A stochastic Newton method for statistical inverse problems, with application to inverse scattering

Omar Ghattas

Joint work with:

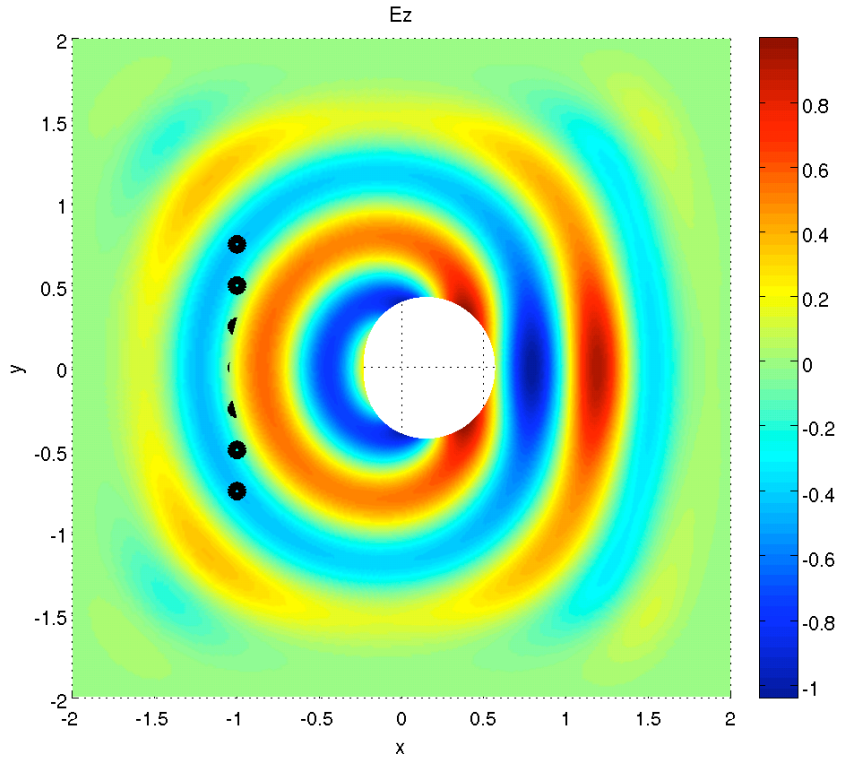
Tan Bui-Thanh Carsten Burstedde James Martin Lucas Wilcox

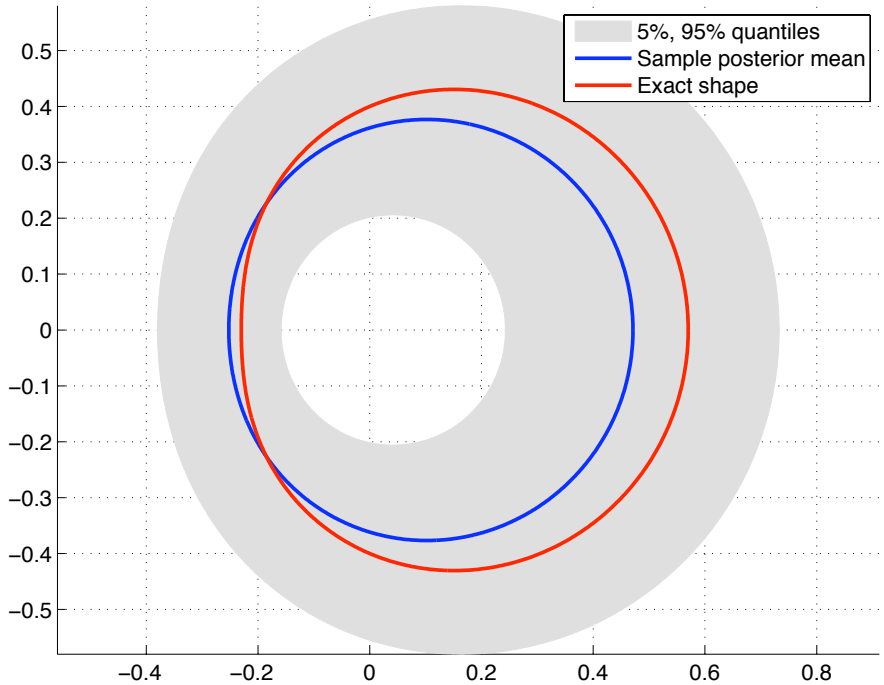
Institute for Computational Engineering and Sciences
Departments of Geological Sciences and Mechanical Engineering
The University of Texas at Austin

KAUST 22 March 2010

Bayesian inference for inverse scattering

Overall goals: Develop algorithms for solving statistical inverse problems that can scale to high dimensional probability spaces and expensive forward models, and tailor them to inverse shape and medium (acoustic, elastic, electromagnetic) scattering problems





Bayesian inference for inverse scattering, cont.

- Bayesian framework for statistical inverse problem: when data and/or model have uncertainties, solution of inverse problem expressed as a posterior probability density function
- Central challenge: for inverse problems characterized by high-dimensional parameter spaces, method of choice is to sample the posterior density using Markov chain Monte Carlo (MCMC)
- For inverse problems characterized by expensive forward simulations, contemporary MCMC methods become prohibitive
- Intractability of MCMC methods for large-scale statistical inverse problems can be traced to their black-box treatment of the parameter-to-observable map
- Goal: develop MCMC methods that exploit the structure of the parameter-to-observation map (including its derivatives), as has been done successfully in PDE-constrained optimization

Outline

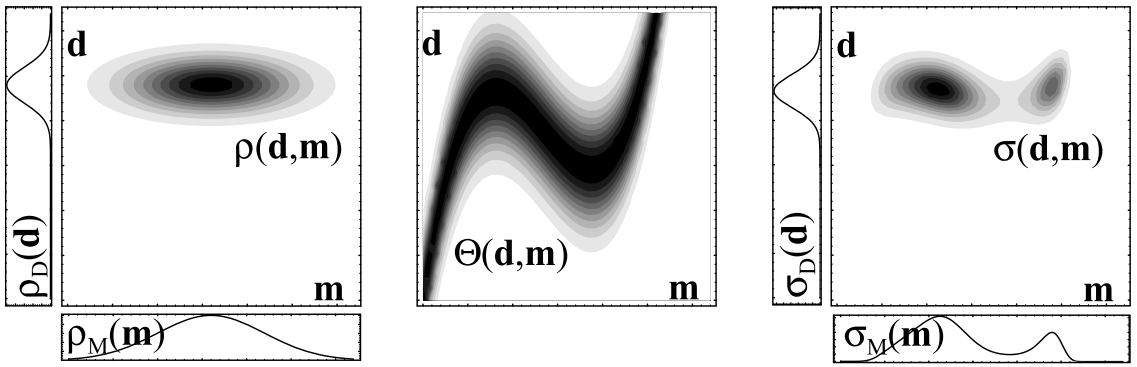
Bayesian formulation for statistical inverse problem

Given:

$$\begin{aligned} \pi_{\text{pr}}(\boldsymbol{x}) &:= \text{prior p.d.f. of model parameters } \boldsymbol{x} \\ \pi_{\text{obs}}(\boldsymbol{y}) &:= \text{prior p.d.f. of the observables } \boldsymbol{y} \\ \pi_{\text{model}}(\boldsymbol{y}|\boldsymbol{x}) &:= \text{conditional p.d.f. relating } \boldsymbol{y} \text{ and } \boldsymbol{x} \end{aligned}$$

Then *posterior p.d.f. of model parameters* is given by:

$$\begin{aligned} \pi_{\text{post}}(\boldsymbol{x}) &\overset{\text{def}}{=} \pi_{\text{post}}(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}) \\ &\propto \pi_{\text{pr}}(\boldsymbol{x}) \int_{\text{y}} \frac{\pi_{\text{obs}}(\boldsymbol{y})\pi_{\text{model}}(\boldsymbol{y}|\boldsymbol{x})}{\mu(\boldsymbol{y})} d\boldsymbol{y} \\ &\propto \pi_{\text{pr}}(\boldsymbol{x})\, \pi(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x}) \end{aligned}$$



From A. Tarantola, *Inverse Problem Theory*, SIAM, 2005

Gaussian additive noise

Given the parameter-to-observable map $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$, a common noise model is Gaussian additive noise:

$$\boldsymbol{y}_{\text{obs}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\text{noise}})$$

If the prior is taken as Gaussian with mean $\boldsymbol{x}_{\text{pr}}$ and covariance $\boldsymbol{\Gamma}_{\text{pr}}$, then the posterior can be written

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

Note that "most likely" point is given by

$$\begin{aligned} \boldsymbol{x}_{\text{MAP}} &\overset{\text{def}}{=} \arg \max_{\boldsymbol{x}} \ \pi_{\text{post}}(\boldsymbol{x}) \\ &= \arg \min_{\boldsymbol{x}} \ \tfrac{1}{2} \parallel \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}_{\text{obs}} \parallel^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} + \tfrac{1}{2} \parallel \boldsymbol{x} - \boldsymbol{x}_{\text{pr}} \parallel^2_{\boldsymbol{\Gamma}^{-1}_{\text{pr}}} \end{aligned}$$

This is an (appropriately weighted) deterministic inverse problem!

Gaussian additive noise, linear inverse problem

Suppose further the parameter-to-observable map is linear, i.e.

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}$$

Then the posterior can be written

$$\pi_{\text{post}}(\boldsymbol{x}) \propto \exp\left(-\tfrac{1}{2} \parallel \boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}_{\text{obs}} \parallel^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \tfrac{1}{2} \parallel \boldsymbol{x} - \boldsymbol{x}_{\text{pr}} \parallel^2_{\boldsymbol{\Gamma}^{-1}_{\text{pr}}}\right)$$

The posterior is then Gaussian with

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_{\text{MAP}}, \boldsymbol{\Gamma}_{\text{post}})$$

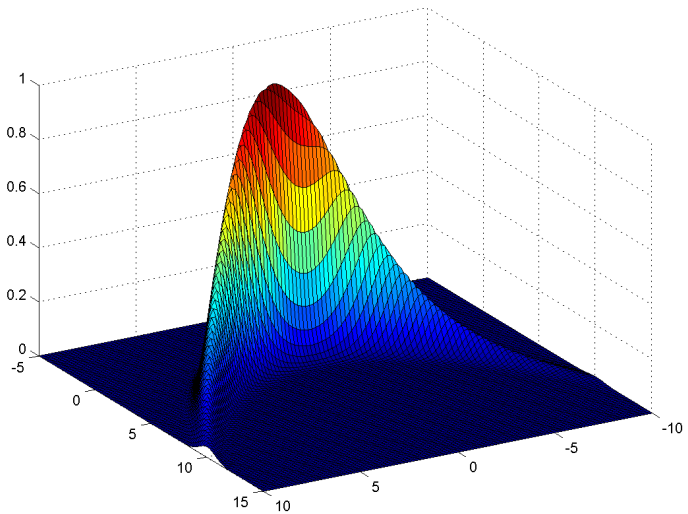
The covariance is the inverse Hessian of the negative log posterior:

$$\begin{aligned}\boldsymbol{\Gamma}^{-1}_{\text{post}} &= \boldsymbol{F}^T \boldsymbol{\Gamma}^{-1}_{\text{noise}} \boldsymbol{F} + \boldsymbol{\Gamma}^{-1}_{\text{pr}} \\ &= \nabla^2_{\boldsymbol{x}}(-\log \pi_{\text{post}})\end{aligned}$$

I.e., the covariance is given by the inverse Hessian of the regularized misfit function that is minimized by deterministic methods

Outline

MCMC sampling

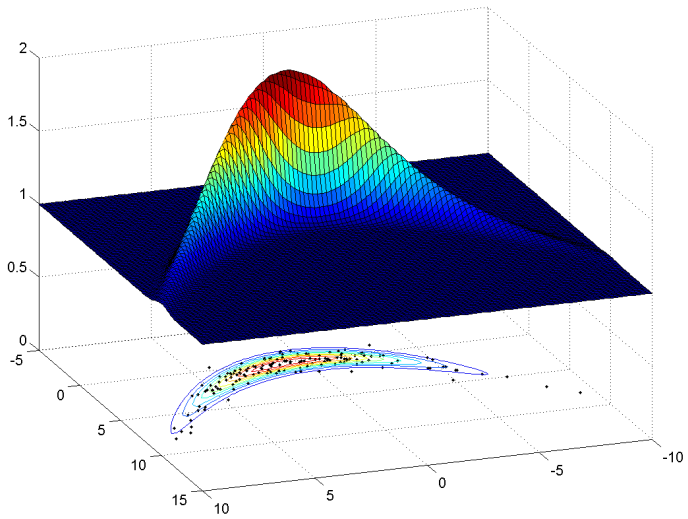


Example Probability Density

Given a probability density $\pi(\boldsymbol{x})$:

- How do we interrogate the distribution?
- Often high dimensional
- Computationally expensive

MCMC sampling



Sampled Probability Density

Given a probability density $\pi(\boldsymbol{x})$:

- How do we interrogate the distribution?
- Often high dimensional
- Computationally expensive

The MCMC Approach

- Replace $\pi(\boldsymbol{x})$ by a sample chain $\{\boldsymbol{x}_k\}$
- Compute using ergodic averages

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^n} f(x)\pi(dx) \approx \frac{1}{N}\sum_{j=1}^{N} f(x_k)$$

Metropolis-Hastings algorithm

1. $\boldsymbol{x}_k \leftarrow \boldsymbol{x}_0$
2. $k \leftarrow 0$
3. Choose a point $\boldsymbol{y}$ from the proposal density $q(\boldsymbol{x}_k, \ \cdot \ )$
4. $\alpha \leftarrow \min\left(1, \dfrac{\pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x}_k)}{\pi(\boldsymbol{x}_k)q(\boldsymbol{x}_k, \boldsymbol{y})}\right)$
5. If $\alpha > \text{rand}([0,1])$ Then

   Accept: $\boldsymbol{x}_{k+1} = \boldsymbol{y}$

   Otherwise

   Reject: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$

   End If
6. $k \leftarrow k + 1$
7. Repeat from step 3

Some proposal functions

The best proposal function is just the pdf itself:

- $q(\boldsymbol{x}_k, \boldsymbol{y}) = \pi(\boldsymbol{y})$
- $\alpha(\boldsymbol{x}_k, \boldsymbol{y}) = \min\left(1, \frac{\pi(\boldsymbol{y})\pi(\boldsymbol{x}_k)}{\pi(\boldsymbol{x}_k)\pi(\boldsymbol{y})}\right) \equiv 1$
- Would like to use an approximation $\tilde{\pi}(\boldsymbol{y})$

Gaussian random walks:

- $q(\boldsymbol{x}_k, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$
- Lots of freedom in choosing $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$
- Both can depend on $\boldsymbol{x}_k$

Many others:

- Hybrid Monte Carlo
- Gibbs sampling

Approaches to reducing the cost of MCMC

Basic difficulty: evaluating posterior density requires forward solve; philosophies to circumvent:

- *Reduce-then-sample:*
  - Reduced model of the forward problem
    - POD (e.g. Wang and Zabaras, Willcox et al., Patera et al.)
  - Reduced model of the outputs
    - PC (e.g Marzouk, Najm, and Rahn, Zabaras et al., Marzouk and Xiu)
    - Gaussian process model (e.g. O'Hagan and Kennedy, Higdon)
  - "Preconditioned" MCMC using reduced order models
    - Higdon, Lee, and Holloman
    - Christen and Fox
    - Efendiev, Hou, and Luo
    - Efendiev, Datta-Gupta, Ginting, Ma, and Mallick
- We will pursue: *Sample-then-reduce*
  - Exploit structure of $\pi_{\text{post}}(\boldsymbol{x})$, in particular local covariance related to Hessian of deterministic inverse problem
  - Build on fast algorithms from deterministic inverse solvers

Outline

Background: Langevin dynamics

Langevin dynamics

- Stochastic differential equation (continuous in time)
  - $\pi(\boldsymbol{x})$ is a stationary solution
  - $\Rightarrow$ Trajectories sample $\pi(\boldsymbol{x})$
- Uses derivative information of $\pi(\boldsymbol{x})$
- Can be preconditioned for better performance

Discrete Langevin dynamics

- Discretization with timestep Δt introduces bias
- Use as proposal distribution for Metropolis-Hastings MCMC (U. Grenander and M. Miller, 1994)

Langevin-MCMC used for PDE-based inverse problems by A. Stuart and Y. Efendiev

Preconditioned Langevin MCMC

Given the target density $\pi(\boldsymbol{x})$, the associated Langevin SDE is given by:

$$d\boldsymbol{X}_t = -\boldsymbol{A}\nabla_{\boldsymbol{x}}(-\log\pi)dt + \sqrt{2}\boldsymbol{A}^{1/2}d\boldsymbol{W}_t$$

Discretize with a timestep Δt to derive Langevin proposal:

$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k - \boldsymbol{A}\nabla_{\boldsymbol{x}}(-\log\pi)\Delta t + \sqrt{2\Delta t}\boldsymbol{A}^{1/2}\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$$

Notes:

- Preconditioner $\boldsymbol{A}$ must be symmetric positive definite
- Process is ergodic (convergence of time averages)
- $\boldsymbol{W}_t$ is i.i.d. vector of standard Brownian motions
- $\boldsymbol{W}_t$ has independent increments given by
  - $\boldsymbol{W}_{(t+\Delta t)} - \boldsymbol{W}_t \sim \mathcal{N}\big(\boldsymbol{0}, \Delta t\ \boldsymbol{I}\big)$

Stochastic Newton's method

Langevin MCMC proposal given by:

$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k - \boldsymbol{A}\nabla_{\boldsymbol{x}}(-\log\pi)\Delta t + \sqrt{2\Delta t}\boldsymbol{A}^{1/2}\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$$

Take $\boldsymbol{A}$ to be the inverse of the (local) Hessian (of the negative log posterior) and set $\Delta t = 1$:

$$\begin{aligned}\boldsymbol{A} \equiv \boldsymbol{H}(\boldsymbol{x})^{-1} &= \nabla^2_{\boldsymbol{x}}(-\log\pi(\boldsymbol{x}))^{-1} \\ &= \left(\boldsymbol{F}^T\boldsymbol{\Gamma}^{-1}_{\text{noise}}\boldsymbol{F} + \boldsymbol{\Gamma}^{-1}_{\text{pr}}\right)^{-1} \text{ (e.g. for Gaussian noise and prior)}\end{aligned}$$

Then we have the stochastic equivalent of Newton's method:

$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k - \boldsymbol{H}^{-1}\nabla_{\boldsymbol{x}}(-\log\pi) + \mathcal{N}(\boldsymbol{0},\boldsymbol{H}^{-1})$$

Stochastic Newton: Optimal sampling of Gaussians

When the target density $\pi(\boldsymbol{x})$ is Gaussian, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$:

$$-\log \pi(\boldsymbol{x}) \equiv \tfrac{1}{2} \parallel \boldsymbol{x} - \boldsymbol{\mu} \parallel_{\boldsymbol{\Gamma}^{-1}}$$

Apply Stochastic Newton:

$$\begin{aligned}
\boldsymbol{x}_{k+1}^{\text{prop}} &= \boldsymbol{x}_k - \boldsymbol{H}^{-1} \nabla_{\boldsymbol{x}} (-\log \pi) + \boldsymbol{H}^{-1/2} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \\
&= \boldsymbol{x}_k - \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1} (\boldsymbol{x}_k - \boldsymbol{\mu}) + \boldsymbol{\Gamma}^{1/2} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \\
&= \boldsymbol{\mu} + \boldsymbol{\Gamma}^{1/2} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \\
&= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})
\end{aligned}$$

Samples $\boldsymbol{x}_k$ act like *independent* draws from the true pdf

Deterministic vs. Stochastic Newton

Deterministic Newton:

- Given a cost function $-\log \pi(\boldsymbol{x})$
- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}^{-1} \nabla_{\boldsymbol{x}} (-\log \pi)$
- Minimizes local quadratic approximation at each step

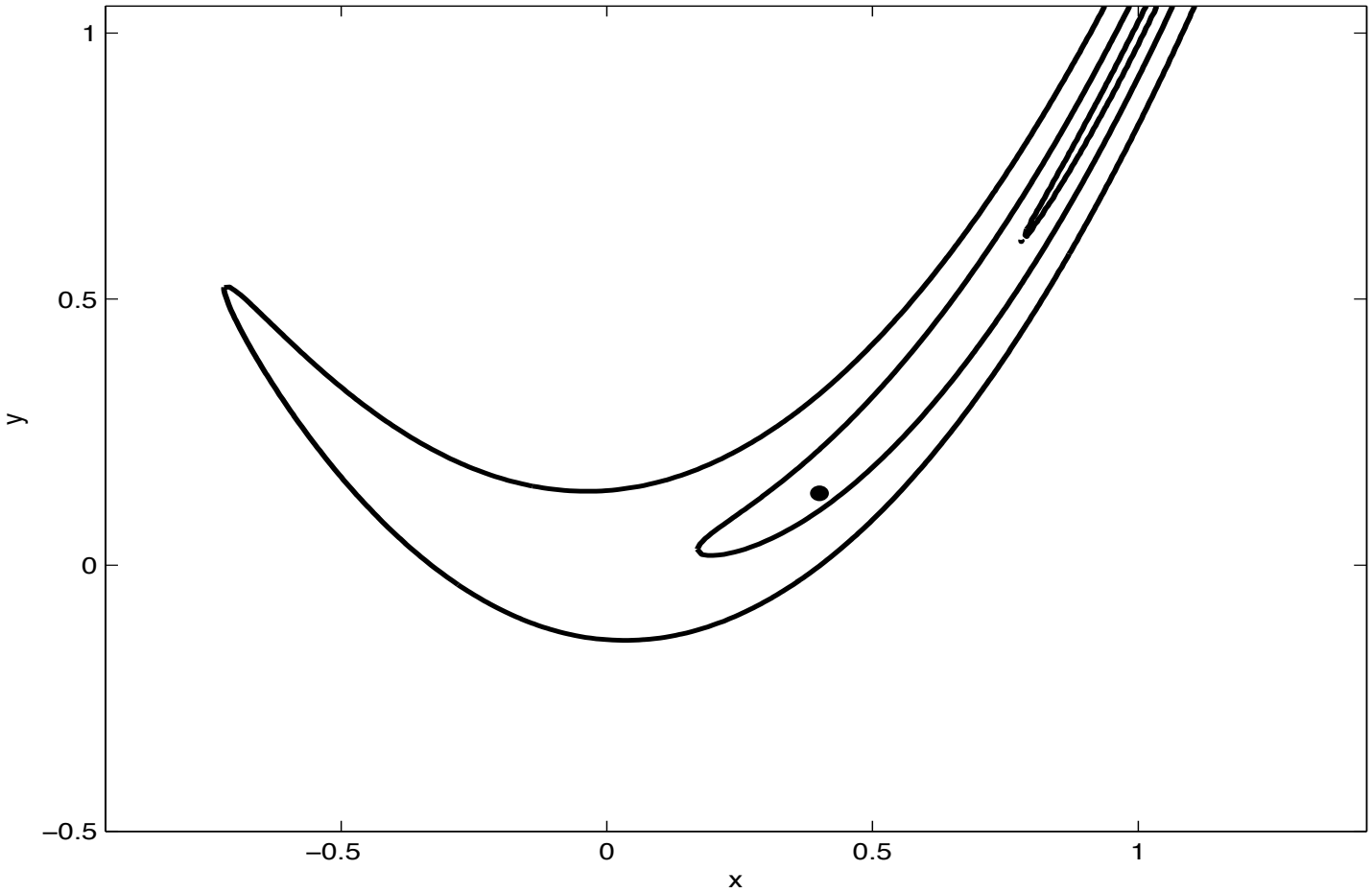
Stochastic Newton:

- Given a probability density $\pi(\boldsymbol{x})$
- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}^{-1} \nabla_{\boldsymbol{x}} (-\log \pi) + \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}^{-1})$
- Samples local Gaussian approximation at each step

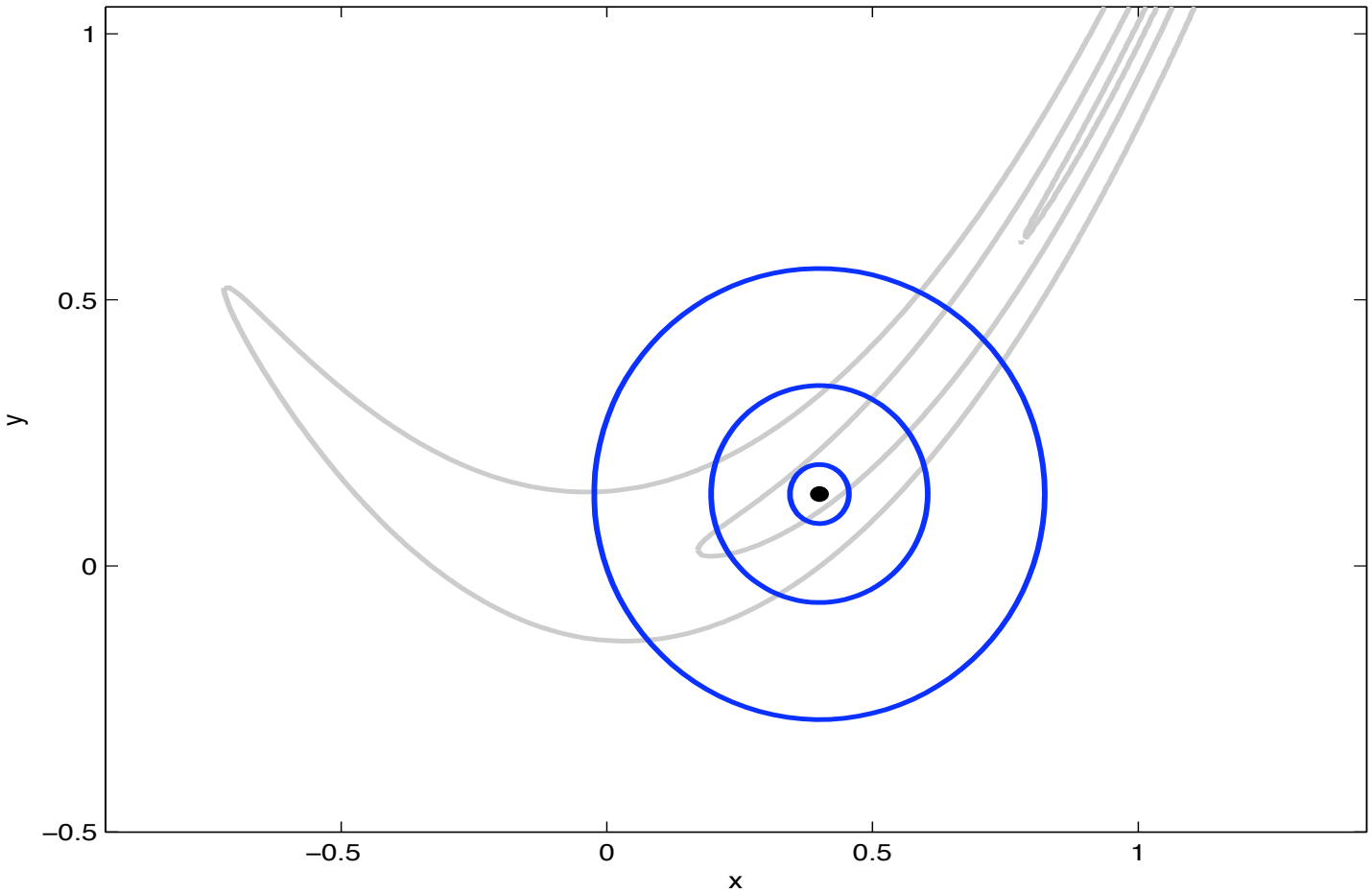
Unpreconditioned Langevin resembles steepest descent

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \Delta t \nabla_{\boldsymbol{x}} (-\log \pi) + \sqrt{2\Delta t}\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

Rosenbrock illustration

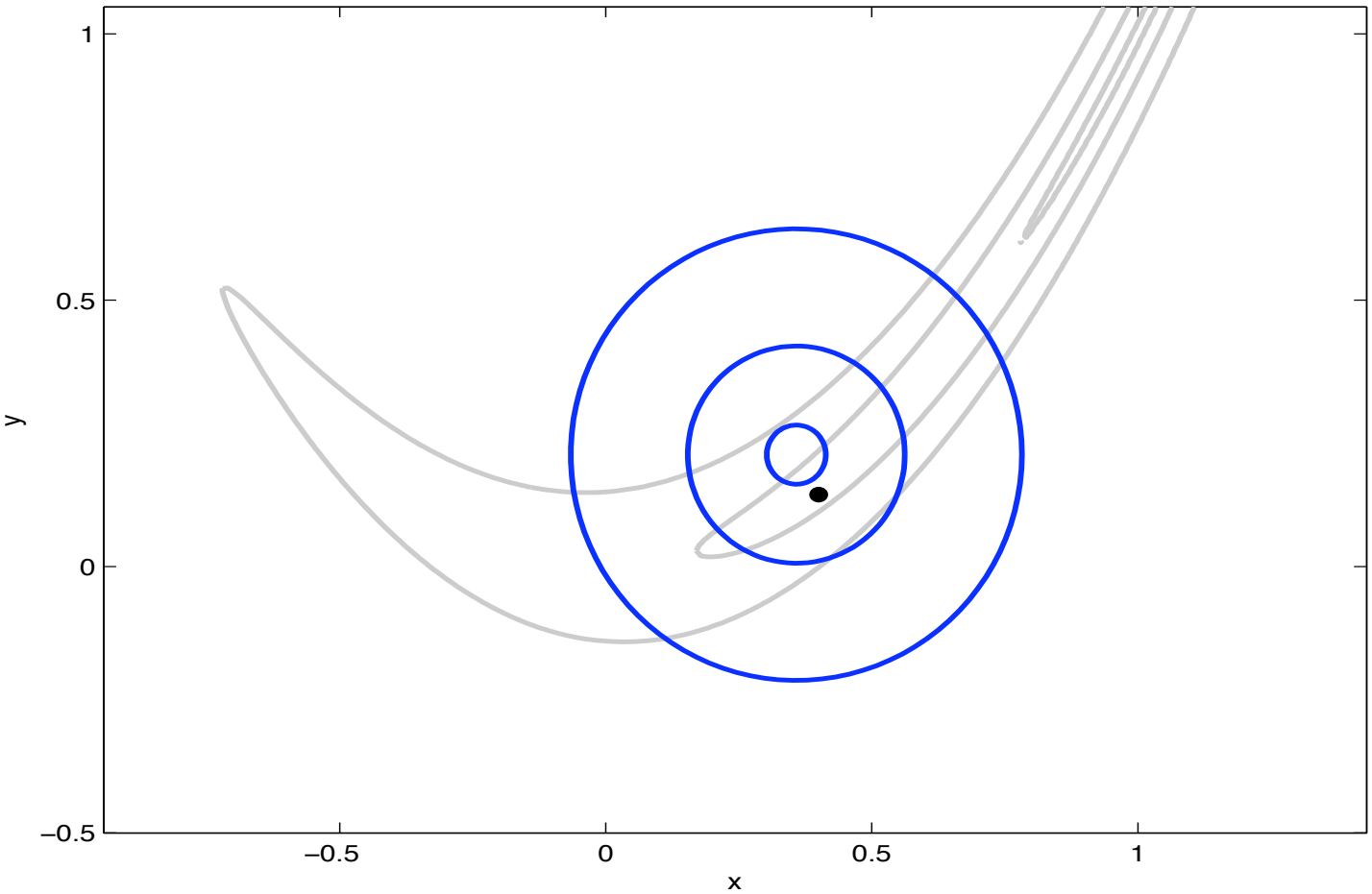


Rosenbrock illustration: Random walk



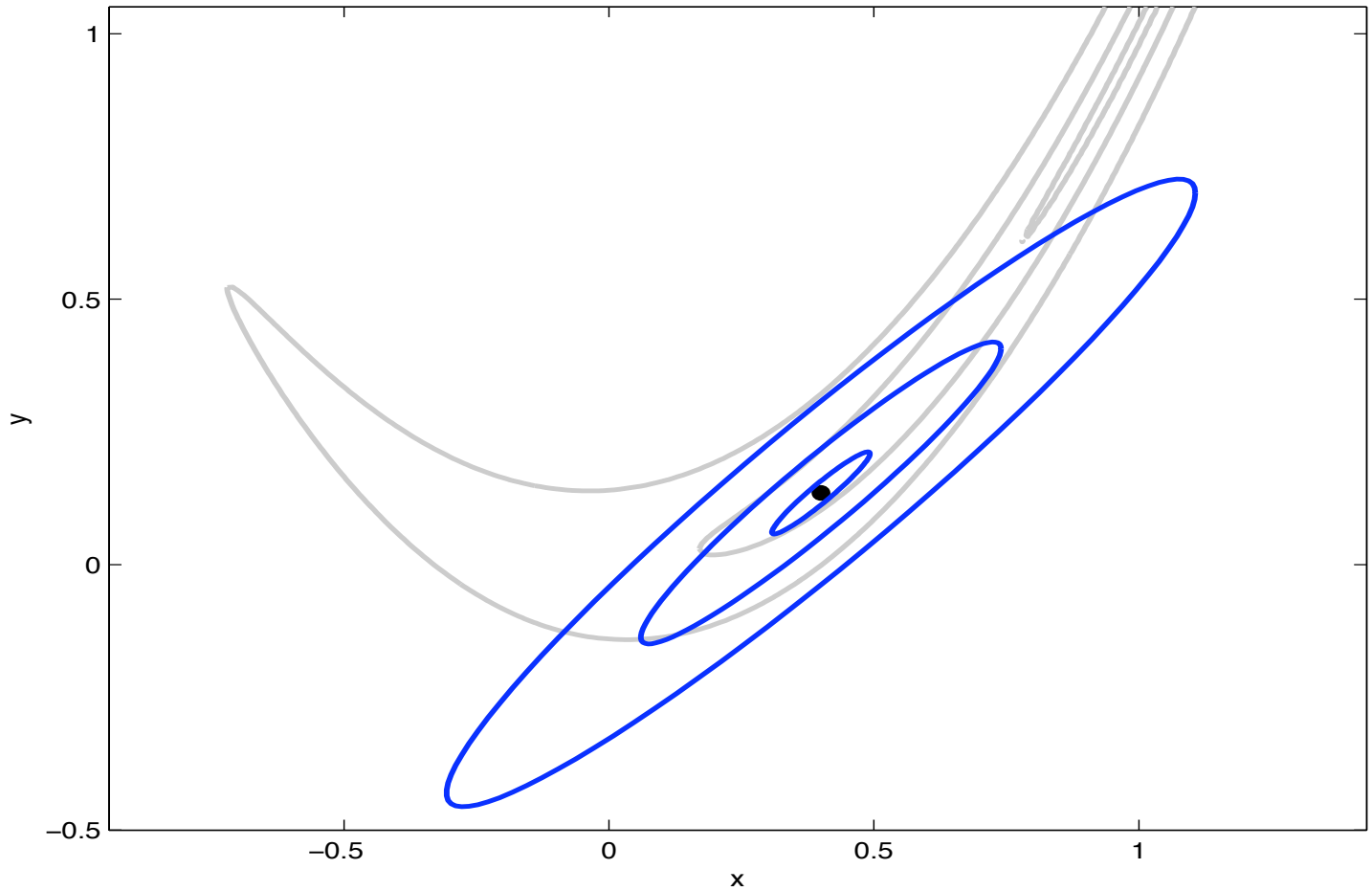
$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k + \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Rosenbrock illustration: Unpreconditioned Langevin



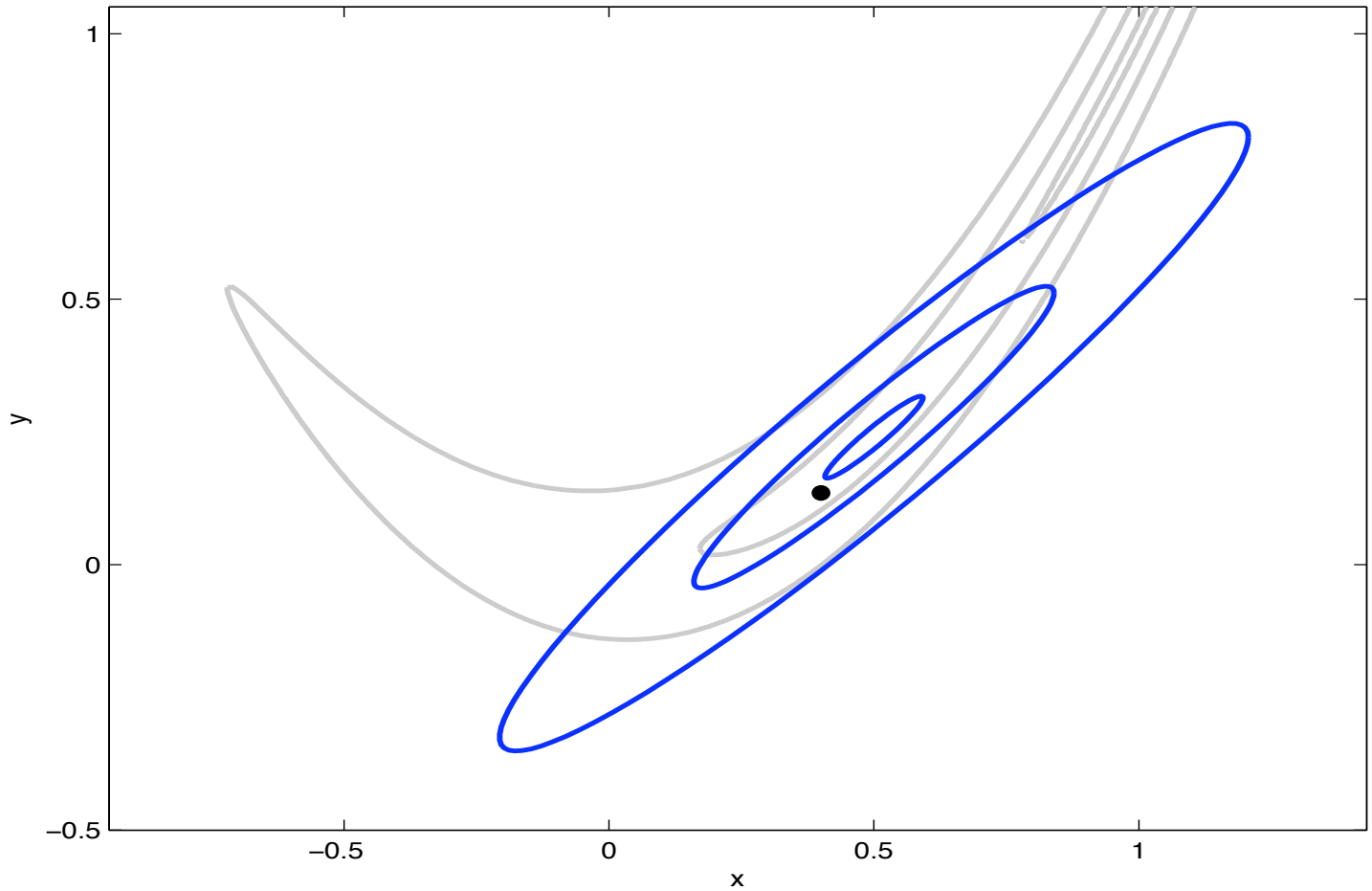
$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k - \Delta t \nabla_{\boldsymbol{x}}(-\log \pi) + \sqrt{2\Delta t}\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Rosenbrock illustration: Hessian-based Gaussian proposal



$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k + \mathcal{N}(\mathbf{0}, \boldsymbol{H}^{-1})$$

Rosenbrock illustration: Hessian-preconditioned Langevin



$$\boldsymbol{x}_{k+1}^{\text{prop}} = \boldsymbol{x}_k - \boldsymbol{H}^{-1}\nabla_{\boldsymbol{x}}(-\log\pi) + \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}^{-1})$$

Stochastic Newton: Large-scale issues

At each MCMC step we need to

- solve systems of form $\boldsymbol{H}\boldsymbol{v} = \boldsymbol{b}$
- evaluate matvecs of form $\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{w}$

Key idea: never form $\boldsymbol{H}$; instead:

- recognize that $\boldsymbol{H}$ is sum of data misfit term, which is often equivalent to a compact operator, and (the inverse of) a smoothing prior, which is often equivalent to a differential operator:

$$\boldsymbol{F}^T \boldsymbol{\Gamma}_{\text{noise}}^{-1} \boldsymbol{F} + \boldsymbol{\Gamma}_{\text{pr}}^{-1}$$

- develop fast algorithms for low rank (in particular, truncated spectral decomposition) approximation of data misfit operator; often require constant number of forward/adjoint solves, independent of problem size
- combine with Sherman-Morrison-Woodbury to invert/factor (requires constant number of forward/adjoint solves)
- construct fast (multilevel) preconditioners for Hessian

Outline

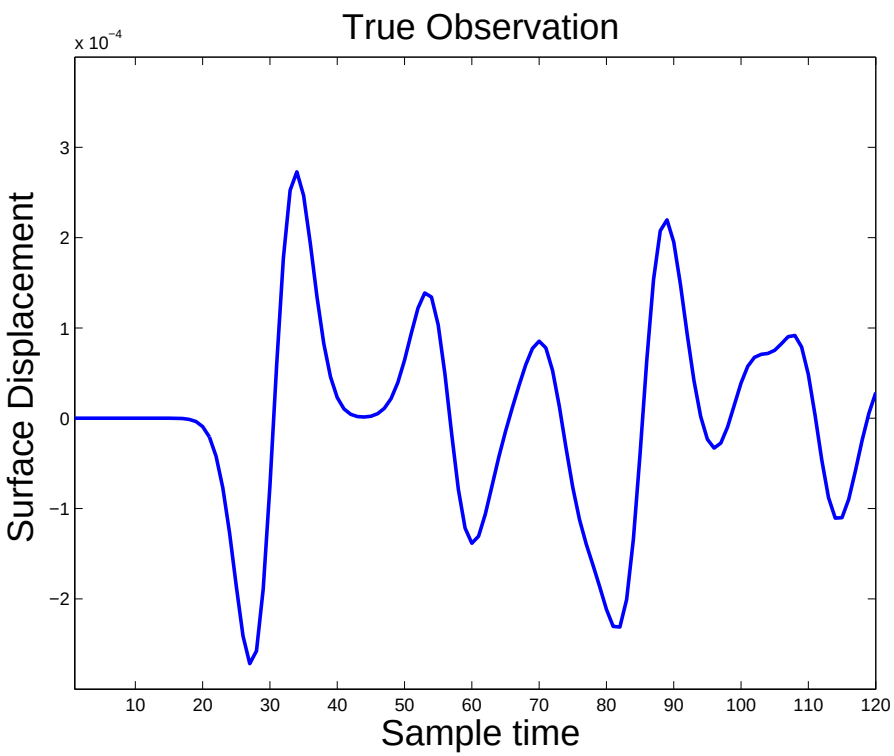
1D inverse elastic medium scattering

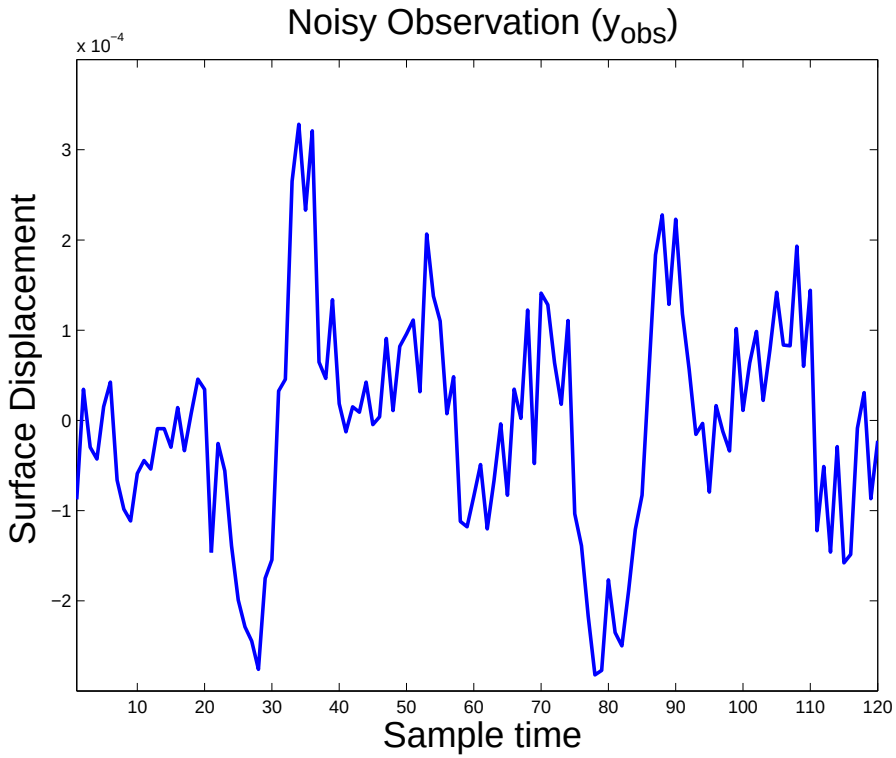
Forward problem: Given elastic modulus $a(x)$, solve 1D wave equation for Ricker wavelet source $g(t)$ to obtain observed waveform $y(0,t)$ at top surface:

$$\begin{aligned}
\rho \frac{\partial^2 y}{\partial t^2} - \frac{\partial}{\partial x}\left(a(x)\frac{\partial y}{\partial x}\right) &= \delta(x-0)g(t) \\
a \left.\frac{\partial y}{\partial x}\right|_{x=0} &= 0 \\
\sqrt{\rho a} \left.\frac{\partial y}{\partial t}\right|_{x=1} &= -a \left.\frac{\partial y}{\partial x}\right|_{x=1} \\
y|_{t=0} &= 0 \\
\dot{y}|_{t=0} &= 0
\end{aligned}$$

Inverse problem: Given observed noisy waveform and layered parametrization of elastic medium modulus $a(x)$ with associated prior, recover layer moduli and associated uncertainty

Sample noisy observations





Prior and likelihood distributions

Gaussian smoothness prior between layers:

- Covariance matrix $\mathbf{\Gamma}$ between layers i and j:

$$\Gamma_{ij} = \theta_1 \exp\left(\frac{-(x_i - x_j)^2}{2\theta_2^2}\right)$$

- Prior mean is 5 for all layers

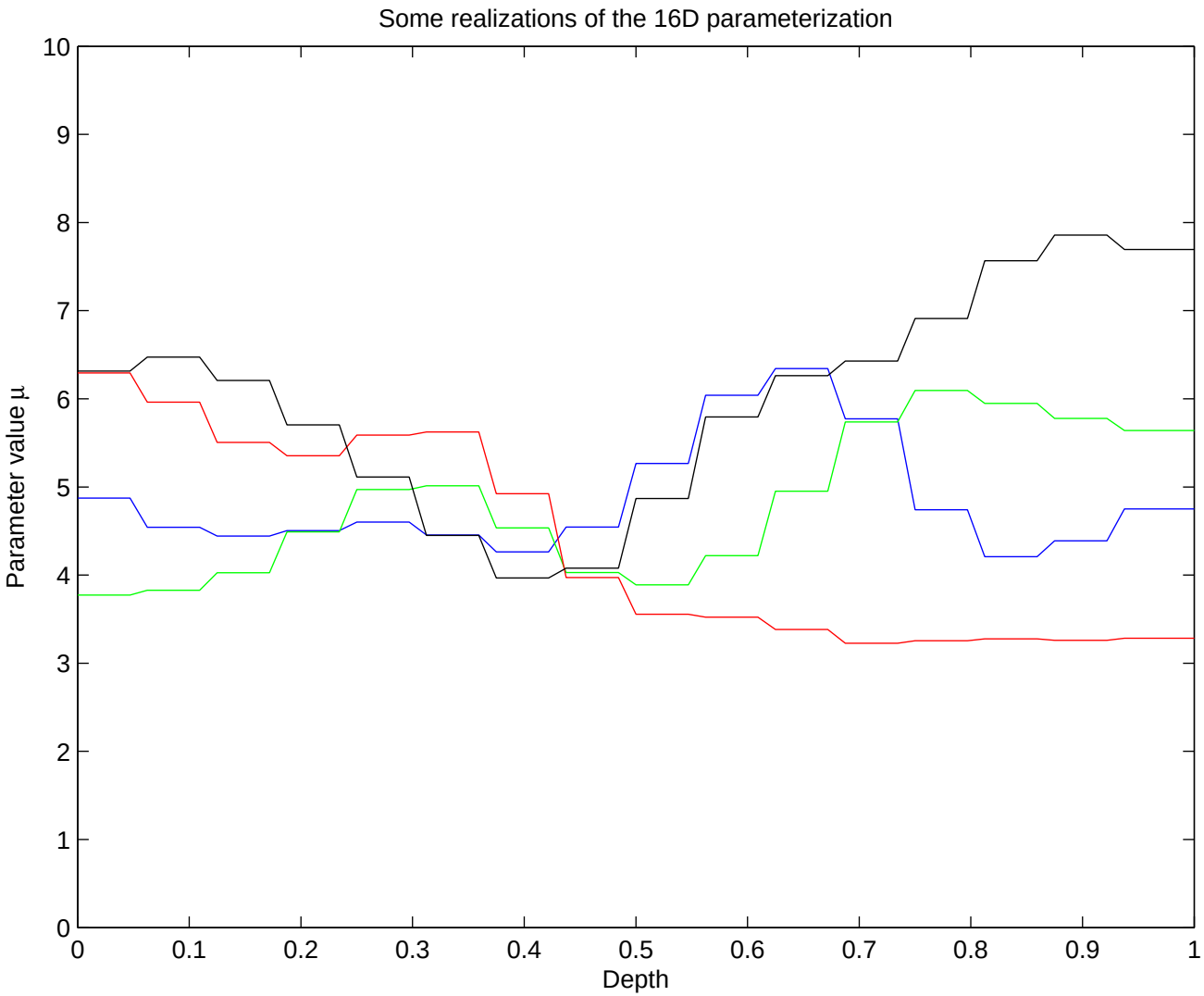
Gaussian likelihood function:

$$\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\boldsymbol{a}) = \exp\left(-\tfrac{1}{2}(\boldsymbol{f}(\boldsymbol{a}) - \boldsymbol{y}_{\text{obs}})^T \mathbf{\Gamma}_{\text{noise}}^{-1}(\boldsymbol{f}(\boldsymbol{a}) - \boldsymbol{y}_{\text{obs}})\right)$$

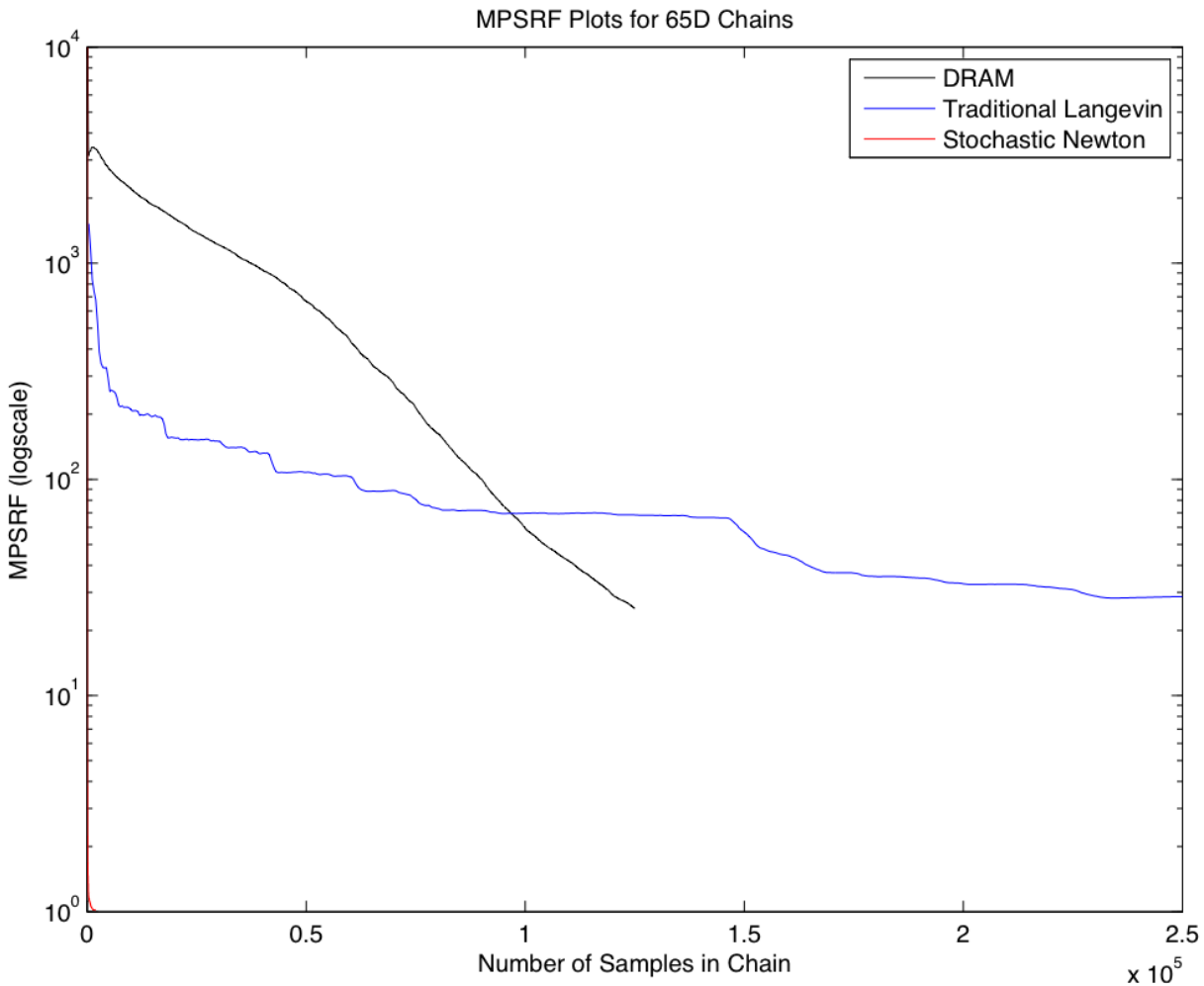
We wish to sample the posterior distribution:

$$\pi_{\text{post}}(\boldsymbol{a}|\boldsymbol{y}_{\text{obs}}) \propto \pi_{\text{pr}}(\boldsymbol{a})\pi_{\text{like}}(\boldsymbol{y}_{\text{obs}}|\boldsymbol{a})$$

Some realizations of 16-layer prior



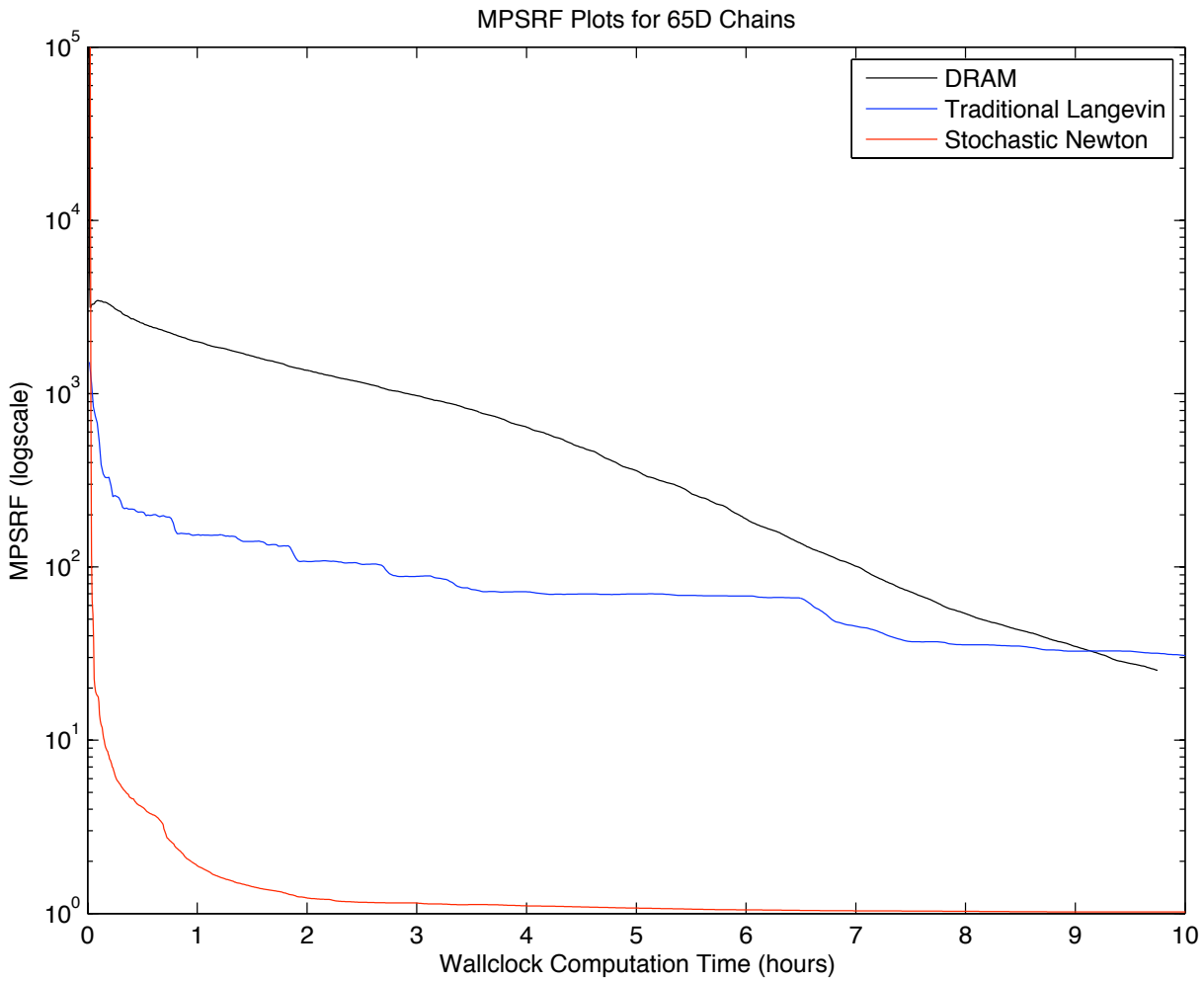
Convergence comparison for 65-layer problem



Multivariate potential scale reduction factor (MPSRF) convergence statistic for 65-layer problem

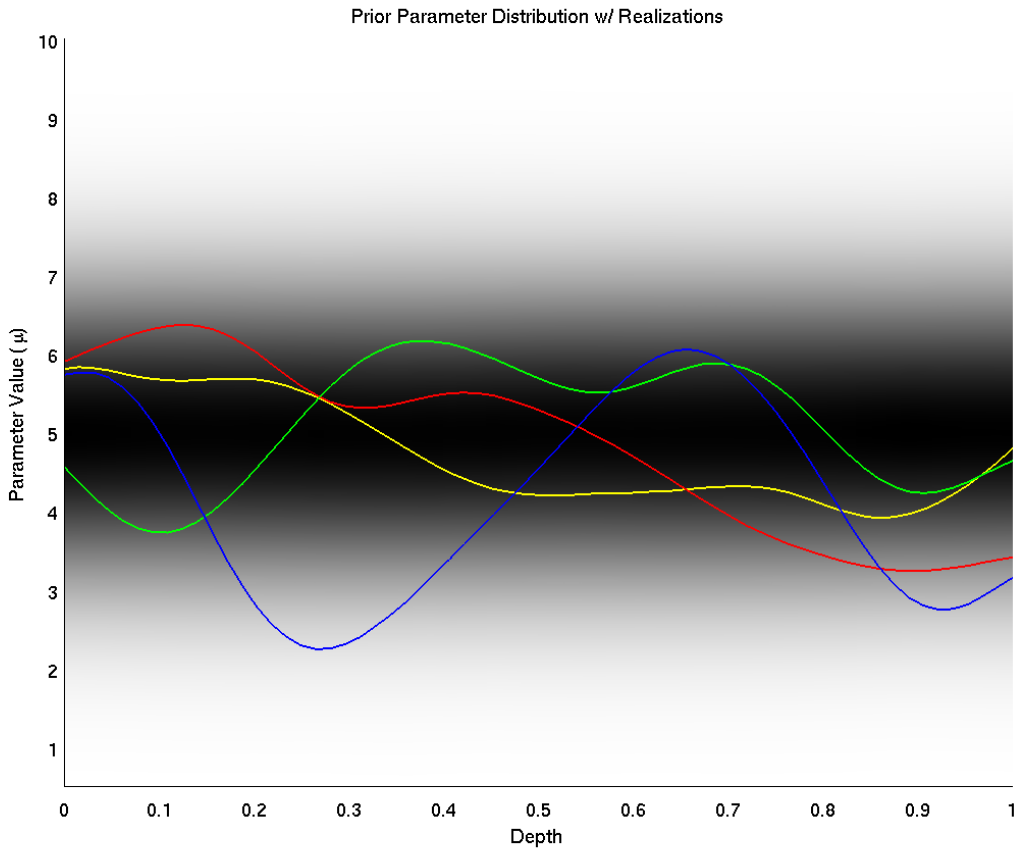
unpreconditioned Langevin vs. stochastic Newton vs. Adaptive Metropolis

Rescaled 65-layer MPSRF convergence



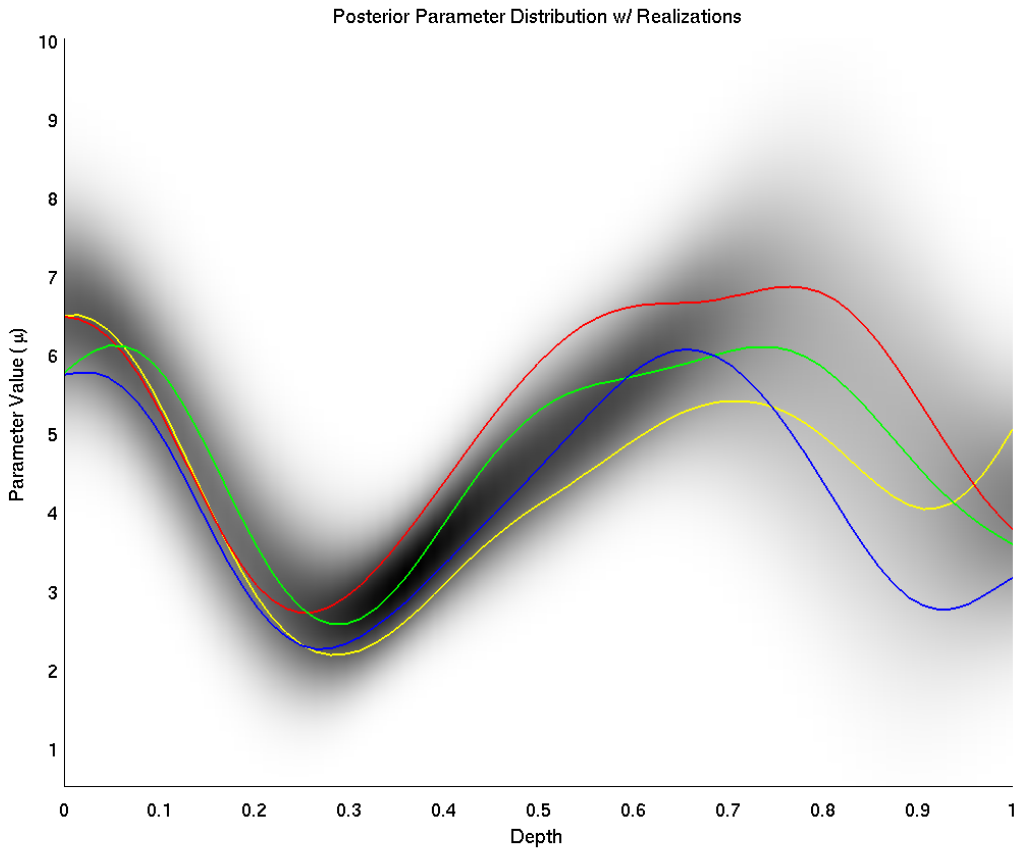
MPSRF statistic for 65-layer problem as function of wall clock time (dense implementation – not recommended!)

65-layer prior



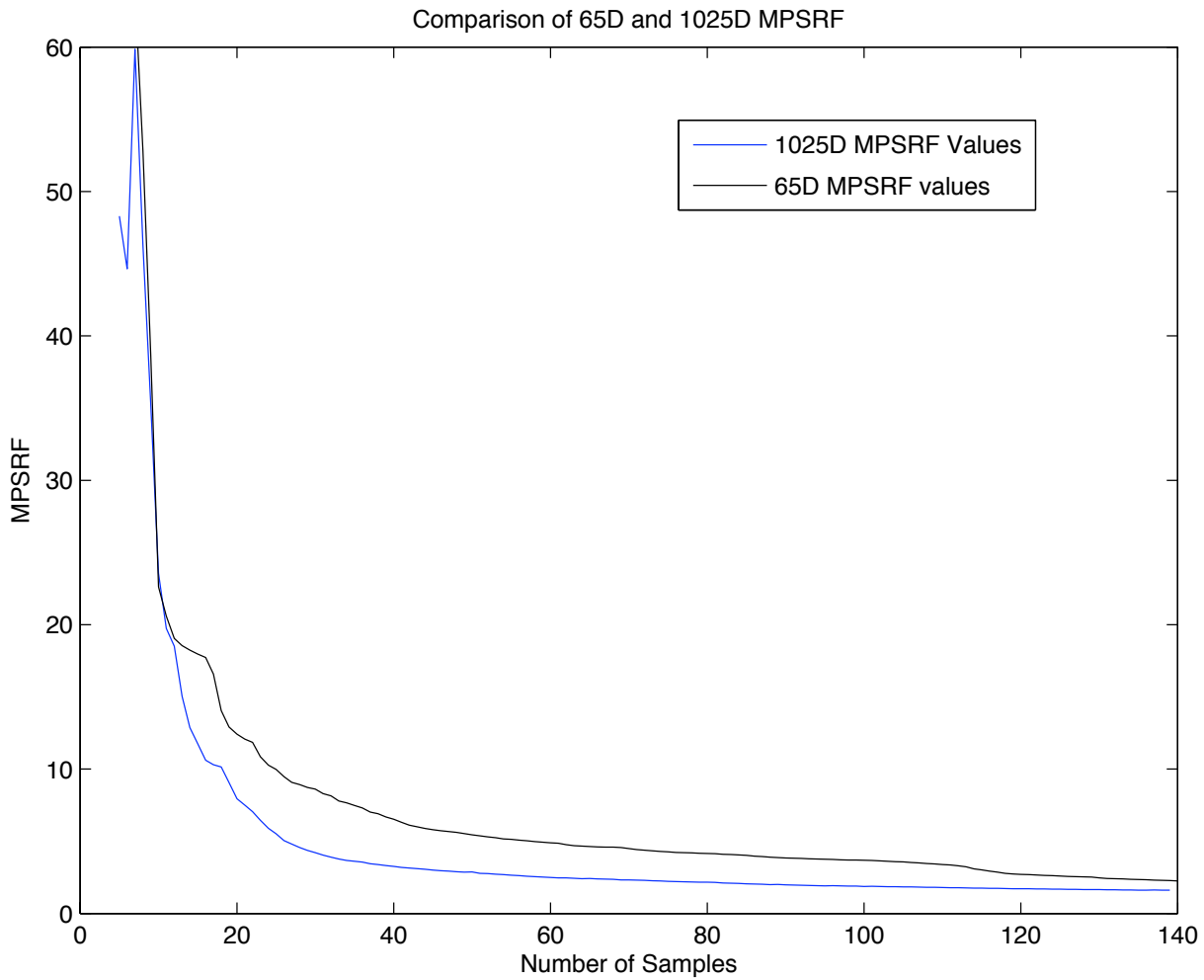
Density plot of marginal pdfs of prior of elastic moduli of 65 layers
Blue curve is "truth" modulus used to synthesize observations
Other colors are draws from prior

65-layer posterior



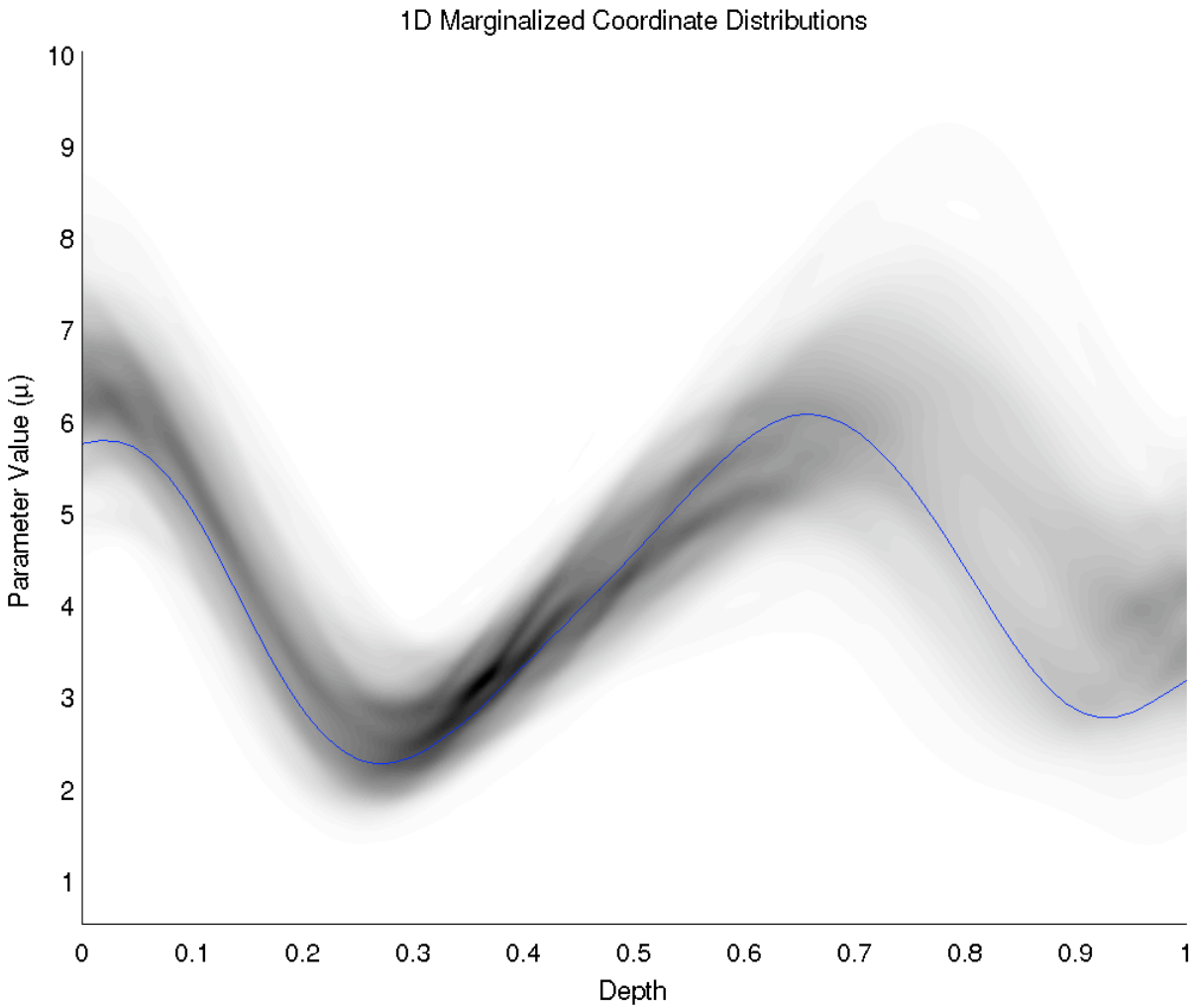
Density plot of marginal pdfs of posterior of elastic moduli of 65 layers
Blue curve is "truth" modulus used to synthesize observations
Other colors are draws from posterior

MPSRF convergence statistic for 1025-layer problem



MPSRF statistic for 1025-layer problem compared with 65-layer (1025-layer results based on fast low-rank implementation)

1025 layer posterior



Density plot of marginal pdfs of posterior of elastic moduli of 1025 layers
Blue curve is "truth" modulus used to synthesize observations

Outline

Statistical inverse electromagnetic obstacle scattering

Maxwell's equations:

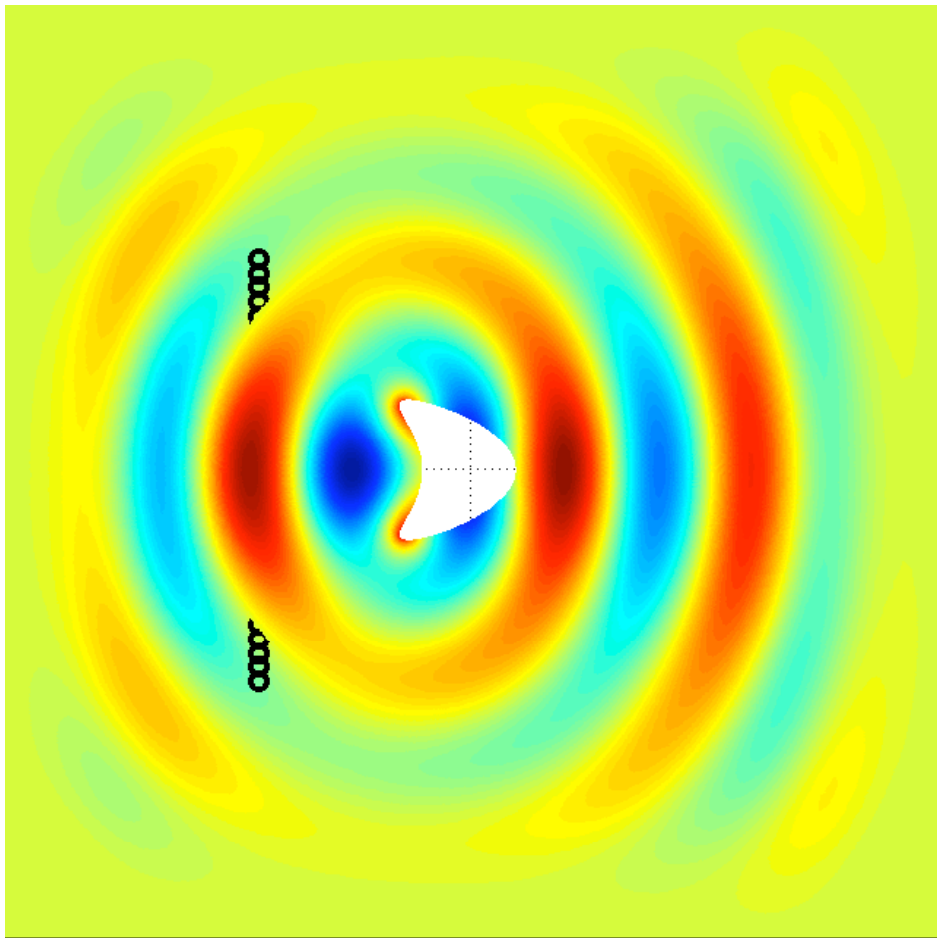
$$\nabla \times \boldsymbol{E} = -\mu \frac{\partial \boldsymbol{H}}{\partial t}$$

$$\nabla \times \boldsymbol{H} = \epsilon \frac{\partial \boldsymbol{E}}{\partial t}$$

$$\nabla \cdot \boldsymbol{E} = 0$$

$$\nabla \cdot \boldsymbol{H} = 0$$

- $\mathbf{E}$ . . . Electric field
- $\mathbf{H}$ . . . Magnetic field
- μ . . . permeability
- ϵ . . . permittivity



Scattered electric field E_z

Problem setup

- Forward code based on C version of Matlab code from J. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods*
- Extended to include adjoint-based shape gradient and shape Hessian
- Discontinuous Galerkin 3rd-order spectral elements in space
- Fourth-order, five-stage explicit Runge Kutta scheme in time
- Prior favors shape with small area

$$\pi_{\text{pr}} = \exp\left(-\frac{\beta}{2}\int_0^{2\pi} r^2 \, d\theta\right), \quad \beta = 0.1$$

- Shape is parametrized by 6 cosine modes

$$r(\theta) = \sum_{i=0}^{5} a_i \cos(i\theta)$$

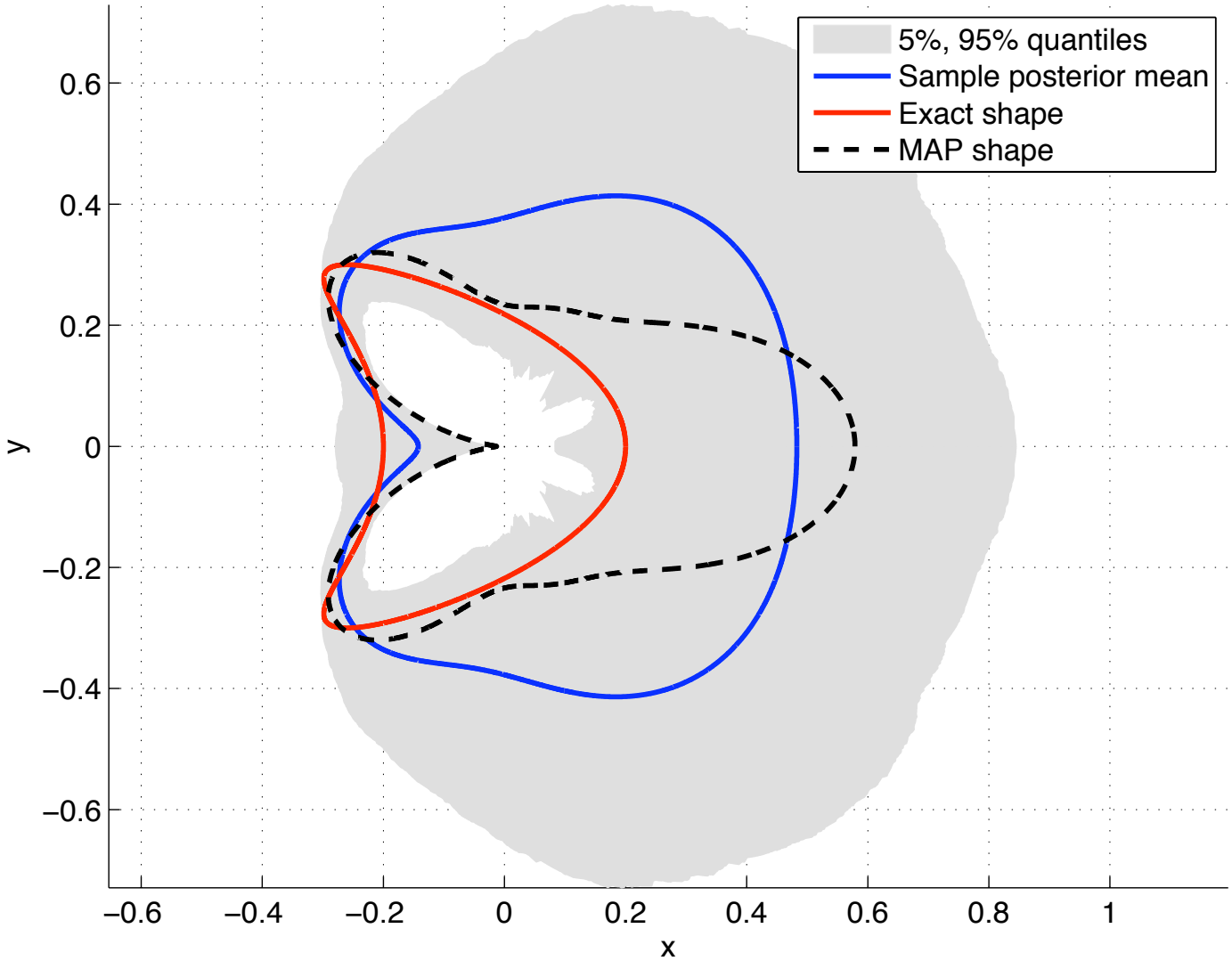
Problem setup, cont.

- Computational domain $\Omega = \{(x, y) : -1 \leq x, y \leq 1\}$
- PML domain $\Omega_{PML} = \{(x, y) : 1 \leq |x|, |y| \leq 2\}$
- Kite shape to generate synthetic observations:

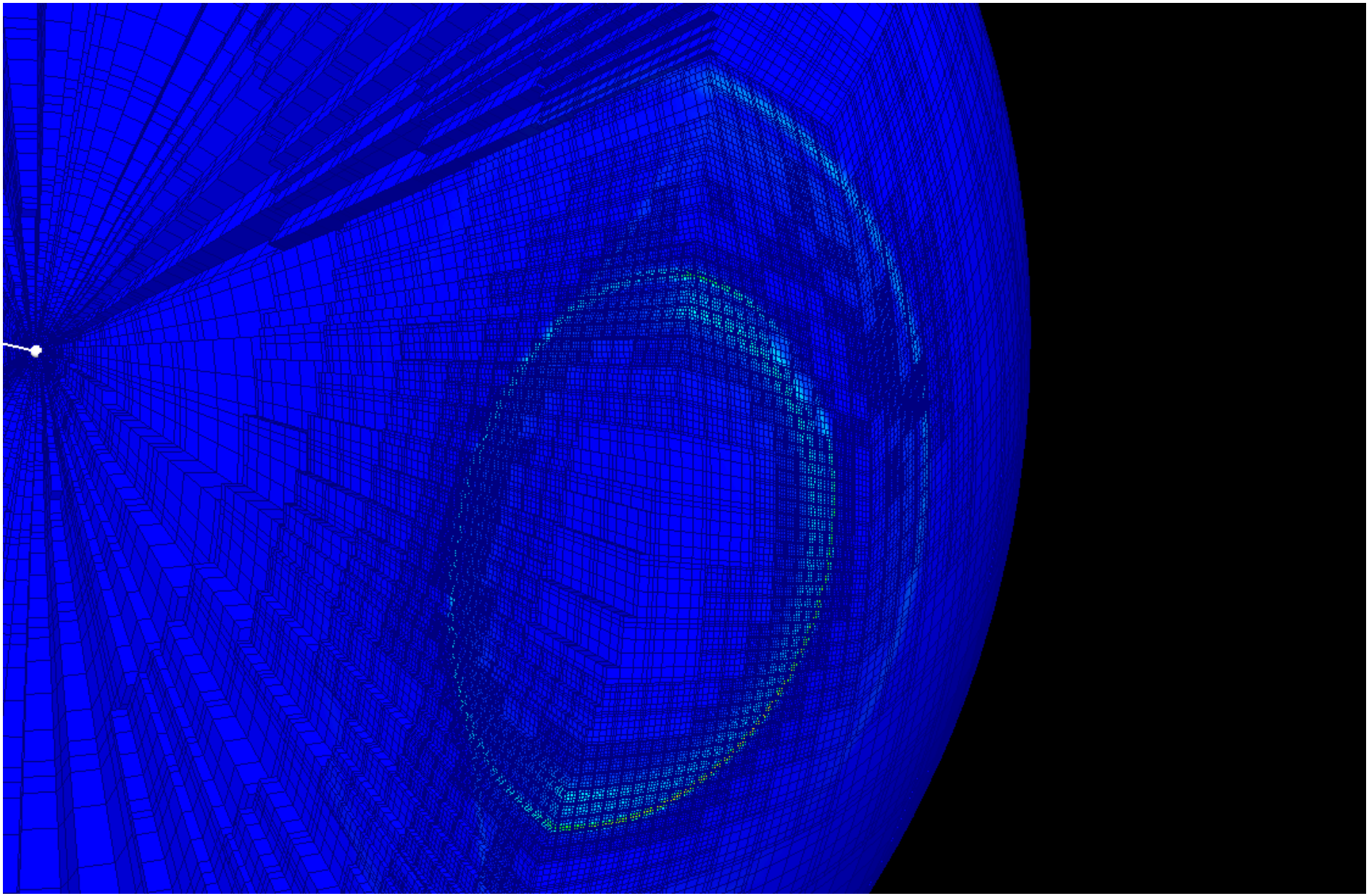
$$x = 0.2\left[\cos(\theta) + 0.65(\cos(2\theta) - 1)\right], \quad y = 0.3\sin(\theta)$$

- Incident wave $E_z^I = \cos\left(8(t - x)\right), H_x = 0, H_y = 0$ from left
- 31 observation points: $x = -0.9, y = linspace\left(-0.9, 0.9, 31\right)$
- E_z, H_x, H_y are observed in $0 \leq t \leq \pi$ with 5% Gaussian noise
- $\Delta t = 10^{-3}$ with 3324 time steps
- Mesh size $h_{\text{min}} = 0.05$ ($\sim$135,000 DOF)

Results using stochastic Newton for 6-parameter problem



3D Inverse elastic-acoustic wave propagation



Conclusions

- Target: PDE-based Bayesian inverse problems
- Stochastic Newton (inverse-Hessian-preconditioned Langevin MCMC)
  - motivated by connection to deterministic Newton method
  - exactly samples a Gaussian posterior
  - dense implementation shows vast improvement over black-box MCMC (adaptive Metropolis)
  - low rank implementation able to solve 1025-dimension inverse scattering problem
- Current work aimed at capitalizing on advances in deterministic PDE-based optimization and inverse methods to improve stochastic Newton
  - inexact Newton (Steihaug-Eisenstat-Walker ideas)
  - trust region methods
  - exploit "compact + differential" structure of Hessians (e.g. low rank approximations, Fredholm-multigrid type preconditioners)
- Exploiting deterministic PDE inverse problem structure should play an important role in scaling MCMC to high dimensions and expensive forward problems