables
- 24 shape parameters
- 1 million MCMC simulations for the Gaussian process response surface



	Offline time
Gaussian process	33 hours
Exact Posterior	0 hours

	Online time
Gaussian process	0.96 hours
Exact Posterior	8802.35 hours

Details in: Bui-Thanh, T., Ghattas, O., and Higdon, D., *Adaptive Hessian-based Non-stationary Gaussian Process Response Surface Method for Probability Density Approximation with Application to Bayesian Solution of Large-scale Inverse Problems*, **SIAM Journal on Scientific Computing**, 34(6), pp. A2837–A2871, 2012.

Sample-then-reduce

- Approach: Gaussian approximation and MCMC
- Application: Full Wave Form Seismic Wave Inversion

Full wave form seismic wave inversion

$$\frac{\partial \boldsymbol{E}}{\partial t} = \frac{1}{2}\left(\nabla \boldsymbol{v} + \nabla^T \boldsymbol{v}\right),$$
$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = \nabla \cdot (\mathsf{C}\boldsymbol{E}) + \boldsymbol{f}$$

Animated by Greg Abram

Strain-velocity formulation

- $\boldsymbol{E}$: strain tensor,
- $\boldsymbol{v}$: velocity vector,
- ρ: density,
- $\boldsymbol{e}_i$: ith unit vector,

- I: fourth-order identity tensor,
- $\boldsymbol{I}$: second-order identity tensor,
- $\boldsymbol{f}$: external volumetric forces,
- C: four-order material tensor.

Inverse problem statement

- Earth surface velocity at given locations is recorded
- Infer the wave velocities $c_s = \sqrt{\mu/\rho}$ and $c_p = \sqrt{(\lambda + 2\mu)/\rho}$

Infinite dimensional Bayesian statistical inference

Bayes' theorem in infinite dimensions

A Bayes' theorem in infinite dimensional spaces (Stuart 2010)

$$\frac{d\mu}{d\mu_0}(\mathbf{u}) \propto \exp\left(-\Phi\left(\boldsymbol{y}^{\text{obs}}, \mathbf{u}\right)\right)$$

defines the Radon-Nikodym derivative of the posterior probability measure μ with respect to the prior measure μ_0.

- μ_0: prior probability measure
- μ: posterior probability measure
- $\Phi\left(\boldsymbol{y}^{\text{obs}}, \mathbf{u}\right)$: misfit functional
- $\mathbf{u}$: unknown parameter
- $\boldsymbol{y}^{\text{obs}}$: observation data

Infinite dimensional Bayesian statistical inference

Prior smoothness

Definitions and assumptions

- Prior distribution of $\mathbf{u}$ is a Gaussian measure $\mu_0 = \mathcal{N}\left(\mathbf{u}_0, \mathcal{C}_0\right)$ on $L^2\left(\Omega\right)$
- $\mathcal{C}_0 = \mathcal{A}^{-\alpha}$ is the prior covariance operator: trace class operator
- $\mathcal{A}$ is a Laplace-like operator, e.g., $\mathcal{A} = \theta I - \beta\Delta$.
- $\mathbf{u}_0$ lies in the Cameron-Martin space $E = \mathcal{R}\left(\mathcal{C}_0^{1/2}\right) = \mathcal{H}^\alpha$ of $\mathcal{C}_0$

Infinite dimensional Bayesian statistical inference

Prior smoothness

Definitions and assumptions

- Prior distribution of $\mathbf{u}$ is a Gaussian measure $\mu_0 = \mathcal{N}\left(\mathbf{u}_0, \mathcal{C}_0\right)$ on $L^2\left(\Omega\right)$
- $\mathcal{C}_0 = \mathcal{A}^{-\alpha}$ is the prior covariance operator: trace class operator
- $\mathcal{A}$ is a Laplace-like operator, e.g., $\mathcal{A} = \theta I - \beta \Delta$.
- $\mathbf{u}_0$ lies in the Cameron-Martin space $E = \mathcal{R}\left(\mathcal{C}_0^{1/2}\right) = \mathcal{H}^\alpha$ of $\mathcal{C}_0$

Prior smoothness

Assume $\alpha > d/2$ and $m_0 \in \mathcal{H}^\alpha$, then the Gaussian measure $\mu_0 = \mathcal{N}\left(m_0, \mathcal{A}^{-\alpha}\right)$ has full measure on $C\left(\overline{\Omega}\right)$, namely, $\mu_0\left(C\left(\overline{\Omega}\right)\right) = 1$.

We choose $\alpha = 2$ for $d = 3$

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution about the MAP

- Compute the MAP
- Linearize the forward map about the MAP $\boldsymbol{y}^{\text{obs}} = f_0 + A(\mathbf{u}) + \eta$

Posterior becomes a Gaussian measure

$$\mathbf{u}|\boldsymbol{y}^{\text{obs}} \sim \mu = \mathcal{N}(\mathbf{m}, \mathcal{C}),$$

posterior mean

$$\mathbf{m} = \mathbb{E}[\mathbf{u}] = \mathbf{u}_0 + C_0 A^* (\Gamma + A\mathcal{C}_0 A^*)^{-1} \left(\boldsymbol{y}^{\text{obs}} - f_0 - A\mathbf{u}_0\right)$$

posterior covariance operator

$$\mathcal{C} = \left(A^* \Gamma^{-1} A + \mathcal{C}_0^{-1}\right)^{-1}$$

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution: Low rank approximation

posterior covariance operator: A low rank approximation

$$\begin{aligned}
\mathcal{C} &= \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1} \\
&= \mathcal{C}_0^{1/2}\left(\mathcal{C}_0^{1/2}\boldsymbol{A}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\,\mathcal{C}_0^{1/2} + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&\approx \mathcal{C}_0^{1/2}\left(\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^* + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&= \mathcal{C}_0 - \mathcal{C}_0^{1/2}\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^*\mathcal{C}_0^{1/2}
\end{aligned}$$

- Low rank approximation only involves incremetal forward and incremental adjoint solve
- Then use Sherman-Morrison-Woodbury
- Relative to the prior uncertainty, the posterior uncertainty is reduced when observations are made.

The gradient computation

- Gradient expression (for general tensor C) given by

$$\mathcal{G}(\mathsf{C}) := \int_0^T \left[\frac{1}{2}(\nabla \boldsymbol{w} + \nabla \boldsymbol{w}^T) \otimes \boldsymbol{E}\right] dt + \mathcal{R}'(\mathsf{C})$$

- where $\boldsymbol{v}, \boldsymbol{E}$ satisfy the *forward wave propagation equations*

$$\begin{aligned}
\rho \frac{\partial \boldsymbol{v}}{\partial t} - \nabla \cdot (\mathsf{C}\boldsymbol{E}) &= \boldsymbol{f} && \text{in } \Omega \times (0,T) \\
-\mathsf{C}\frac{\partial \boldsymbol{E}}{\partial t} + \frac{1}{2}\mathsf{C}(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^T) &= \mathbf{0} && \text{in } \Omega \times (0,T) \\
\rho \boldsymbol{v} = \mathsf{C}\boldsymbol{E} &= \mathbf{0} && \text{in } \Omega \times \{t = 0\} \\
\mathsf{C}\boldsymbol{E}\boldsymbol{n} &= \mathbf{0} && \text{on } \Gamma \times (0,T)
\end{aligned}$$

- $\boldsymbol{w}, \boldsymbol{D}$ (adjoint velocity, strain) satisfy the *adjoint wave propagation equations*

$$\begin{aligned}
-\rho \frac{\partial \boldsymbol{w}}{\partial t} - \nabla \cdot (\mathsf{C}\boldsymbol{D}) &= -\mathcal{B}(\boldsymbol{v} - \boldsymbol{v}^{\text{obs}}) && \text{in } \Omega \times (0,T) \\
\mathsf{C}\frac{\partial \boldsymbol{D}}{\partial t} + \frac{1}{2}\mathsf{C}(\nabla \boldsymbol{w} + \nabla \boldsymbol{w}^T) &= \mathbf{0} && \text{in } \Omega \times (0,T) \\
\rho \boldsymbol{w} = \mathsf{C}\boldsymbol{D} &= \mathbf{0} && \text{in } \Omega \times \{t = T\} \\
\mathsf{C}\boldsymbol{D}\boldsymbol{n} &= \mathbf{0} && \text{on } \Gamma \times (0,T)
\end{aligned}$$

Computation of action of Hessian in given direction

- Action of the Hessian operator in direction $\tilde{\mathsf{C}}$ at a point C given by

$$\mathcal{H}(\mathsf{C})\tilde{\mathsf{C}} := \int_0^T \left[\frac{1}{2}(\nabla\tilde{\boldsymbol{w}} + \nabla\tilde{\boldsymbol{w}}^T) \otimes \boldsymbol{E} + \frac{1}{2}(\nabla\boldsymbol{w} + \nabla\boldsymbol{w}^T) \otimes \tilde{\boldsymbol{E}}\right] dt + \mathcal{R}''(\mathsf{C})\tilde{\mathsf{C}}$$

- where $\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{E}}$ satisfy the *incremental forward wave propagation equations*

$$\begin{aligned}
\rho\frac{\partial\tilde{\boldsymbol{v}}}{\partial t} - \nabla\cdot(\mathsf{C}\tilde{\boldsymbol{E}}) &= \nabla\cdot(\tilde{\mathsf{C}}\boldsymbol{E}) && \text{in } \Omega\times(0,T)\\
-\mathsf{C}\frac{\partial\tilde{\boldsymbol{E}}}{\partial t} + \frac{1}{2}\mathsf{C}(\nabla\tilde{\boldsymbol{v}} + \nabla\tilde{\boldsymbol{v}}^T) &= \boldsymbol{0} && \text{in } \Omega\times(0,T)\\
\rho\tilde{\boldsymbol{v}} = \mathsf{C}\tilde{\boldsymbol{E}} &= \boldsymbol{0} && \text{in } \Omega\times\{t=0\}\\
\mathsf{C}\tilde{\boldsymbol{E}}\boldsymbol{n} &= -\tilde{\mathsf{C}}\boldsymbol{E}\boldsymbol{n} && \text{on } \Gamma\times(0,T)
\end{aligned}$$

- and $\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{D}}$ satisfy the *incremental adjoint wave propagation equations*

$$\begin{aligned}
-\rho\frac{\partial\tilde{\boldsymbol{w}}}{\partial t} - \nabla\cdot(\mathsf{C}\tilde{\boldsymbol{D}}) &= \nabla\cdot(\tilde{\mathsf{C}}\boldsymbol{D}) - \mathcal{B}\tilde{\boldsymbol{v}} && \text{in } \Omega\times(0,T)\\
\mathsf{C}\frac{\partial\tilde{\boldsymbol{D}}}{\partial t} + \frac{1}{2}\mathsf{C}(\nabla\tilde{\boldsymbol{w}} + \nabla\tilde{\boldsymbol{w}}^T) &= \boldsymbol{0} && \text{in } \Omega\times(0,T)\\
\rho\tilde{\boldsymbol{w}} = \mathsf{C}\tilde{\boldsymbol{D}} &= \boldsymbol{0} && \text{in } \Omega\times\{t=T\}\\
\mathsf{C}\tilde{\boldsymbol{D}}\boldsymbol{n} &= -\tilde{\mathsf{C}}\boldsymbol{D}\boldsymbol{n} && \text{on } \Gamma\times(0,T)
\end{aligned}$$

posterior covariance operator: A low rank approximation

$$
\begin{aligned}
\mathcal{C} &= \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1} \\
&= \mathcal{C}_0^{1/2}\left(\mathcal{C}_0^{1/2}\boldsymbol{A}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\,\mathcal{C}_0^{1/2} + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&\approx \mathcal{C}_0^{1/2}\left(\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^* + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&= \mathcal{C}_0 - \mathcal{C}_0^{1/2}\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^*\mathcal{C}_0^{1/2}
\end{aligned}
$$

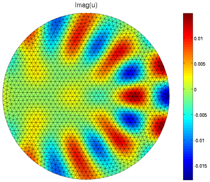
- Low rank approximation only involves incremetal forward and incremental adjoint solve
- Then use Sherman-Morrison-Woodbury
- Relative to the prior uncertainty, the posterior uncertainty is reduced when observations are made.

Linearized Bayesian solution: Why low rank approximation?

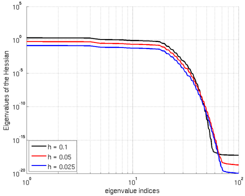
Compactness of the Hessian in inverse acoustic scattering

Theorem

Let $(1-n) \in C_0^{m,\alpha}$*, where* n *is the refractive index,* $m \in \mathbb{N} \cup \{0\}, \alpha \in (0,1)$*. The Hessian is a compact operator everywhere.*



Coupled FEM-BEM method



Eigenvalues of Gauss-Newton Hessian

Details in:

- T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part II: Inverse medium scattering of acoustic waves.* **Inverse Problems**, 28, 055002, 2012.
- T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part I: Inverse shape scattering of acoustic waves.* **Inverse Problems 2012 Highlights Collection**, 28, 055001, 2012.

Infinite dimensional Bayesian statistical inference

Linearized Bayesian solution: Low rank approximation

posterior covariance operator: A low rank approximation

$$
\begin{aligned}
\mathcal{C} &= \left(A^*\Gamma^{-1}A + \mathcal{C}_0^{-1}\right)^{-1} \\
&= \mathcal{C}_0^{1/2}\left(\mathcal{C}_0^{1/2}\boldsymbol{A}^*\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\,\mathcal{C}_0^{1/2} + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&\approx \mathcal{C}_0^{1/2}\left(\boldsymbol{V}_r\boldsymbol{\Lambda}_r\boldsymbol{V}_r^* + \boldsymbol{I}\right)^{-1}\mathcal{C}_0^{1/2} \\
&= \mathcal{C}_0 - \mathcal{C}_0^{1/2}\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^*\mathcal{C}_0^{1/2}
\end{aligned}
$$

- Low rank approximation only involves incremetal forward and incremental adjoint solve
- Then use Sherman-Morrison-Woodbury
- Relative to the prior uncertainty, the posterior uncertainty is reduced when observations are made.

Convergence for non-conforming hp-discretization

Theorem

Assume $\mathfrak{q}^e \in \left[H^{s_e}\left(\mathsf{D}^e\right)\right]^d, s_e \geq 3/2$ *with* $d = 6$ *for electromagnetic case and* $d = 12$ *for elastic–acoustic case. In addition, suppose* $\mathfrak{q}_d(0) = \Pi\mathfrak{q}(0)$*, and the mesh is affine and non-conforming. Then, the discontinuous Galerkin spectral element solution* $\mathfrak{q}_d$ *converges to the exact solution* $\mathfrak{q}$*, i.e., there exists a constant* C *that depends only on the angle condition of* D^e*,* s*, and the material constants* μ *and* ε *(*λ *and* μ *for elastic–acoustic case) such that*

$$
\begin{aligned}
\left\|\mathfrak{q}\left(t\right)-\mathfrak{q}_d\left(t\right)\right\|_{\mathcal{D}^{N_{el}},d} \leq & C\sum_e \frac{h_e^{\sigma_e}}{N_e^{s_e}}\left\|\mathfrak{q}\left(t\right)\right\|_{\left[H^{s_e}\left(\mathsf{D}^e\right)^d\right]} \\
& + C\sum_e t\frac{h_e^{\sigma_e-1/2}}{N_e^{s_e-1/2}}\max_{[0,t]}\left\|\mathfrak{q}\left(t\right)\right\|_{\left[H^{s_e}\left(\mathsf{D}^e\right)\right]^d},
\end{aligned}
$$

with $h_e = diam\left(\mathsf{D}^e\right)$*,* $\sigma_e = \min\left\{p_e+1, s_e\right\}$*, and* $\left\|\cdot\right\|_{H^s\left(\mathsf{D}^e\right)}$ *denoting the usual Sobolev norm*

Details in: T. Bui-Thanh and O. Ghattas, *Analysis of an hp-non-conforming discontinuous Galerkin spectral element method for wave propagations*, **SIAM Journal on Numerical Analysis**, 50(3), pp. 1801–1826, 2012.

Scalability of global seismic wave propagation on Jaguar

Strong scaling: 3rd order DG, 16,195,864 elements, 9.3 billion DOFs

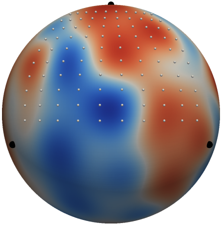
#cores	time [ms]	elem/core	efficiency [%]
1024	5423.86	15817	100.0
4096	1407.81	3955	96.3
8192	712.91	1978	95.1
16384	350.43	989	96.7
32768	211.86	495	80.0
65536	115.37	248	73.5
131072	57.27	124	74.0
262144	29.69	62	71.4

Strong scaling: 6th order DG, 170 million elements, 525 billion DOFs

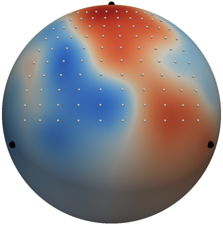
# cores	meshing time (s)	wave prop per step (s)	par eff wave	Tflops
32,640	6.32	12.76	1.00	25.6
65,280	6.78	6.30	1.01	52.2
130,560	17.76	3.12	1.02	105.5
223,752	$<$25	1.89	0.99	175.6

An example of global seismic inversion

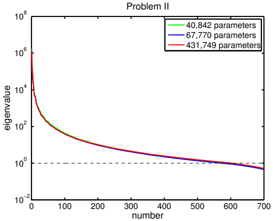
- inversion field: c_p in acoustic wave equation
- prior mean: PREM (radially symmetric model)
- "truth" model: S20RTS (Ritsema et al.), (laterally heterogeneous)
- Piecewise-trilinear on same mesh as forward/adjoint 3rd order dG fields
- dimensions: 1.07 million parameters, 630 million field unknowns
- Final time: $T = 1000s$ with 2400 time steps
- A single forward solve takes 1 minute on 64K Jaguar cores



"truth", sources (black)

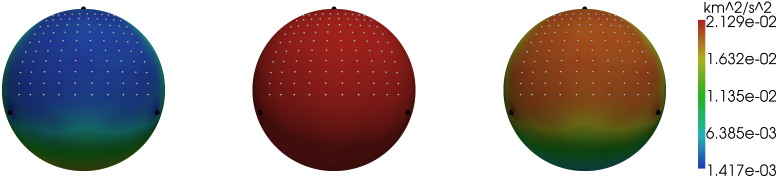


MAP, receivers (white)

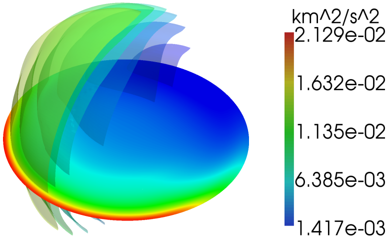


Hessian eigenvalues

Uncertainty quantification

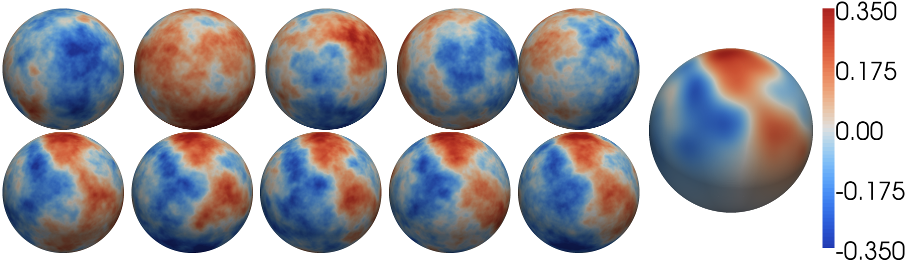


$$\mathfrak{C} \approx \mathfrak{C}_0 - \mathfrak{C}_0^{1/2} \boldsymbol{V}_r \boldsymbol{D}_r \boldsymbol{V}_r^* \mathfrak{C}_0^{1/2}$$



A slice through the equator and isosurfaces in the left hemisphere of variance reduction

Samples from prior and posterior distributions

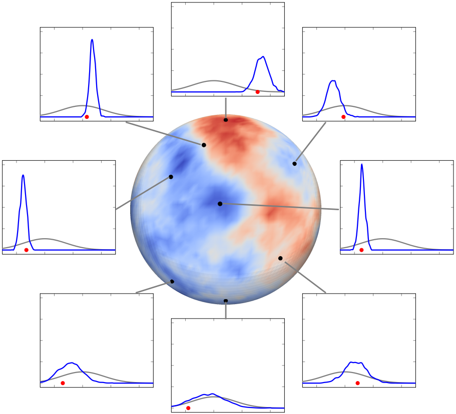


- **Top row**: samples from prior
- **Bottom row**: samples from posterior
- **Far right**: MAP estimate

Details in:

- Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., and Wilcox, L.C.,*Extreme-scale UQ for Bayesian inverse problems governed by PDEs*, ACM/IEEE Supercomputing SC12, **Gordon Bell Prize Finalist**, 2012.
- Bui-Thanh, T., Ghattas, O., Martin, J., and Stadler, G., *A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case*, **SIAM Journal on Scientific Computing**, Submitted, 2012.

MCMC Simulation for Seismic inversion



- prior distribution
- posterior distribution
- posterior sample

- Use Gaussian approximation as proposal
- 15,587 samples, acceptance rate 0.28
- 96 hours on 2048 cores

Discretization of infinite dimensional Bayesian inversion

Error analysis and uncertainty quantification for 2D inverse shape acoustic scattering

- Shape $r = \exp(\mathbf{u})$, where $\mathbf{u} \in C^{s,\alpha}[0, 2\pi]$, $s \geq 2$ and $0 \leq \alpha \leq 1$
- Discretize μ_0 using Karhunen-Loève truncation with m terms
- Discretize the forward equation using n-th order Nyström scheme

Theorem

$$d_{Hellinger}(\mu, \mu_{n,m}) \leq c\left(\frac{1}{(2n)^{s-1}} + \frac{\log m}{m^{s-1+\alpha}}\right),$$

$$\|\mathcal{E}_M\|_{L^2[0,2\pi]} \leq c\left(\frac{1}{(2n)^{s-1}} + \frac{\log m}{m^{s-1+\alpha}}\right),$$

$$\|\mathcal{E}_C\|_{L^2[0,2\pi]\otimes L^2[0,2\pi]} \leq c\left(\frac{1}{(2n)^{s-1}} + \frac{\log m}{m^{s-1+\alpha}}\right).$$

Details in: Bui-Thanh, T., and Ghattas, O., *An Analysis of Infinite Dimensional Bayesian Inverse Shape Acoustic Scattering and its Numerical Approximation*, **SIAM Journal on Uncertainty Quantification**, Submitted, 2012.

Conclusions

Conclusions: Reduce-Then-Sample

Inverse shape electromagnetic scattering

1. Statistical inversion via the Bayesian framework
2. Expensive forward solve
3. Monte Carlo sampling the posterior in high dimensions is too expensive

Approach and main results

- Hessian-based Piecewise Gaussian approximation to the posterior
- Automatically partion high dimensional parameter spaces without meshing
- Inverse solution comes with quantifiable uncertainty
- More than three order of magnitudes saving in time

Conclusions: Sample-Then-Reduce

Full wave form seismic inversion

1. Infinite dimensional Bayesian inference
2. Doubly infinite dimensional problem: both state and parameter live in infinite dimensional spaces
3. Very expensive forward solve even on supercomputers

Approach and Main results

- Discontinuous Galerkin for forward PDE
- Continuous FEM for prior with multigrid
- Exploit the ill-posedness and hence the compactness of the Hessian
- Able to solve statistical inverse problem with more than one million parameters with more than three orders of magnitude speedup
- Gaussian approximation seems to be good in this case
- Inverse solution comes with quantifiable uncertainty and more