Towards Scalable Methods for Large-scale Statistical Inverse Problems

Omar Ghattas

joint work with:

Carsten Burstedde, Pearl Flath, James Martin, Lucas Wilcox

other collaborators:

Youssef Marzouk, Bart van Bloeman Waanders, Karen Willcox

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

October 28, 2008

Motivation

Key computational kernels for uncertainty quantification

- estimate model parameters and their uncertainty from data (statistical inverse problem)
- propagate parameter uncertainty through model to predict quantities of interest and their uncertainty (forward uncertainty propagation problem)

Challenge: framework is often intractable for

- high-dimensional input parameter spaces
- expensive forward problems

Focus of this talk: How structure-exploiting methods can help overcome the curse of dimensionality for the statistical inverse problem

Outline

1. Lessons from deterministic inverse problems
2. Bayesian framework for statistical inverse problems
3. MCMC sampling
4. Langevin methods and stochastic Newton
5. Results for an inverse wave propagation problem

Outline

1. Lessons from deterministic inverse problems
2. Bayesian framework for statistical inverse problems
3. MCMC sampling
4. Langevin methods and stochastic Newton
5. Results for an inverse wave propagation problem

A model deterministic linear inverse problem: State estimation for atmospheric transport

$$\min_{u,q} \sum_j \int_\Omega \int_0^T (u - u^*)^2 \delta(\boldsymbol{x} - \boldsymbol{x_j})\, d\boldsymbol{x}\, dt + \frac{\beta}{2} \int_\Omega q^2\, d\boldsymbol{x}$$

$$\begin{aligned} u_t - k\Delta u + \boldsymbol{v} \cdot \nabla u &= 0 \text{ in } \Omega \times (0, T) \\ u &= q \text{ in } \Omega \times \{t = 0\} \\ k\nabla u \cdot \boldsymbol{n} &= 0 \text{ in } \Gamma_N \times (0, T) \\ u &= 0 \text{ on } \Gamma_D \times (0, T) \end{aligned}$$

u	contaminant concentration	q	initial condition
$\boldsymbol{v}$	wind velocity	k	diffusion coefficient
T	length of time window	β	regularization constant
$\boldsymbol{x}_j$	jth sensor location	u^*	observed concentration

Inverse atmospheric transport: Optimality conditions

State equation:

$$\begin{aligned} u_t - k\Delta u + \boldsymbol{v}\cdot\nabla u &= 0 \text{ in } \Omega\times(0,T) \\ u &= q \text{ in } \Omega\times\{t=0\} \\ k\nabla u\cdot\boldsymbol{n} &= 0 \text{ on } \Gamma_N\times(0,T) \\ u &= 0 \text{ on } \Gamma_D\times(0,T) \end{aligned}$$

Adjoint equation (for adjoint concentration p):

$$\begin{aligned} -p_t - k\Delta p - \nabla\cdot(p\boldsymbol{v}) &= -\sum_j (u-u^*)\delta(\boldsymbol{x}-\boldsymbol{x_j}) \text{ in } \Omega\times(0,T) \\ p &= 0 \text{ in } \Omega\times\{t=T\} \\ (k\nabla p + p\boldsymbol{v})\cdot\boldsymbol{n} &= 0 \text{ on } \Gamma_N\times(0,T) \\ p &= 0 \text{ on } \Gamma_D\times(0,T) \end{aligned}$$

Control equation:

$$-\beta q - p|_{t=0} = 0 \text{ in } \Omega$$

Inverse atmospheric transport: Construction of Hessian

Discretized optimality conditions:

$$\begin{pmatrix} B^TB & 0 & A^T \\ 0 & \beta I & -T^T \\ A & -T & 0 \end{pmatrix} \begin{pmatrix} u \\ q \\ p \end{pmatrix} = \begin{pmatrix} B^TBu^* \\ 0 \\ 0 \end{pmatrix}$$

Elimination of u and p blocks yields the equation for q:

$$(G^TG + \beta I)q = -G^TBu^*$$

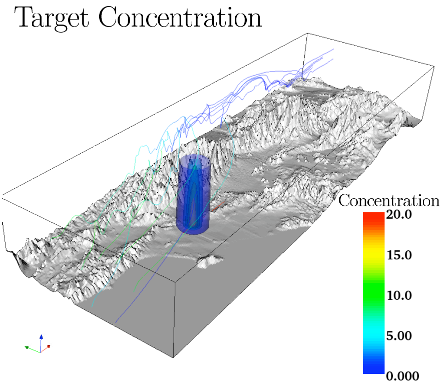
where

$$\begin{aligned} G &= BA^{-1}T \text{ is the input–output map} \\ H &= G^TG + \beta I \text{ is the (reduced) Hessian} \\ &= T^TA^{-T}B^TBA^{-1}T + \beta I \end{aligned}$$

Use CG to solve system, form hessian-vector product on the fly at cost of one forward/adjoint PDE solve per iteration.

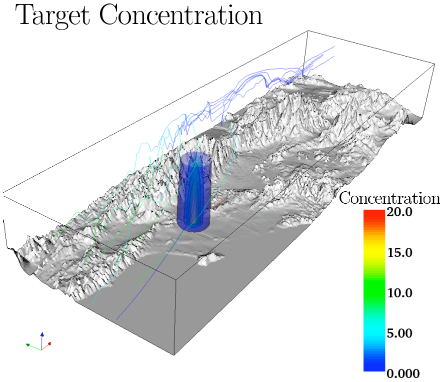
Inverse atmospheric transport: Typical solution

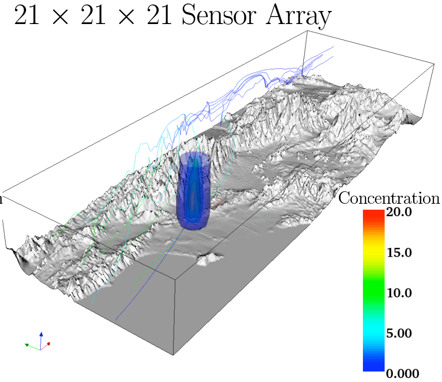
Solution of an airborne contaminant inverse problem in the Greater Los Angeles Basin with onshore winds; mesh Peclet = 10



Inverse atmospheric transport: Typical solution

Solution of an airborne contaminant inverse problem in the Greater Los Angeles Basin with onshore winds; mesh Peclet = 10


Target Concentration


21 × 21 × 21 Sensor Array

Inverse atmospheric transport: Scalability

Fixed size scalability of unpreconditioned and multigrid preconditioned inversion; problem size is 257^4

CPUs	no preconditioner		multigrid	
	hours	efficiency	hours	efficiency
128	5.65	1.00	2.22	1.00
512	1.41	1.00	0.76	0.73
1024	0.74	0.95	0.48	0.58

Isogranular scalability of unpreconditioned and multigrid preconditioned inversion:

grid	problem size		CPUs	no precond.		multigrid	
	q	(u, p, q)		hours	iter	hours	iter
129^4	2.15E+6	5.56E+8	16	2.13	23	1.05	8
257^4	1.70E+7	8.75E+9	128	5.65	23	2.22	6
513^4	1.35E+8	1.39E+11	1024	—	—	4.89	5

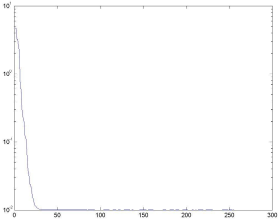
Presented at SC’05; built on PETSc library

A peek into why CG is so effective for Hessians with “compact + identity” structure

At iteration k, CG solves the weighted least squares problem

$$\min_{P_k} ||e_k|| = \sum_i P_k \left[\lambda_i\right]^2 \xi_i^2 \lambda_i$$

where P_k is polynomial of order k and $e_0 = \sum_i \xi_i v_i, \quad H v_i = \lambda_i v_i$



Example spectrum of data misfit ($\beta = 0$) portion of Hessian ($H \overset{\text{def}}{=} G^T G + \beta I$)

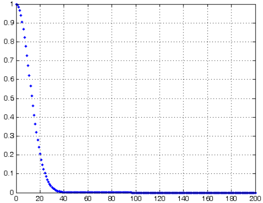
Analytical example of Hessian spectrum

1D advection-diffusion with periodic boundary conditions, inversion for initial condition with final time observations

$$\min_q \int_0^L (u - u^*(T))^2 dx + \frac{\beta}{2} \int_0^L q^2 dx$$

$$\begin{aligned} \text{where: } u_t - k u_{xx} + v u_x &= 0 \text{ in } (0,L) \times (0,T) \\ k u_x(0,t) &= k u_x(L,t) \text{ for } t \in (0,T) \\ u(0,t) &= u(L,t) \text{ for } t \in (0,T) \\ u &= q \text{ in } (0,L) \times \{t = 0\} \end{aligned}$$

Hessian: jth eigenfunction: $e^{2\pi i j x/L}$, jth eigenvalue: $e^{-8j^2\pi^2 kT/L^2}$



Outline

1. Lessons from deterministic inverse problems
2. Bayesian framework for statistical inverse problems
3. MCMC sampling
4. Langevin methods and stochastic Newton
5. Results for an inverse wave propagation problem

Bayesian formulation for statistical inversion

(Tarantola framework)

- Given:
  - a forward model $\mathbf{g}(\mathbf{m}) = \mathbf{d}$ relating model parameters $\mathbf{m}$ with observables $\mathbf{d}$, and its uncertainty
  - actual observations $\mathbf{d}_{\text{obs}}$ and their uncertainty
  - a “prior” estimate, of model parameters, $\mathbf{m}_{\text{prior}}$, and its uncertainty
- Seek a statistical characterization of model parameters consistent with observations, forward model, and prior model

Bayesian formulation for statistical inversion

Given:

$$\begin{aligned} \rho_{\mathrm{M}}(\mathbf{m}) \quad &:= \quad \text{prior p.d.f. of model parameters} \\ \rho_{\mathrm{D}}(\mathbf{d}) \quad &:= \quad \text{prior p.d.f. of the data} \\ \theta(\mathbf{d}|\mathbf{m}) \quad &:= \quad \text{conditional p.d.f. relating } \mathbf{d} \text{ and } \mathbf{m} \end{aligned}$$

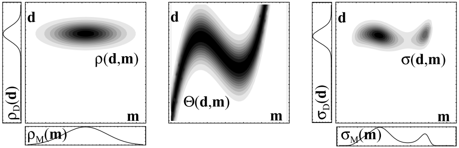
Then posterior p.d.f. of model parameters is given by:

$$\sigma_{\mathrm{M}}(\mathbf{m}) = \; k \;\; \rho_{\mathrm{M}}(\mathbf{m}) \int_{\mathfrak{D}} \frac{\rho_{\mathrm{D}}(\mathbf{d})\, \theta(\mathbf{d}|\mathbf{m})}{\mu_{\mathrm{D}}(\mathbf{d})} d\mathbf{d}$$

$L(\mathbf{m}) :=$ likelihood function

constant for linear data space

normalization constant



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

Bayesian formulation for statistical inversion

Gaussian uncertainties, nonlinear forward model

If forward model uncertainty is Gaussian:

$$\theta(\mathbf{d}|\mathbf{m}) = \text{const.} \exp\left(-\frac{1}{2}\,(\mathbf{d}-\mathbf{g}(\mathbf{m}))^T\ \mathbf{C}_{\mathrm{T}}^{-1}\ (\mathbf{d}-\mathbf{g}(\mathbf{m}))\right)$$

forward model covariance

and observation uncertainty is Gaussian:

$$\rho_{\mathrm{D}}(\mathbf{d}) = \text{const.} \exp\left(-\frac{1}{2}\,(\mathbf{d}-\mathbf{d}_{\mathrm{obs}})^T\ \mathbf{C}_{\mathrm{d}}^{-1}\ (\mathbf{d}-\mathbf{d}_{\mathrm{obs}})\right)$$

observation covariance

and prior model parameter uncertainty is Gaussian:

$$\rho_{\mathrm{M}}(\mathbf{m}) = \text{const.} \exp\left(-\frac{1}{2}\,(\mathbf{m}-\mathbf{m}_{\mathrm{prior}})^T\ \mathbf{C}_{\mathrm{M}}^{-1}\ (\mathbf{m}-\mathbf{m}_{\mathrm{prior}})\right)$$

prior model parameter covariance

Then the posterior model parameter p.d.f. is given by:

$$\sigma_{\mathrm{M}}(\mathbf{m}) = k \exp\left(-\,S\,(\mathbf{m})\right)$$

not Gaussian!

where the misfit function is:

$$S(\mathbf{m}) := \frac{1}{2}\,(\mathbf{g}(\mathbf{m})-\mathbf{d}_{\mathrm{obs}})^T\ \mathbf{C}_{\mathrm{D}}^{-1}\ (\mathbf{g}(\mathbf{m})-\mathbf{d}_{\mathrm{obs}}) + \frac{1}{2}\,(\mathbf{m}-\mathbf{m}_{\mathrm{prior}})^T\,\mathbf{C}_{\mathrm{M}}^{-1}\,(\mathbf{m}-\mathbf{m}_{\mathrm{prior}})$$

$\mathbf{C}_{\mathrm{D}} = \mathbf{C}_{\mathrm{T}} + \mathbf{C}_{\mathrm{d}}$ forward model and measurement uncertainty combine

Bayesian formulation for statistical inversion

Gaussian uncertainties, linear forward problem

If modeling, measurement, and prior uncertainties are all Gaussian, and if in addition the forward problem is linear, i.e.,

$$\mathbf{G}\,\mathbf{m} = \mathbf{d}$$

Then the posterior p.d.f. for the model parameters is also Gaussian:

$$\sigma_{\mathrm{M}}(\mathbf{m}) = k\exp\left(-\,S\left(\mathbf{m}\right)\right)$$

where

$$\begin{aligned} 2\,S(\mathbf{m}) := & \left(\mathbf{G}\,\mathbf{m} - \mathbf{d}_{\mathrm{obs}}\right)^{T}\mathbf{C}_{\mathrm{D}}^{-1}\left(\mathbf{G}\,\mathbf{m} - \mathbf{d}_{\mathrm{obs}}\right) \\ & + \left(\mathbf{m} - \mathbf{m}_{\mathrm{prior}}\right)^{T}\mathbf{C}_{\mathrm{M}}^{-1}\left(\mathbf{m} - \mathbf{m}_{\mathrm{prior}}\right) \end{aligned}$$

Bayesian formulation for statistical inversion

Gaussian uncertainties, linear forward problem (continued)

Since the posterior p.d.f. for the model parameters is Gaussian, its mean can be found by maximizing the p.d.f., which is equivalent to solving the weighted least squares optimization problem:

$$\tilde{\mathbf{m}} = \arg\min S(\mathbf{m}) := \|\mathbf{G}\,\mathbf{m} - \mathbf{d}_{\text{obs}}\|^2_{\mathbf{C}_{\text{D}}^{-1}} + \|\mathbf{m} - \mathbf{m}_{\text{prior}}\|^2_{\mathbf{C}_{\text{M}}^{-1}}$$

Note the connection with the regularization approach to inverse problems: $\mathbf{C}_{\text{M}}^{-1}$ plays the role of the regularizer.

The posterior parameter covariance is given by the inverse of the Hessian:

$$\tilde{\mathbf{C}}_{\text{M}} = \left(\mathbf{G}^T\,\mathbf{C}_{\text{D}}^{-1}\,\mathbf{G} + \mathbf{C}_{\text{M}}^{-1}\right)^{-1}$$

Note also the posterior p.d.f. for the data is also Gaussian, with mean and covariance given by:

$$\tilde{\mathbf{d}} = \mathbf{G}\,\tilde{\mathbf{m}} \qquad \tilde{\mathbf{C}}_{\text{D}} = \mathbf{G}\,\tilde{\mathbf{C}}_{\text{M}}\,\mathbf{G}^T$$

Outline

1 Lessons from deterministic inverse problems

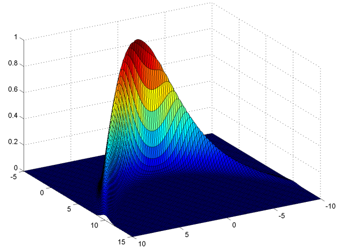
2 Bayesian framework for statistical inverse problems

3 MCMC sampling

4 Langevin methods and stochastic Newton

5 Results for an inverse wave propagation problem

MCMC sampling

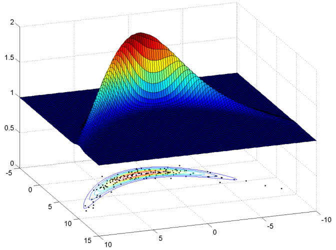


Example Probability Density

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 $\alpha > \mathrm{rand}([0,1])$ 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:

- $q(\boldsymbol{x}_k, \boldsymbol{y}) = \pi(\boldsymbol{y})$
- $\alpha(\boldsymbol{x}_k, \boldsymbol{y}) = \min\left(1, \frac{\pi(\boldsymbol{y})\pi(\boldsymbol{x}_k)}{\pi(\boldsymbol{x}_k)\pi(\boldsymbol{y})}\right) \equiv 1$
- Can also use any 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

Delayed Rejection Adaptive Metropolis (DRAM)

Delayed rejection

- If proposal $\boldsymbol{y}$ is rejected, don't give up
- Use new proposal: $q_1(\boldsymbol{x}_k, \boldsymbol{y}, \boldsymbol{y}_1)$
- q_1 is typically much more conservative
- Second accept/reject step with similar criterion
- We can delay rejection multiple times

Adaptation of proposal covariance

- Modify proposal $\boldsymbol{\Gamma}$ with current sample covariance
- Non-Markovian, but still converges

Approaches to reducing the cost of each sample

- Reduced model of the forward problem
  - POD (e.g. Wang and Zabaras, Willcox et al.)
  - Stochastic Galerkin (e.g Marzouk, Najm and Rahn)
- Reduced model of the outputs (i.e. response surface)
  - Gaussian process model (e.g. O'Hagan and Kennedy)
  - Stochastic response surface (Balakrishnan, Roy, Ierapetritou, Flach, Georgopoulos)
- "Preconditioned" MCMC using reduced order models
  - Higdon, Lee, and Holloman
  - Christen and Fox
  - Efendiev, Hou, and Luo
  - Efendiev, Datta-Gupta, Ginting, Ma, and Mallick

Desired properties for speeding up sampling algorithms

- Scale to high-dimensional parameter spaces
- Take advantage of known properties of the misfit function (e.g., gradient, low rank approximation of Hessian, ...)
- Reuse techniques developed for the deterministic inverse problem
- Build on experience from linear inverse problem

Outline

1 Lessons from deterministic inverse problems

2 Bayesian framework for statistical inverse problems

3 MCMC sampling

4 Langevin methods and stochastic Newton

5 Results for an inverse wave propagation problem

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 algorithm

Preconditioned Langevin MCMC

Given a probability density of the form:

$$\pi(\boldsymbol{x}) = c \; \exp(-V(\boldsymbol{x}))$$

The associated Langevin SDE is given by:

$$d\boldsymbol{X}_t = -\boldsymbol{A}\nabla V \, dt + \sqrt{2}\boldsymbol{A}^{1/2} d\boldsymbol{W}_t$$

Discretize with a timestep Δt to derive Langevin proposal:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{A}\nabla V \, \Delta t + \sqrt{2\Delta t}\boldsymbol{A}^{1/2}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

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
  - $\boldsymbol{W}_{(t+\Delta t)} - \boldsymbol{W}_t \sim \mathcal{N}\big(\boldsymbol{0}, \Delta t \; \boldsymbol{I}\big)$
- $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ is the i.i.d. standard normal pdf

Stochastic Newton's method

- Standard Langevin MCMC proposal given by:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{A}\nabla V \Delta t + \sqrt{2\Delta t}\boldsymbol{A}^{1/2}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

- For preconditioner $\boldsymbol{A}$, use the inverse of the (local) Hessian $\boldsymbol{H}(\boldsymbol{x}) = \nabla^2 V$ to precondition; set $\Delta t = 1$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}^{-1}\nabla V + \boldsymbol{H}^{-1/2}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

- This becomes the stochastic equivalent of Newton's method

Stochastic Newton: Optimal sampling of Gaussians

- Consider a Gaussian density $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$:

$$\begin{aligned} V &= \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\ \nabla V &= \boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\ \boldsymbol{H} = \nabla^2 V &= \boldsymbol{\Gamma}^{-1} \end{aligned}$$

- Apply Stochastic Newton:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k - \boldsymbol{H}^{-1}\nabla V + \boldsymbol{H}^{-1/2}\mathcal{N}(\mathbf{0}, \boldsymbol{I}) \\ &= \boldsymbol{x}_k - \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}_k - \boldsymbol{\mu}) + \boldsymbol{\Gamma}^{1/2}\mathcal{N}(\mathbf{0}, \boldsymbol{I}) \\ &= \boldsymbol{\mu} + \boldsymbol{\Gamma}^{1/2}\mathcal{N}(\mathbf{0}, \boldsymbol{I}) \\ &= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \end{aligned}$$

- Samples $\boldsymbol{x}_k$ act like independent draws from the true PDF

Classical vs. Stochastic Newton

Classical Newton:

- Given a cost function: $V(\boldsymbol{x})$
- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}^{-1}\nabla V$
- Optimizes best fit quadratic in one step

Stochastic Newton:

- Given a probability density: $\exp\big(-V(\boldsymbol{x})\big)$
- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}^{-1}\nabla V + \boldsymbol{H}^{-1/2}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
- Samples best fit Gaussian in one step

Vanilla flavor Langevin resembles steepest descent

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \Delta t \nabla V + \sqrt{\Delta t}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

Stochastic Newton: Computational considerations

Need local Hessians and gradients for each MCMC step

- Can be expensive... even to reject proposal points
- Use adjoints if available
- Exploit all of the machinery of large scale deterministic inverse problems (low rank approximations, inexact Newton methods, trust region ideas, etc.)

$\boldsymbol{H}$ not always positive definite away from $\boldsymbol{x}_{\mathrm{map}}$

- Current implementation:
  - Compute eigenvector decomposition: $\boldsymbol{H} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T$
  - Replace small or negative eigenvalues with floor threshold
  - Apply the square root as $\boldsymbol{H}^{1/2} = \boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{V}^T$
- Future: Inexact Newton and/or Gauss-Newton

Outline

1 Lessons from deterministic inverse problems

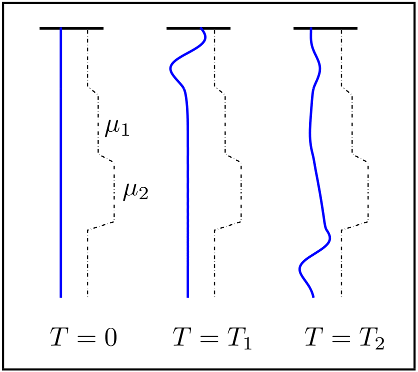
2 Bayesian framework for statistical inverse problems

3 MCMC sampling

4 Langevin methods and stochastic Newton

5 Results for an inverse wave propagation problem

1D seismic wave propagation



1D Seismic wave propagation

Seismic wave propagation forward model:

- 1D wave equation
- Ricker wavelet source at surface
- Measure reflected wavefield
- Add receiver noise

Inverse Problem:

- Discretize medium into n layers
- Reconstruct shear modulus of medium
- Bayesian inversion framework

The forward model

Given $\mu(x)$, we solve the 1D wave equation:

$$\rho\frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial x}\Big(\mu(x)\frac{\partial}{\partial x}u\Big) = \delta(x-0)\cdot F(t)$$

$$\begin{aligned}
\sqrt{\rho\mu}\ \frac{\partial u}{\partial t}\Big|_{x=1} &= -\mu\cdot\ \frac{\partial u}{\partial z}\Big|_{x=1} \\
\mu\ \frac{\partial u}{\partial z}\Big|_{x=0} &= 0 \\
u\big|_{t=0} &= 0 \\
\dot{u}\big|_{t=0} &= 0
\end{aligned}$$

and observe the displacement field $u(0,t)$ at surface

Bayesian inversion setting

Uncertainty quantification problem:

- Layered medium with two parameters (μ_1, μ_2)
- Uniform prior on $[0.5, 10] \times [0.5, 10]$

$$\pi_{\text{pr}}(\boldsymbol{\mu}) \propto 1$$

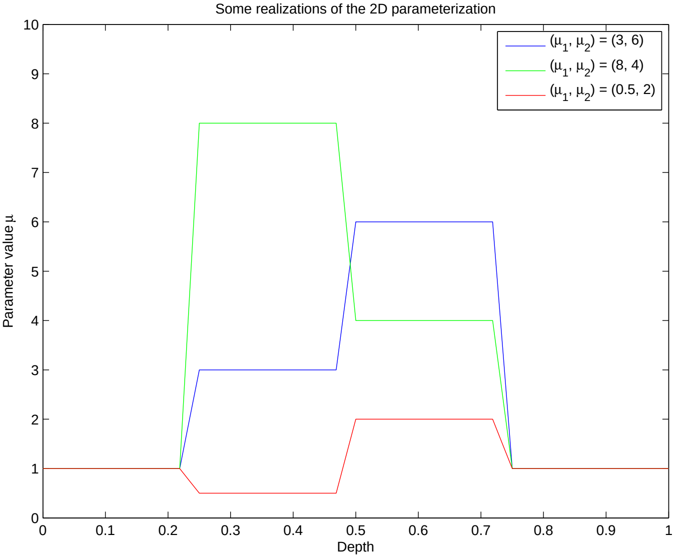
- Gaussian likelihood function:

$$\pi_{\text{like}}(y_{\text{obs}}|\boldsymbol{\mu}) = \exp\Big(- \frac{1}{2}(y(\boldsymbol{\mu}) - y_{\text{obs}})^T \Gamma_{\text{noise}}^{-1}(y(\boldsymbol{\mu}) - y_{\text{obs}})\Big)$$

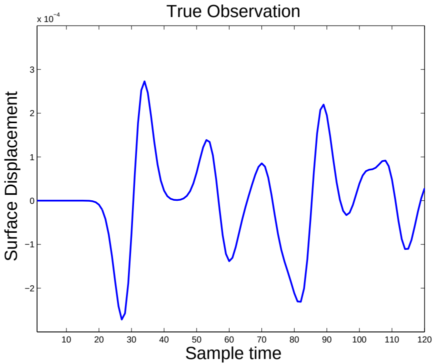
- Ultimately we wish to sample the posterior distribution:

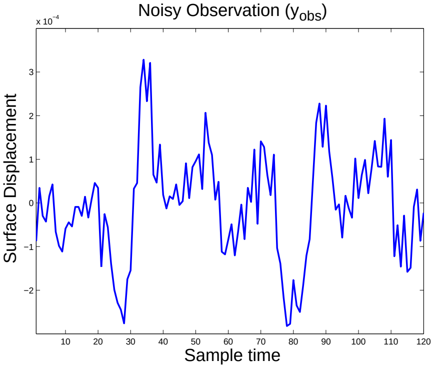
$$\pi_{\text{post}}(\boldsymbol{\mu}|y_{\text{obs}}) \propto \pi_{\text{pr}}(\boldsymbol{\mu})\pi_{\text{like}}(y_{\text{obs}}|\boldsymbol{\mu})$$

2D parameterization of input space

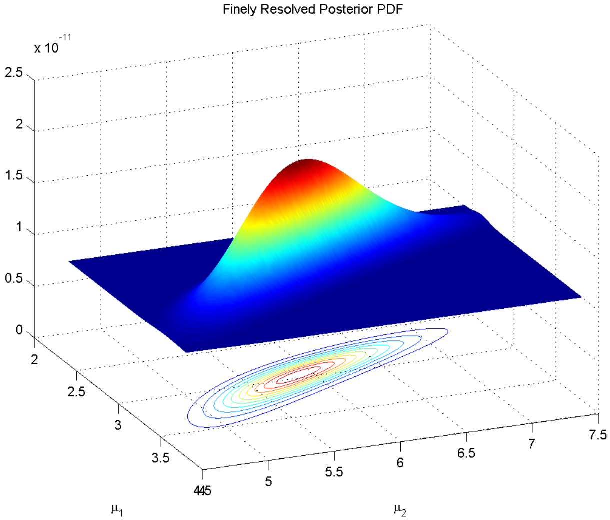


Sample observation data

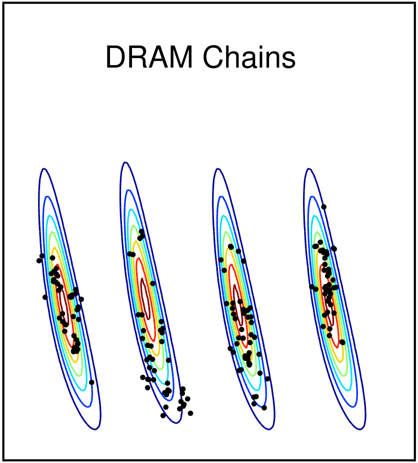




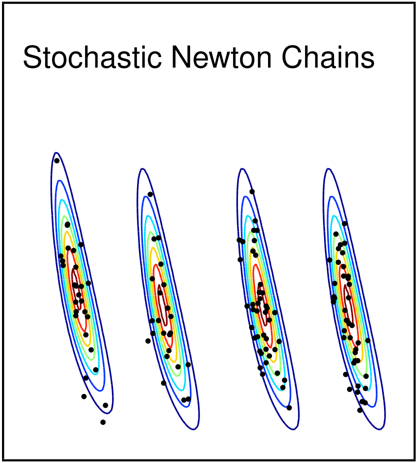
2D posterior density



DRAM chains vs. Stochastic Newton chains

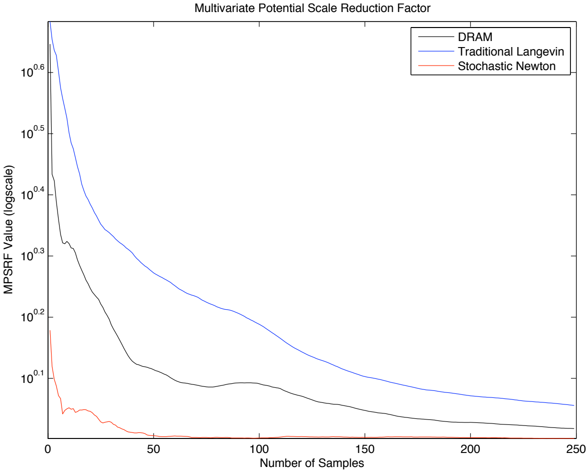


DRAM Chains



Stochastic Newton Chains

2D MPSRF for DRAM vs. Stochastic Newton



MPSRF statistic for 2 layer parametrization

Stochastic Newton samples cost $\approx 3\times$ one DRAM sample

16 Layer inversion problem

As before, solve Bayesian inverse problem for 16 layer parameters

- Gaussian smoothness prior between layers:
    - Covariance $\mathbf{\Gamma}$ between layers i and j:

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

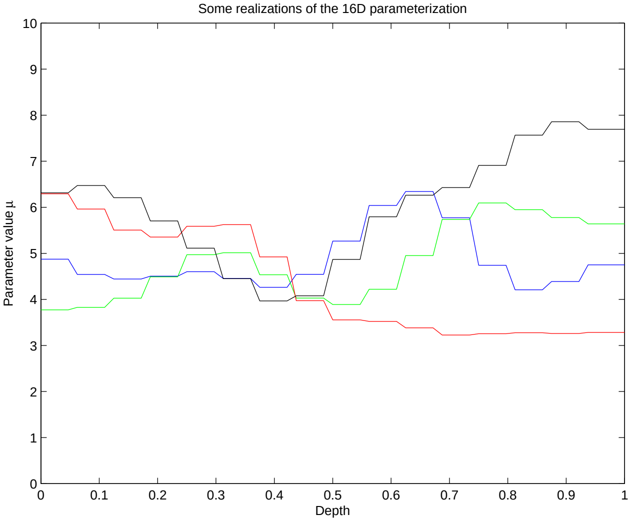
    - Prior mean is constant $\mu(\boldsymbol{x}) = 5$
- Gaussian likelihood function:

$$\pi_{\text{like}}(y_{\text{obs}}|\boldsymbol{\mu}) = \exp\Big(-\frac{1}{2}(F(\boldsymbol{\mu}) - y_{\text{obs}})^T \Gamma_{\text{noise}}^{-1}(F(\boldsymbol{\mu}) - y_{\text{obs}})\Big)$$

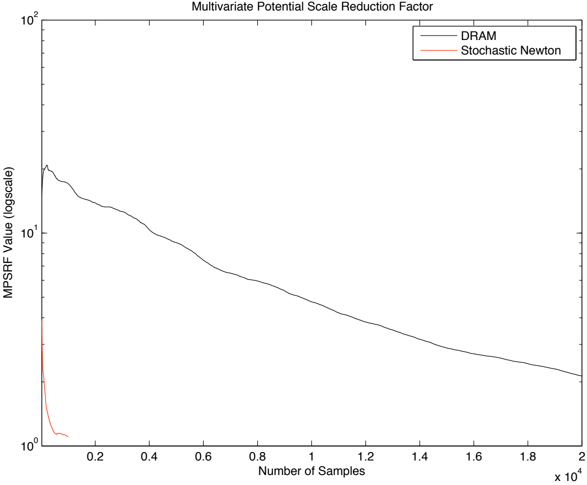
- Again we wish to sample the posterior distribution:

$$\pi_{\text{post}}(\boldsymbol{\mu}|y_{\text{obs}}) \propto \pi_{\text{pr}}(\boldsymbol{\mu})\pi_{\text{like}}(y_{\text{obs}}|\boldsymbol{\mu})$$

16D parameterization of input space



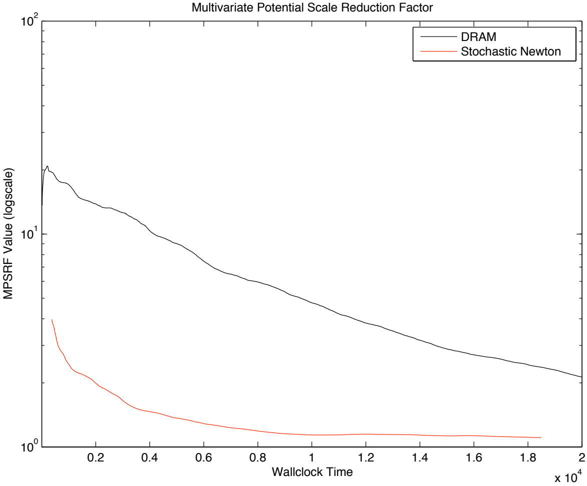
16D MPSRF plots



MPSRF statistic for 16 layer parametrization

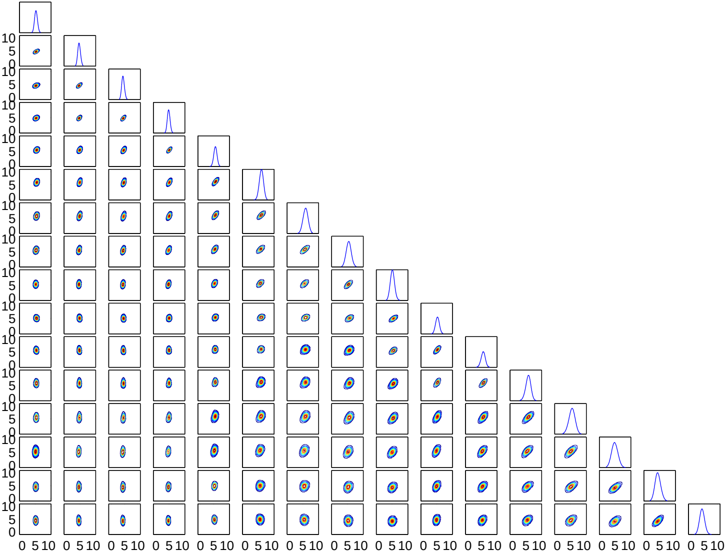
Stochastic Newton samples cost $\approx 18.5\times$ one DRAM sample

16D MPSRF plots (rescaled)

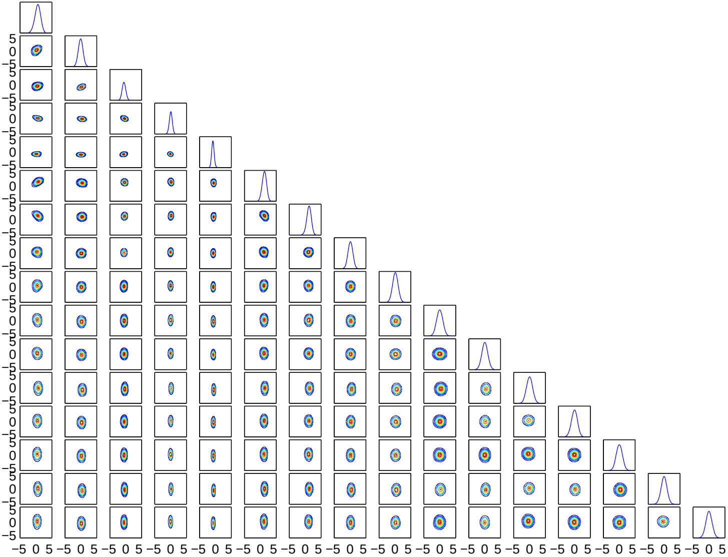


MPSRF statistic for 16 layer parametrization
Axes rescaled to reflect total computation time

KDE cross correlation plots – physical basis



KDE cross correlation plots – prior basis



65D inversion problem

Solve Bayesian inverse problem for 65-layer parametrization

- Gaussian smoothness prior between grid points:
  - Covariance $\mathbf{\Gamma}$ between grid points i and j:

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

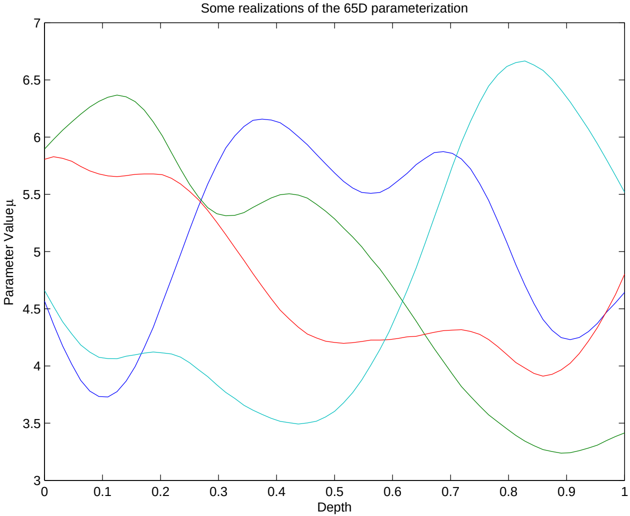
  - Prior mean is constant $\mu(\boldsymbol{x}) = 5$
- Gaussian likelihood function:

$$\pi_{\text{like}}(y_{\text{obs}}|\boldsymbol{\mu}) = \exp\Big(-\frac{1}{2}(y(\boldsymbol{\mu}) - y_{\text{obs}})^T \Gamma_{\text{noise}}^{-1}(y(\boldsymbol{\mu}) - y_{\text{obs}})\Big)$$

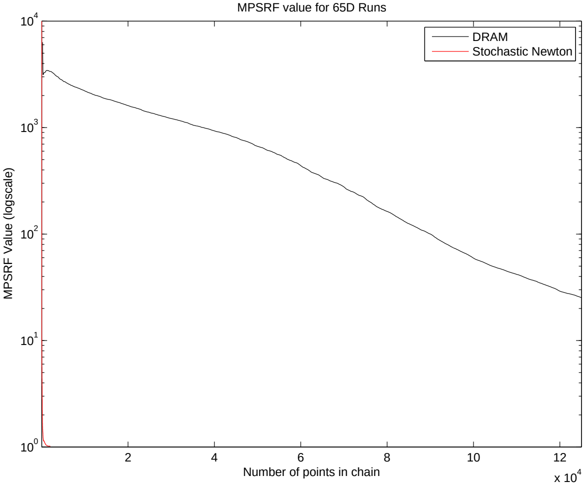
- Again we wish to sample the posterior distribution:

$$\pi_{\text{post}}(\boldsymbol{\mu}|y_{\text{obs}}) \propto \pi_{\text{pr}}(\boldsymbol{\mu})\pi_{\text{like}}(y_{\text{obs}}|\boldsymbol{\mu})$$

65D parametrization of input space



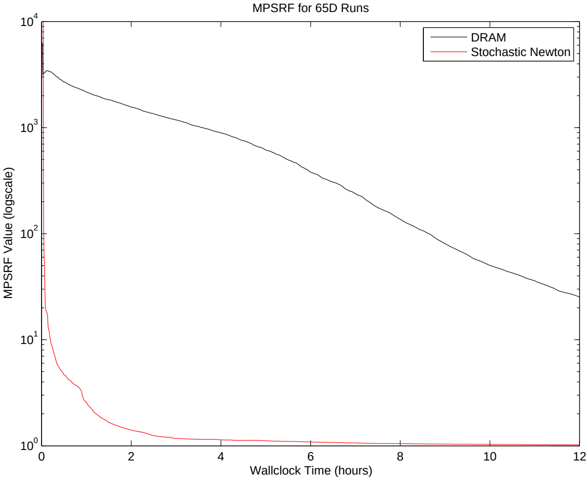
65D MPSRF plots



MPSRF statistic for 65D parameterization

Stochastic Newton samples cost $\approx 82\times$ one DRAM sample

65D MPSRF plots (rescaled)



MPSRF statistic for 65 layer parameterization

Axes rescaled to reflect total computation time

Conclusions

- For linear statistical inverse problems, fast methods can be constructed that exploit low rank approximations of the Hessian
- Hessian-preconditioned Langevin MCMC (aka Stochastic Newton)
  - motivated by connection to deterministic Newton method
  - exactly samples a Gaussian posterior
  - naive implementation shows several orders of magnitude improvement over DRAM
- Can capitalize on several decades of advances in deterministic PDE-based optimization and inverse methods to vastly improve stochastic Newton, e.g.
  - inexact Newton (Eisenstat-Walker, negative curvature ideas)
  - trust region methods
  - exploit “compact + differential” structure of Hessians (e.g. low rank approximations, Fredholm-multigrid type preconditioners)
- I believe that exploiting deterministic PDE inverse problem structure is mandatory for scaling MCMC to high dimensions and expensive forward problems