Bayesian Inversion for the WIPP groundwater flow problem

Björn Sprungk

joint work:
O. Ernst (U Chemnitz), K. A. Cliffe, (U Nottingham)

Workshop on Multiscale Inverse Problems
Mathematics Institue, University of Warwick
June 17–19, 2013

TECHNISCHE UNIVERSITÄT
CHEMNITZ
FAKULTÄT FÜR MATHEMATIK

Outline

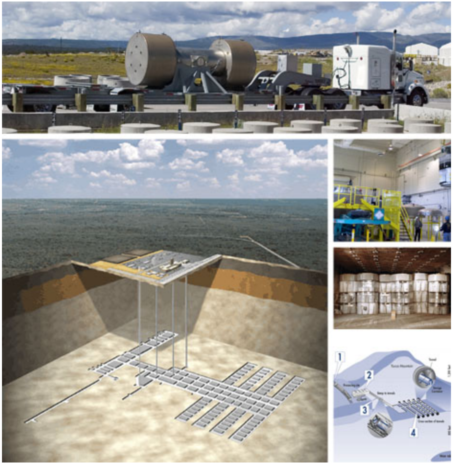
Next

1. Bayesian Inversion for the WIPP groundwater flow problem
   - Problem setting and motivation
   - Obtaining the prior
   - Bayesian Inversion using MCMC
   - Bayesian Inversion via EnKF
2. Gauss-Newton (preconditioned) MH-MCMC methods
3. Conclusions

Next

UQ Problem

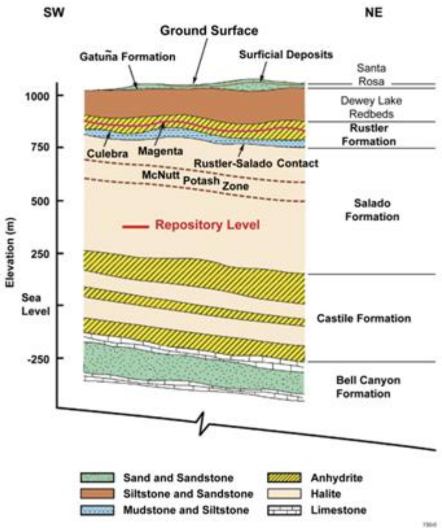
Radioactive Waste Repository Site Assessment



- Waste Isolation Pilot Plant (WIPP) Carlsbad, New Mexico

UQ Problem

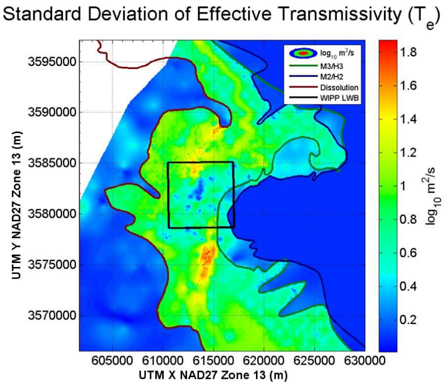
Radioactive Waste Repository Site Assessment



- Waste Isolation Pilot Plant (WIPP) Carlsbad, New Mexico
- Groundwater transport of radionuclides in transmissive Culebra layer

UQ Problem

Radioactive Waste Repository Site Assessment

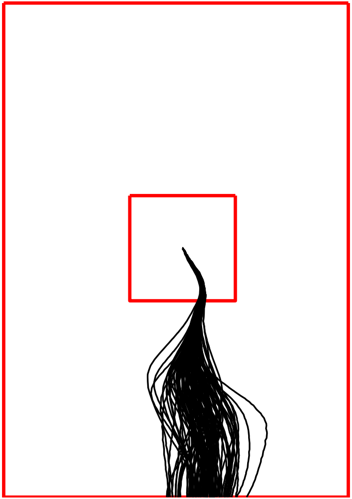


source: Sandia National Labs

- Waste Isolation Pilot Plant (WIPP) Carlsbad, New Mexico
- Groundwater transport of radionuclides in transmissive Culebra layer
- Uncertainty in hydraulic conductivity

UQ Problem

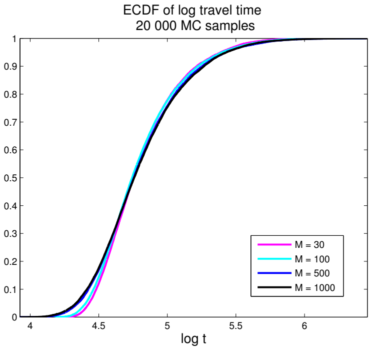
Radioactive Waste Repository Site Assessment



- Waste Isolation Pilot Plant (WIPP) Carlsbad, New Mexico
- Groundwater transport of radionuclides in transmissive Culebra layer
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time to boundary of WIPP site

UQ Problem

Radioactive Waste Repository Site Assessment



- Waste Isolation Pilot Plant (WIPP) Carlsbad, New Mexico
- Groundwater transport of radionuclides in transmissive Culebra layer
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time to boundary of WIPP site
- **Approach:** Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.

Groundwater Flow Model

Stationary Darcy flow	$\mathbf{q} = -K\nabla p$	$\mathbf{q}$: Darcy flux
		K : hydraulic conductivity
		p : hydraulic head
mass conservation	$\mathrm{div}\ \mathbf{u} = 0$	$\mathbf{u}$: pore velocity
	$\mathbf{q} = \phi\mathbf{u}$	ϕ : porosity
transmissivity	$T = Kb$	b : aquifer thickness
Particle transport	$\dot{\mathbf{x}}(t) = -\frac{T(\mathbf{x})}{b\phi}\nabla p(\mathbf{x})$	$\mathbf{x}$: particle position
	$\mathbf{x}(0) = \mathbf{x}_0$	$\mathbf{x}_0$: release location

Quantity of interest s**:** $\log_{10}$ of travel time of particle to reach WIPP boundary, in particular, its cumulative distribution function (cdf).

PDE with Random Coefficient

Mixed form of Darcy equations:

$$T^{-1}(\mathbf{x})\mathbf{u}(\mathbf{x}) + \nabla p(\mathbf{x}) = 0, \qquad p|_{\partial D} = p_0,$$
$$\operatorname{div} \mathbf{u}(\mathbf{x}) = 0, \quad \mathbf{x} \in D.$$

Model T as a random field (RF) $T = T(\mathbf{x}, \omega)$, $\omega \in \Omega$, with respect to underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$.

PDE with Random Coefficient

Mixed form of Darcy equations:

$$T^{-1}(\mathbf{x})\mathbf{u}(\mathbf{x}) + \nabla p(\mathbf{x}) = 0, \qquad p|_{\partial D} = p_0,$$
$$\operatorname{div} \mathbf{u}(\mathbf{x}) = 0, \quad \mathbf{x} \in D.$$

Model T as a random field (RF) $T = T(\mathbf{x}, \omega)$, $\omega \in \Omega$, with respect to underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$.

Assumptions:

- T has finite mean and covariance

$$\overline{T}(\mathbf{x}) = \mathbf{E}\left[T(\mathbf{x}, \cdot)\right], \qquad \mathbf{x} \in D,$$
$$\mathrm{Cov}_T(\mathbf{x}, \mathbf{y}) = \mathbf{E}\left[\left(T(\mathbf{x}, \cdot) - \overline{T}(\mathbf{x})\right)\left(T(\mathbf{y}, \cdot) - \overline{T}(\mathbf{y})\right)\right], \qquad \mathbf{x}, \mathbf{y} \in D.$$

- T is lognormal, i.e., $Z(\mathbf{x}, \omega) := \log T(\mathbf{x}, \omega)$ is a Gaussian RF.
- Cov_Z is stationary, isotropic, and of Matérn type.

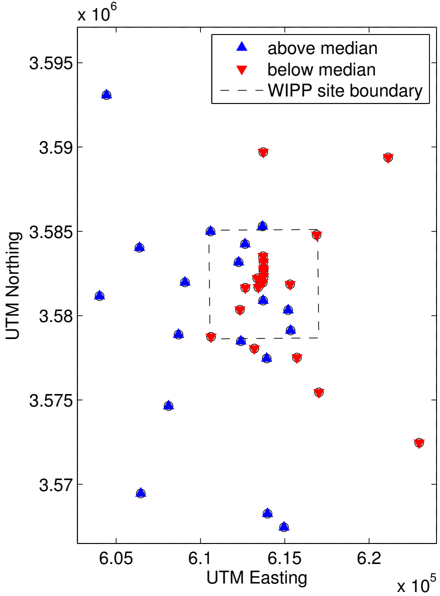
Next

1. Bayesian Inversion for the WIPP groundwater flow problem
   - Problem setting and motivation
   - Obtaining the prior
   - Bayesian Inversion using MCMC
   - Bayesian Inversion via EnKF

2. Gauss-Newton (preconditioned) MH-MCMC methods

3. Conclusions

WIPP Data



- transmissivity measurements at 38 test wells
- head measurements at 33 test wells, used to obtain boundary data via statistical interpolation (kriging)
- constant layer thickness of $b = 8$m
- constant porosity of $\phi = 0.16$
- SANDIA Nat. Labs reports [Caufman et al., 1990] [La Venue et al., 1990]

Prior Probabilistic Model of Transmissivity

Merge transmissivity data with statistical model

Given: Measurements of log transmissivity $(\mathbf{x}_j, \kappa_j)$, $j = 1, \ldots, N$

Assumptions:

- Linear model $\overline{\kappa}(\mathbf{x}) = \sum_{i=1}^{n} \beta_i f_i(\mathbf{x})$ for mean
- Matérn covariance structure for fluctuations around mean

Procedure:

(1) $\{f_i\}_{i=1}^{n}$ yield point estimates of parameters in Matérn covariance function via restricted maximum likelihood estimation (REML).

(2) Krige RF $\log T$: best linear prediction of $\kappa(\mathbf{x})$ based on measurements (for Gaussian RF coincides with conditioning on observations).

(3) Approximate $\log T$ by truncated Karhunen-Loève expansion.

Matérn Family of Covariance Kernels

$$c(\mathbf{x},\mathbf{y}) = c_{\boldsymbol{\theta}}(r) = \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,r}{\rho}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu}\,r}{\rho}\right), \quad r = \|\mathbf{x}-\mathbf{y}\|_2$$

K_ν : modified Bessel function of order ν

Parameters $\boldsymbol{\theta} = (\sigma^2, \rho, \nu)$

σ^2 : variance

ρ : correlation length

ν : smoothness parameter

Special cases:

$\nu = \frac{1}{2}$:	$c(r) = \sigma^2 \exp(-\sqrt{2}r/\rho)$	exponential covariance
$\nu = 1$:	$c(r) = \sigma^2 \left(\frac{2r}{\rho}\right) K_1\left(\frac{2r}{\rho}\right)$	Bessel covariance
$\nu \to \infty$:	$c(r) = \sigma^2 \exp(-r^2/\rho^2)$	Gaussian covariance

Smoothness: Realizations $Z(\cdot,\omega)$ are k times differentiable $\Leftrightarrow \nu > k$.

Kriging

Best unbiased linear prediction

Given RF κ with known covariance and observations $\{\kappa(\mathbf{x}_j) = \kappa_j\}_{j=1}^N$, approximate $\mathbf{E}\left[\kappa | \{\kappa(\mathbf{x}_j) = \kappa_j\}_{j=1}^N\right]$ by linear prediction

$$\hat{\kappa}(\mathbf{x}) = m_0(\mathbf{x}) + \sum_{j=1}^N m_j(\mathbf{x})\, \kappa(\mathbf{x}_j)$$

such that it is unbiased $\mathbf{E}\left[\hat{\kappa}(\mathbf{x})\right] = \mathbf{E}\left[\kappa(\mathbf{x})\right]$ and optimal

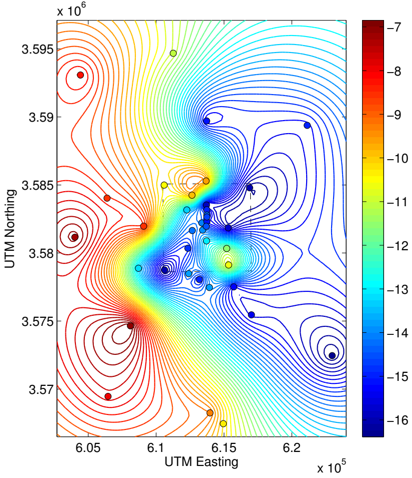
$$\mathbf{E}\left[\left(\hat{\kappa}(\mathbf{x}) - \kappa(\mathbf{x})\right)^2\right] \to \min_{m_0,\dots,m_N}!$$

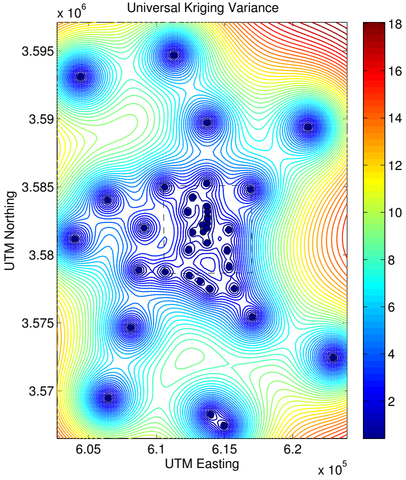
Variants: **Simple Kriging** (assumes known mean)
Universal Kriging (assumes linear regression model for mean)

Kriging

WIPP results

We assume constant but unknown mean $\bar{\kappa}(\mathbf{x}) \equiv \beta$.
REML estimates for covariance: $\sigma^2 = 18.9$, $\rho = 9865$, $\nu = 0.59$.





Parametrization of Input RF

Karhunen-Loève expansion

$$\kappa(\mathbf{x},\omega) = \bar{\kappa}(\mathbf{x}) + \sum_{m=1}^{\infty} \sqrt{\lambda_m}\,\phi_m(\mathbf{x})\,\xi_m(\omega)$$

converges in $L^\infty(D)$ and $L^2_{\mathbb{P}}(\Omega)$, where

$\{(\lambda_m,\phi_m)\}_{m\in\mathbb{N}}$ eigenpairs of covariance operator

$$(C\phi)(\mathbf{x}) = \int_D \phi(\mathbf{y})\ \mathrm{Cov}_\kappa(\mathbf{x},\mathbf{y})\,d\mathbf{y},$$

$\{\xi_m\}_{m\in\mathbb{N}} \subset L^2_{\mathbb{P}}(\Omega)$, $\quad \mathbf{E}[\xi_m] = 0, \quad \mathbf{E}[\xi_k\xi_m] = \delta_{k,m}.$

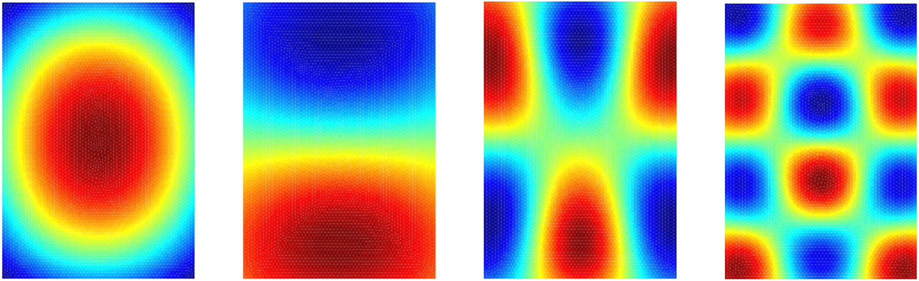
Truncation after M terms yields RF κ_M with

$$\mathbf{E}\left[\|\kappa-\kappa_M\|^2_{L^2(D)}\right] = \sum_{m=M+1}^{\infty} \lambda_m.$$

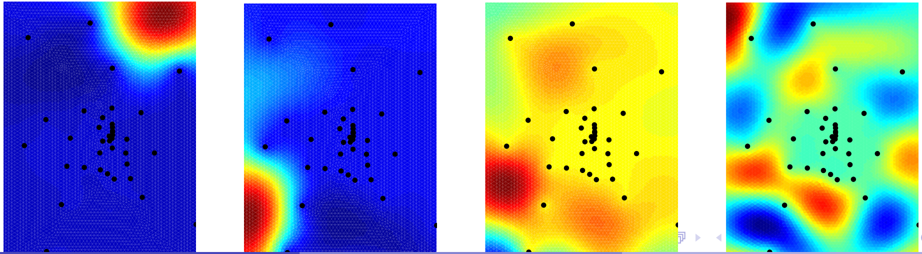
WIPP KL Modes

Conditioned on 38 transmissivity observations

unkriged, $m = 1, 2, 9, 16$



kriged, $m = 1, 2, 9, 16$



Next

The Inverse Problem

Bayesian Approach

$$\exp(-\kappa)\mathbf{u} = -\nabla p, \quad \mathrm{div}\ \mathbf{u} = 0, \quad p|_{\partial D} = p_0$$

Further reduction of uncertainty using head measurements:

- Finite number of observations of p: $\mathbf{y} = Q(p) = Q(p(\kappa)) \in \mathbb{R}^k$.
- Measurement noise: $\mathbf{d} = \mathbf{y} + \varepsilon$, $\varepsilon \sim N(\mathbf{0}, \mathbf{\Gamma})$
- Prior measure μ_0 for $\kappa \in \mathrm{span}\{\phi_1, \dots, \phi_M\} \subset L^\infty(D)$ by measurements of κ

The Inverse Problem

Bayesian Approach

$$\exp(-\kappa)\mathbf{u} = -\nabla p, \quad \mathrm{div}\ \mathbf{u} = 0, \quad p|_{\partial D} = p_0$$

Further reduction of uncertainty using head measurements:

- Finite number of observations of p: $\mathbf{y} = Q(p) = Q(p(\kappa)) \in \mathbb{R}^k$.
- Measurement noise: $\mathbf{d} = \mathbf{y} + \varepsilon$, $\varepsilon \sim N(\mathbf{0}, \mathbf{\Gamma})$
- Prior measure μ_0 for $\kappa \in \mathrm{span}\{\phi_1, \dots, \phi_M\} \subset L^\infty(D)$ by measurements of κ
- Bayes theorem yields conditional distribution $\mu_{\mathbf{d}}$ of $\kappa|\mathbf{d}$

$$\frac{\mathrm{d}\mu_{\mathbf{d}}}{\mathrm{d}\mu_0} \propto \exp(\Phi(\kappa)), \qquad \Phi(\kappa) = \frac{1}{2}\|\mathbf{d} - G(\kappa)\|^2_{\mathbf{\Gamma}^{-1}}, \qquad G = Q \circ p.$$

Goal: Compute cdf of QoI $s(\kappa)$ according to $\kappa \sim \mu_{\mathbf{d}}$.

Bayesian Inversion for WIPP

Sampling from the posterior

Markov Chain Monte Carlo (MCMC)

- Construct Markov chain with stationary distribution $\mu_{\mathbf{d}}$.
- Sample sequence usually highly correlated, need for subsampling.
- Prior model for κ:

$$\kappa_M(\mathbf{x}, \boldsymbol{\xi}) = \phi_0(\mathbf{x}) + \sum_{m=1}^{M} \phi_m(\mathbf{x})\, \xi_m, \qquad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_m) \sim N(0, I),$$

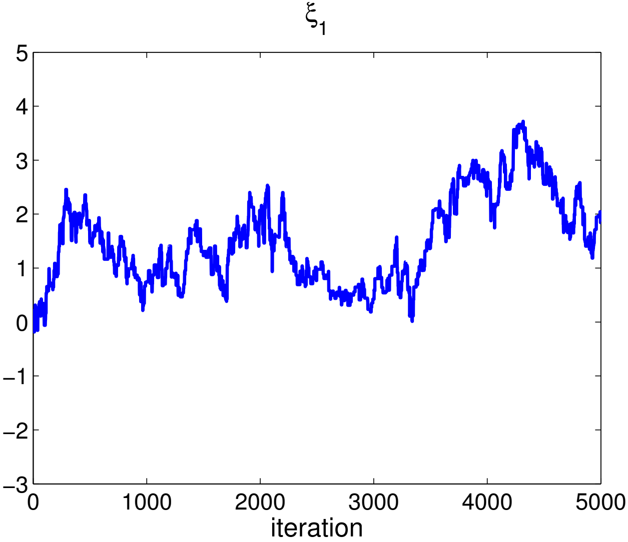
- Parametrization: prior on $\boldsymbol{\xi} \in \mathbb{R}^M$ instead of $\kappa \in L^\infty(D)$

$$\mu_0 \sim N(0, I) \text{ on } \mathbb{R}^M,$$

- Data: head measurements $\mathbf{d} = (p(\mathbf{x}_1), \ldots, p(\mathbf{x}_k)) \in \mathbb{R}^k$, here $k = 33$.
- Due to high dimension M: pCN-MCMC proposed in [Cotter et. al, 2012]

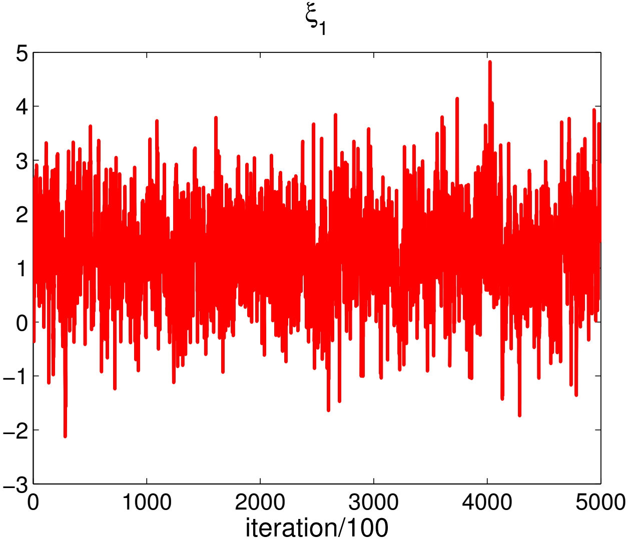
WIPP: MH-MCMC

Preliminary Results for $M = 1000$



WIPP: MH-MCMC

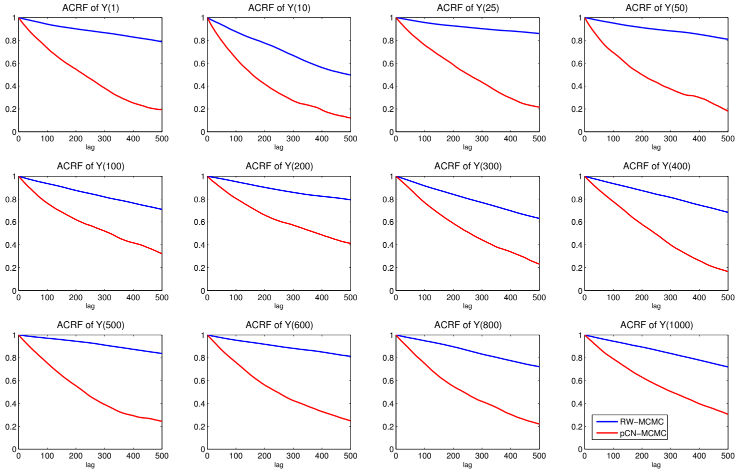
Preliminary Results for $M = 1000$



WIPP: MH-MCMC

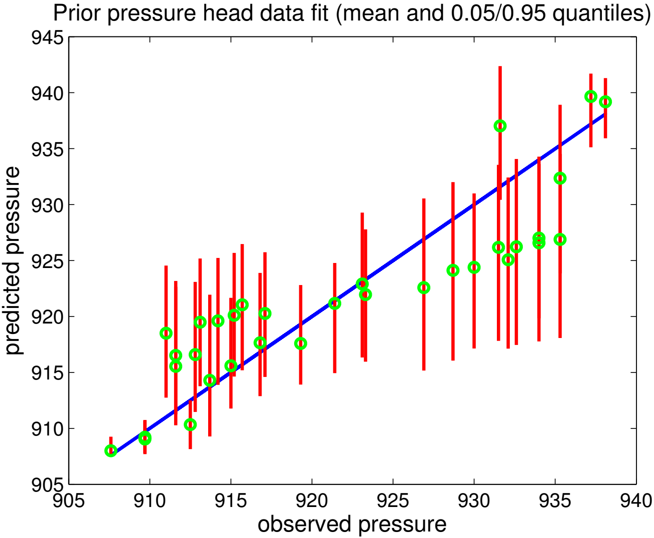
Preliminary Results for $M = 1000$

Comparing pCN and standard Random Walk



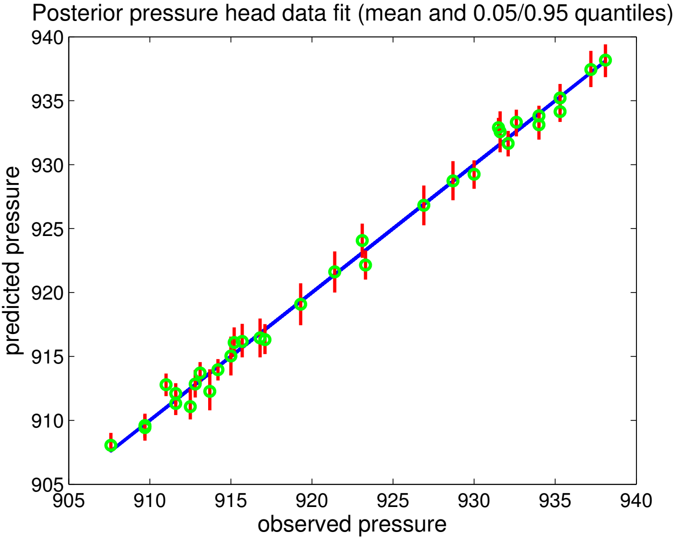
WIPP: Posterior Data Fit

Preliminary Results for $M = 1000$



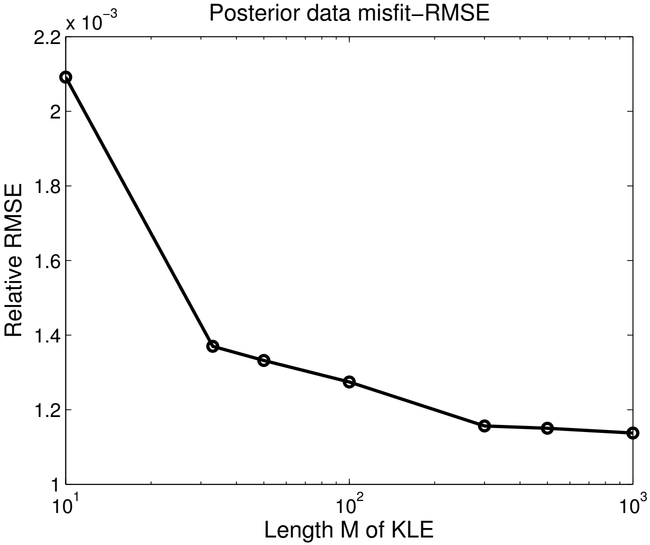
WIPP: Posterior Data Fit

Preliminary Results for $M = 1000$



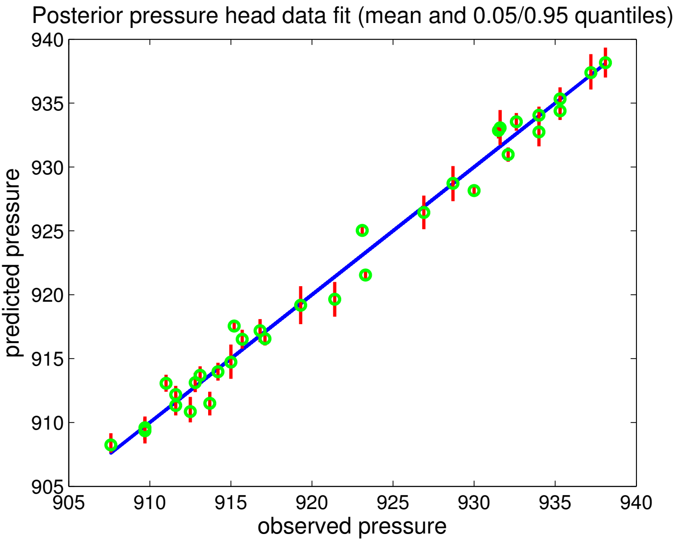
WIPP: Posterior Data Fit

Preliminary Results for $M = 1000$



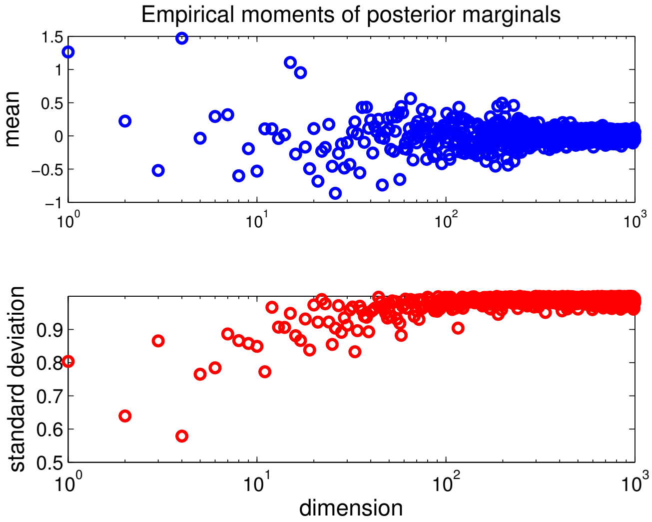
WIPP: Posterior Data Fit

Preliminary Results for $M = 33$



WIPP: Posterior distribution of ξ

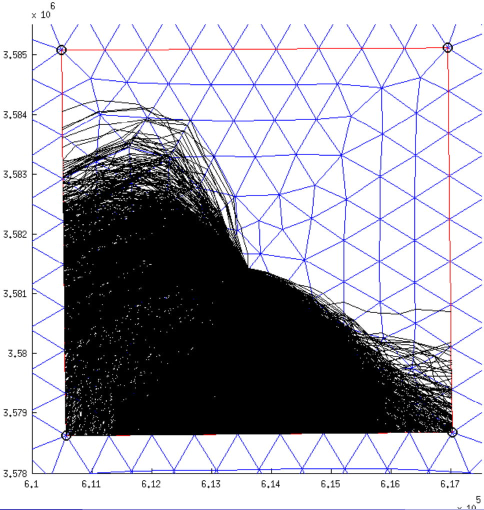
Preliminary Results for $M = 1000$



WIPP: Posterior distribution of QoI

Preliminary Results for $M = 1000$

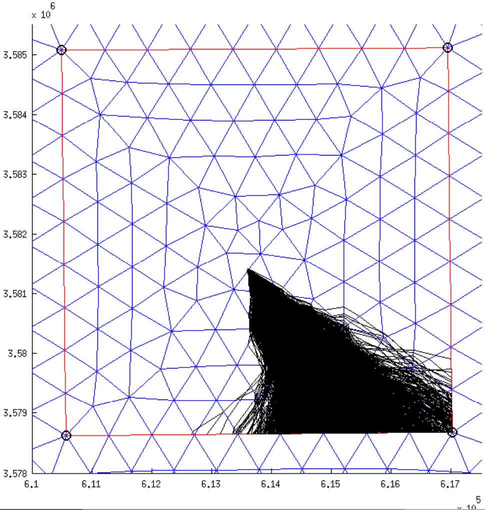
Particle trajectories according to prior



WIPP: Posterior distribution of QoI

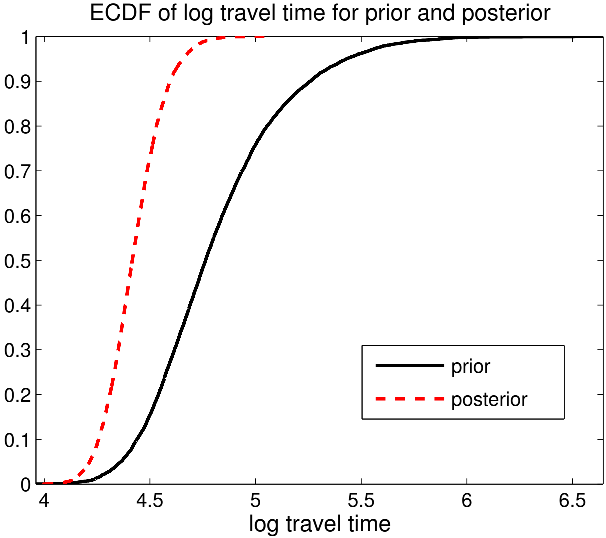
Preliminary Results for $M = 1000$

Particle trajectories according to posterior



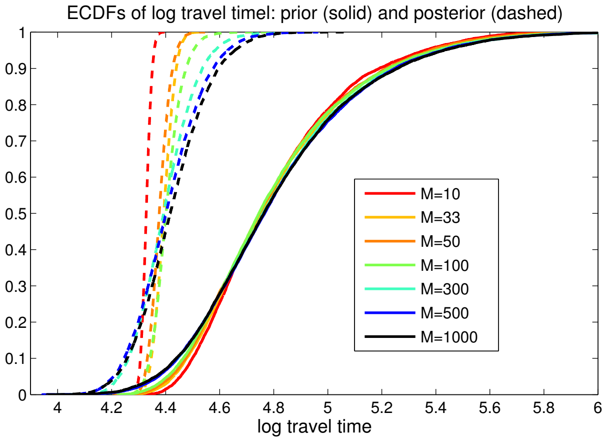
WIPP: Posterior distribution of QoI

Preliminary Results for $M = 1000$



WIPP: Posterior distribution of QoI

Preliminary Results for $M = 1000$



Next

1. Bayesian Inversion for the WIPP groundwater flow problem
   - Problem setting and motivation
   - Obtaining the prior
   - Bayesian Inversion using MCMC
   - Bayesian Inversion via EnKF
2. Gauss-Newton (preconditioned) MH-MCMC methods
3. Conclusions

Inversion via EnKF

Background

- EnKF derived from Kalman filter for state estimation for incompletely observed (stochastic) nonlinear dynamics.
- Yields linear update of state estimate and estimation error by linear update of the ensemble
- State estimation by mean of ensemble, estimation error by covariance of ensemble.
- Can be applied to time-independent problems, too, e.g. elliptic PDEs.

Inversion via EnKF

Background

- EnKF derived from Kalman filter for state estimation for incompletely observed (stochastic) nonlinear dynamics.
- Yields linear update of state estimate and estimation error by linear update of the ensemble
- State estimation by mean of ensemble, estimation error by covariance of ensemble.
- Can be applied to time-independent problems, too, e.g. elliptic PDEs.
- Recent application of linear Bayesian update for UQ for PDE models [Matthies et. al, 2012]

$$\boldsymbol{\xi}_{\mathbf{d}}(\omega) = \boldsymbol{\xi}(\omega) + \mathbf{K}(\mathbf{d} - G(\boldsymbol{\xi}(\omega)) - \varepsilon(\omega)),$$

where $\mathbf{K} = \mathrm{Cov}_{\boldsymbol{\xi},G(\boldsymbol{\xi})}\left[\mathrm{Cov}_{G(\boldsymbol{\xi})} + \mathrm{Cov}_{\boldsymbol{\varepsilon}}\right]^{-1} \in \mathbb{R}^{M\times k}$.

$\Rightarrow$ Yields posterior random variable $\boldsymbol{\xi}_{\mathbf{d}}$ instead of posterior measure $\mu_{\mathbf{d}}$.

Inversion via EnKF

Interpretation in Bayesian (statistics) context

Set $\mathbf{Y}(\omega) = G(\boldsymbol{\xi}(\omega))$, $\mathbf{D}(\omega) = \mathbf{Y}(\omega) + \varepsilon(\omega)$ and

$$\hat{\varphi}(\mathbf{D}) = \mathbb{E}[\boldsymbol{\xi}] + \mathbf{K}\big(\mathbf{D} - \mathbb{E}[G(\boldsymbol{\xi})]\big).$$

- $\hat{\varphi}(\mathbf{D})$ linear approximation of $\mathbb{E}[\boldsymbol{\xi}|\mathbf{D}] = \varphi^*(\mathbf{D})$:

$$\hat{\varphi} = \mathrm{argmin}_{\varphi \in \mathrm{span}\{1,\mathbf{D}\}}\, \mathbb{E}\left[\|\boldsymbol{\xi} - \varphi(\mathbf{D})\|^2\right]$$

Inversion via EnKF

Interpretation in Bayesian (statistics) context

Set $\mathbf{Y}(\omega) = G(\boldsymbol{\xi}(\omega))$, $\mathbf{D}(\omega) = \mathbf{Y}(\omega) + \varepsilon(\omega)$ and

$$\hat{\varphi}(\mathbf{D}) = \mathbb{E}[\boldsymbol{\xi}] + \mathbf{K}\big(\mathbf{D} - \mathbb{E}[G(\boldsymbol{\xi})]\big).$$

- $\hat{\varphi}(\mathbf{D})$ linear approximation of $\mathbb{E}[\boldsymbol{\xi}|\mathbf{D}] = \varphi^*(\mathbf{D})$:

$$\hat{\varphi} = \operatorname{argmin}_{\varphi \in \operatorname{span}\{1,\mathbf{D}\}} \mathbb{E}\left[\|\boldsymbol{\xi} - \varphi(\mathbf{D})\|^2\right]$$

- $\mathbb{E}[\boldsymbol{\xi}_{\mathbf{d}}] = \hat{\varphi}(\mathbf{d})$ and $\operatorname{Cov}(\boldsymbol{\xi}_{\mathbf{d}}) = \mathbb{E}[\|\boldsymbol{\xi} - \hat{\varphi}(\mathbf{D})\|^2]$ (independent of $\mathbf{d}$)

Inversion via EnKF

Interpretation in Bayesian (statistics) context

Set $\mathbf{Y}(\omega) = G(\boldsymbol{\xi}(\omega))$, $\mathbf{D}(\omega) = \mathbf{Y}(\omega) + \varepsilon(\omega)$ and

$$\hat{\varphi}(\mathbf{D}) = \mathbb{E}[\boldsymbol{\xi}] + \mathbf{K}\big(\mathbf{D} - \mathbb{E}[G(\boldsymbol{\xi})]\big).$$

- $\hat{\varphi}(\mathbf{D})$ linear approximation of $\mathbb{E}[\boldsymbol{\xi}|\mathbf{D}] = \varphi^*(\mathbf{D})$:

$$\hat{\varphi} = \operatorname{argmin}_{\varphi \in \operatorname{span}\{1,\mathbf{D}\}} \mathbb{E}\left[\|\boldsymbol{\xi} - \varphi(\mathbf{D})\|^2\right]$$

- $\mathbb{E}[\boldsymbol{\xi}_{\mathbf{d}}] = \hat{\varphi}(\mathbf{d})$ and $\operatorname{Cov}(\boldsymbol{\xi}_{\mathbf{d}}) = \mathbb{E}[\|\boldsymbol{\xi} - \hat{\varphi}(\mathbf{D})\|^2]$ (independent of $\mathbf{d}$)
- In particular, for $\mathbf{r}_{\mathbf{d}} := \varphi^*(\mathbf{d}) - \hat{\varphi}(\mathbf{d})$

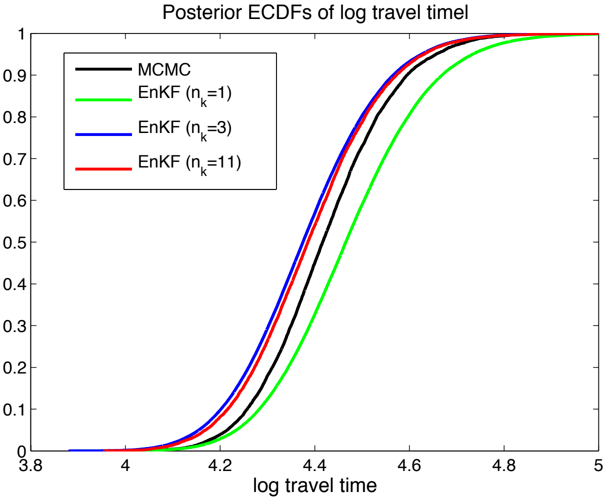
$$\operatorname{Cov}(\boldsymbol{\xi}_{\mathbf{d}}) = \int_{\mathbb{R}^k} \left(\operatorname{Cov}(\boldsymbol{\xi}|\boldsymbol{\delta}) + \mathbf{r}_{\boldsymbol{\delta}}\mathbf{r}_{\boldsymbol{\delta}}^\top\right) \, p_{\mathbf{D}}(\boldsymbol{\delta}) \,\mathrm{d}\boldsymbol{\delta},$$

where $p_{\mathbf{D}}$ density of $\mathbb{P} \circ \mathbf{D}^{-1}$.

Inversion via EnKF

Results

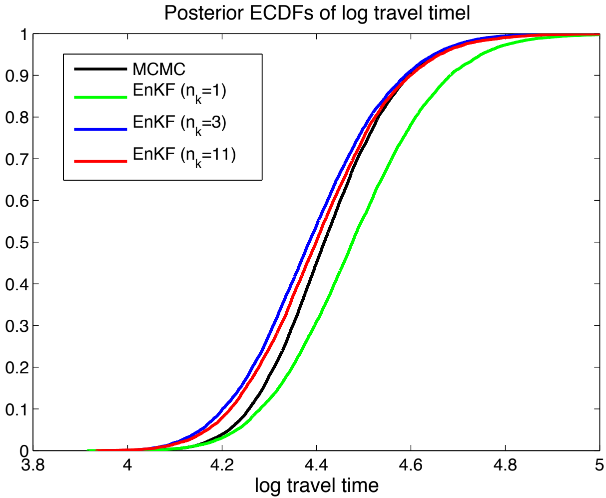
EnKF using n_k even batches $\mathbf{d} = (d_1 \ldots, d_{33}) = (\bar{\mathbf{d}}_1, \ldots, \bar{\mathbf{d}}_{n_k})$ for n_k updates.



Inversion via EnKF

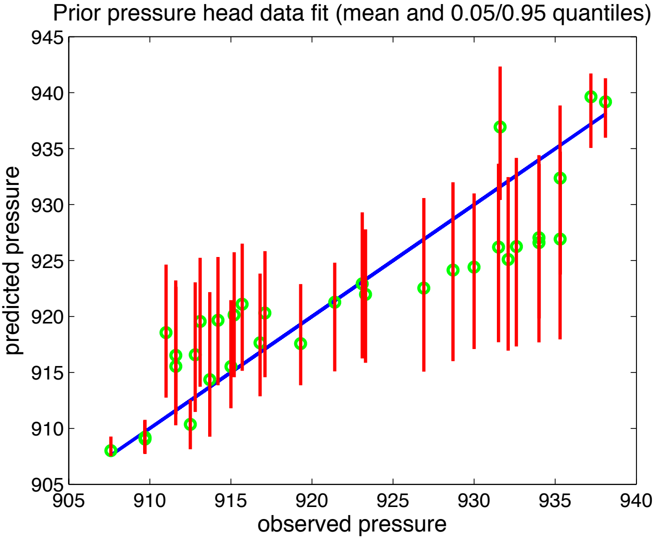
Results

Using Gaussian approximation with final ensemble mean and covariance for UQ.



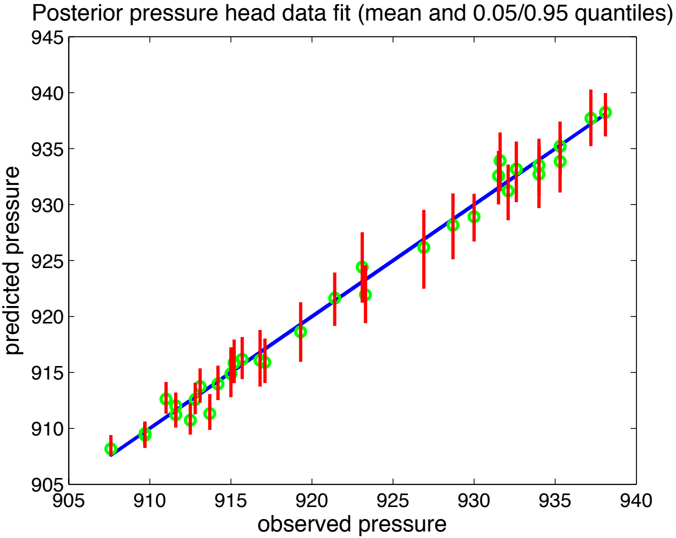
Inversion via EnKF

Results



Inversion via EnKF

Results



Inversion via EnKF

Comparison with MCMC

Errors relative to MCMC

- $$\frac{\|\kappa(\overline{\xi}_{\mathrm{MCMC}}) - \kappa(\overline{\xi}_{\mathrm{EnKF}})\|_{L^2(D)}}{\|\kappa(\overline{\xi}_{\mathrm{MCMC}})\|_{L^2(D)}}$$
- $$\frac{\|\operatorname{Cov}(\kappa(\xi_{\mathrm{MCMC}})) - \operatorname{Cov}(\kappa(\xi_{\mathrm{EnKF}}))\|_{L^2(D)\otimes L^2(D)}}{\|\operatorname{Cov}(\kappa(\xi_{\mathrm{MCMC}}))\|_{L^2(D)\otimes L^2(D)}}$$

no. updates	rel. error mean	rel. error cov	rel. data misfit
1	0.5728	0.2752	0.0026
3	0.3333	0.2578	0.0018
11	0.3516	0.2785	0.0017
MCMC	-	-	0.0011

Observations

- Need $M = 300$ KL modes to explain the data well.
- Main changes from prior to posterior in the first, say, 100 KL modes.
- For very high KL modes basically no change.
- Significant uncertainty reduction by incooperating head data (total variance reduced by 25%).
- Need to take into account $M = 1000$ KL modes for accurate estimation of posterior cdf of QoI.

Observations

- Need $M = 300$ KL modes to explain the data well.
- Main changes from prior to posterior in the first, say, 100 KL modes.
- For very high KL modes basically no change.
- Significant uncertainty reduction by incooperating head data (total variance reduced by 25%).
- Need to take into account $M = 1000$ KL modes for accurate estimation of posterior cdf of QoI.

Perspectives:

- Run inversion (MCMC) only in *active* dimensions where significant change from prior to posterior.
- Use surrogates for solving parametric PDE in these parameters for computing Hastings ratio.
- Estimate posterior cdf of QoI according to posterior in active and prior in inactive dimensions.

Next

Motivation

Observation:

Sufficiently small prior variance yields approx. linear relationship between $\boldsymbol{\xi}$ and $p(\boldsymbol{\xi})$.

Thus, for $\mu_0 \sim N(\mathbf{0}, \Sigma_0)$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$, posterior is approximately Gaussian, too,

$$\mu_{\mathbf{d}} \approx N\left(\mathbf{K}(\mathbf{d} - G(\mathbf{0})),\, \Sigma_0 - \mathbf{K}\mathbf{L}\Sigma_0 \right),$$

where $\mathbf{L} = \nabla_{\boldsymbol{\xi}} G(\mathbf{0})$ and $\mathbf{K} = \Sigma_0 \mathbf{L}^\top \left(\mathbf{L}\Sigma_0\mathbf{L}^\top + \boldsymbol{\Gamma}\right)^{-1}$. [Cliffe & Jackson, 2001]

Motivation

Observation:

Sufficiently small prior variance yields approx. linear relationship between $\boldsymbol{\xi}$ and $p(\boldsymbol{\xi})$.

Thus, for $\mu_0 \sim N(\mathbf{0}, \Sigma_0)$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Gamma})$, posterior is approximately Gaussian, too,

$$\mu_{\mathbf{d}} \approx N\left(\mathbf{K}(\mathbf{d} - G(\mathbf{0})), \Sigma_0 - \mathbf{KL}\Sigma_0 \right),$$

where $\mathbf{L} = \nabla_{\boldsymbol{\xi}} G(\mathbf{0})$ and $\mathbf{K} = \Sigma_0 \mathbf{L}^\top (\mathbf{L}\Sigma_0 \mathbf{L}^\top + \boldsymbol{\Gamma})^{-1}$. [Cliffe & Jackson, 2001]

For target $\mu_{\mathbf{d}} \sim N(\mathbf{0}, \Sigma)$ optimal Random-Walk proposal is $q(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) \sim N(\boldsymbol{\xi}, s^2\Sigma)$. [Roberts & Rosenthal, 2001]

Example

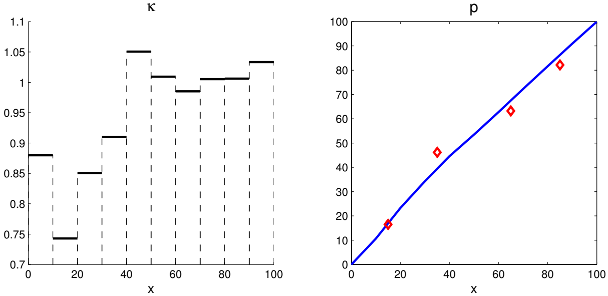
1D toy problem

$$-\nabla(\exp(\kappa)\nabla p) = 0 \text{ in } D = [0,100], \quad p(0) = 0, p(100) = 100,$$

$$\kappa(x, \boldsymbol{\xi}) = \sum_{m=1}^{10} c_m \xi_m \mathbf{1}_{[\frac{m-1}{10}, \frac{m}{10}]}(x/100), \quad \xi_m \sim N(0,1) \text{ iid},$$

$$Qp = (p(15), p(35), p(65), p(85))^\top, \quad \varepsilon \sim N(0, 0.01 I_4).$$

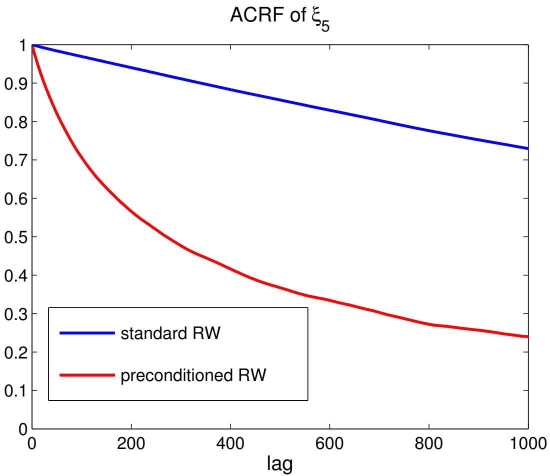
Synthetic data: $\boldsymbol{\xi}_{true}$ draw from μ_0



Example

1D toy problem

$$-\nabla(\exp(\kappa)\nabla p) = 0 \text{ in } D = [0, 100], \quad p(0) = 0, p(100) = 100,$$

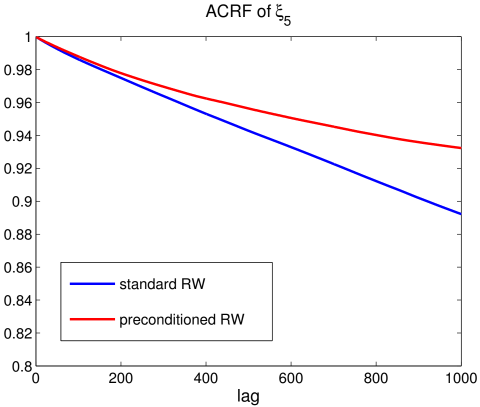


Example

1D toy problem

$$-\nabla(\exp(\kappa)\nabla p) = 0 \text{ in } D = [0, 100], \quad p(0) = 0, p(100) = 100,$$

Synthetic data: $\boldsymbol{\xi}_{true} \sim N(\mathbf{1}, 0.5\, I_{10}), \quad \kappa_{true}(x) = \sum_m \xi_{true,m}\, \mathbf{1}_{[\frac{m-1}{10}, \frac{m}{10}]}(x/100)$



Improvement

Stochastic Gauss-Newton

Local linearization for Random-Walk proposal:

$$q(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) \sim N(\boldsymbol{\xi}, s^2\Sigma(\boldsymbol{\xi})), \qquad \Sigma(\boldsymbol{\xi}) = \Sigma_0 - \Sigma_0 \mathbf{L}_{\boldsymbol{\xi}}^\top (\mathbf{L}_{\boldsymbol{\xi}}\Sigma_0\mathbf{L}_{\boldsymbol{\xi}}^\top + \boldsymbol{\Gamma})^{-1}\mathbf{L}_{\boldsymbol{\xi}}\Sigma_0,$$

where $\mathbf{L}_{\boldsymbol{\eta}} = \nabla_{\boldsymbol{\xi}} G(\boldsymbol{\eta})$.

Improvement

Stochastic Gauss-Newton

Local linearization for Random-Walk proposal:

$$q(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) \sim N(\boldsymbol{\xi}, s^2\Sigma(\boldsymbol{\xi})), \qquad \Sigma(\boldsymbol{\xi}) = \Sigma_0 - \Sigma_0 \mathbf{L}_{\boldsymbol{\xi}}^\top (\mathbf{L}_{\boldsymbol{\xi}}\Sigma_0\mathbf{L}_{\boldsymbol{\xi}}^\top + \boldsymbol{\Gamma})^{-1}\mathbf{L}_{\boldsymbol{\xi}}\Sigma_0,$$

where $\mathbf{L}_{\boldsymbol{\eta}} = \nabla_{\boldsymbol{\xi}} G(\boldsymbol{\eta})$.

Note $\Sigma(\xi) = (\mathbf{L}_{\boldsymbol{\xi}}^T \boldsymbol{\Gamma} \mathbf{L}_{\boldsymbol{\xi}} + \Sigma_0)^{-1}$ is inverse Hessian of

$$\frac{1}{2}\|\mathbf{d} - \tilde{G}(\boldsymbol{\xi})\|^2_{\boldsymbol{\Gamma}^{-1}} + \frac{1}{2}\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2_{\Sigma_0^{-1}}$$

with linearized $\tilde{G}(\boldsymbol{\eta}) = G(\boldsymbol{\xi}) + \nabla_{\boldsymbol{\xi}} G(\boldsymbol{\xi})(\boldsymbol{\eta} - \boldsymbol{\xi})$.

Improvement

Stochastic Gauss-Newton

Local linearization for Random-Walk proposal:

$$q(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) \sim N(\boldsymbol{\xi}, s^2\Sigma(\boldsymbol{\xi})), \qquad \Sigma(\boldsymbol{\xi}) = \Sigma_0 - \Sigma_0 \mathbf{L}_{\boldsymbol{\xi}}^\top \left(\mathbf{L}_{\boldsymbol{\xi}}\Sigma_0\mathbf{L}_{\boldsymbol{\xi}}^\top + \boldsymbol{\Gamma}\right)^{-1}\mathbf{L}_{\boldsymbol{\xi}}\Sigma_0,$$

where $\mathbf{L}_{\boldsymbol{\eta}} = \nabla_{\boldsymbol{\xi}} G(\boldsymbol{\eta})$.

Note $\Sigma(\boldsymbol{\xi}) = (\mathbf{L}_{\boldsymbol{\xi}}^T \boldsymbol{\Gamma} \mathbf{L}_{\boldsymbol{\xi}} + \Sigma_0)^{-1}$ is inverse Hessian of

$$\frac{1}{2}\|\mathbf{d} - \tilde{G}(\boldsymbol{\xi})\|^2_{\boldsymbol{\Gamma}^{-1}} + \frac{1}{2}\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2_{\Sigma_0^{-1}}$$

with linearized $\tilde{G}(\boldsymbol{\eta}) = G(\boldsymbol{\xi}) + \nabla_{\boldsymbol{\xi}} G(\boldsymbol{\xi})(\boldsymbol{\eta} - \boldsymbol{\xi})$.

Yields stochastic Gauss-Newton method (cf. [Martin et al., 2011])

$$q(\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\eta}) \sim N\left(\boldsymbol{\xi} - \frac{s^2}{2}\nabla_{\boldsymbol{\xi}} J(\boldsymbol{\xi})\,\Sigma(\boldsymbol{\xi}),\ \Sigma(\boldsymbol{\xi})\right),$$

where $J(\boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{d} - G(\boldsymbol{\xi})\|^2_{\boldsymbol{\Gamma}^{-1}} + \frac{1}{2}\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2_{\Sigma_0^{-1}}$.

Remarks on stochastic Gauss-Newton (sGN)

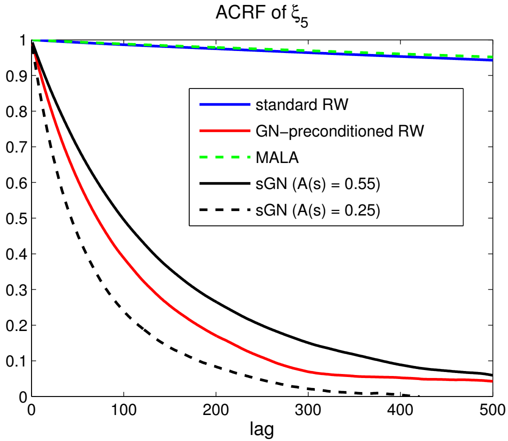
- sGN is preconditioned MALA with preconditioner $\Sigma(\xi) = (\mathbf{L}_\xi^T \mathbf{\Gamma} \mathbf{L}_\xi + \Sigma_0)^{-1}$, cf. [Girolami & Calderhead, 2011].
- $\Sigma(\xi)$ positive semi-definite in contrast to true Hessian of J.
- Computing $\Sigma_0 - \Sigma_0 \mathbf{L}_\xi^\top (\mathbf{L}_\xi \Sigma_0 \mathbf{L}_\xi^\top + \mathbf{\Gamma})^{-1} \mathbf{L}_\xi \Sigma_0$ requires only inverse of $k \times k$-matrix
- $\nabla_\xi G(\eta)$ easily computable via adjoint method.
- Since adjoint equations as original PDE, except for source term, they allow for the same surrogate (e.g. stochastic collocation).

Example

1D toy problem

$$-\nabla(\exp(\kappa)\nabla p) = 0 \text{ in } D = [0, 100], \quad p(0) = 0, p(100) = 100,$$

$$\text{Synthetic data: } \xi_{true} \sim N(\mathbf{1}, 0.5\, I_{10}), \quad \kappa_{true}(x) = \textstyle\sum_m \xi_{true,m}\, \mathbf{1}_{[\frac{m-1}{10}, \frac{m}{10}]}(x/100)$$

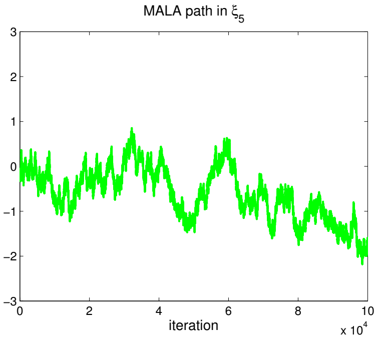


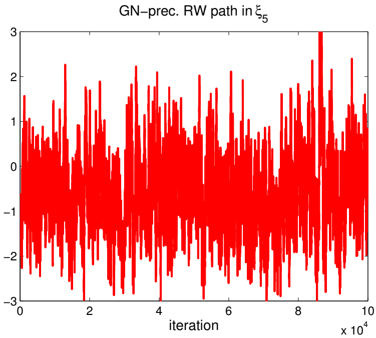
Example

1D toy problem

$$-\nabla(\exp(\kappa)\nabla p) = 0 \text{ in } D = [0, 100], \quad p(0) = 0, p(100) = 100,$$

Synthetic data: $\boldsymbol{\xi}_{true} \sim N(\mathbf{1}, 0.5\, I_{10}), \quad \kappa_{true}(x) = \sum_m \xi_{true,m}\, \mathbf{1}_{[\frac{m-1}{10}, \frac{m}{10}]}(x/100)$



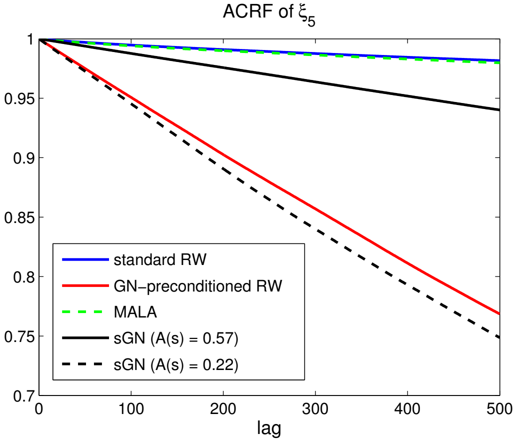


Example

1D toy problem

$$-\nabla(\exp(3\kappa)\nabla p) = 0 \text{ in } D = [0, 100], \quad p(0) = 0, p(100) = 100,$$

$$\text{Synthetic data: } \xi_{true} \sim \mu_0, \quad \kappa_{true}(x) = \sum_m 3 c_m \xi_{true,m} \mathbf{1}_{[\frac{m-1}{10}, \frac{m}{10}]}(x/100)$$



Next

Concluding Remarks

- Bayesian Inversion / MCMC challenging for WIPP problem due to high parameter dimension and high chain correlation.
- Possible remedy: reduced set of active dimensions for Bayesian inversion, advanced MCMC methods and surrogates for solving PDE.

Concluding Remarks

- Bayesian Inversion / MCMC challenging for WIPP problem due to high parameter dimension and high chain correlation.
- Possible remedy: reduced set of active dimensions for Bayesian inversion, advanced MCMC methods and surrogates for solving PDE.

- sGN methods exploit geometrical structure of posterior as stochastic Newton and Riemann manifold MCMC methods for possibly less cost
- Extension and further analysis of sGN (and preconditioned RW) to infinite dimensions to be done.

Concluding Remarks

- Bayesian Inversion / MCMC challenging for WIPP problem due to high parameter dimension and high chain correlation.
- Possible remedy: reduced set of active dimensions for Bayesian inversion, advanced MCMC methods and surrogates for solving PDE.

- sGN methods exploit geometrical structure of posterior as stochastic Newton and Riemann manifold MCMC methods for possibly less cost
- Extension and further analysis of sGN (and preconditioned RW) to infinite dimensions to be done.

- Further goal: adaptive refinement of stochastic collocation surrogates for MCMC (necessary if posterior locates in tails of prior).

Thank you for your attention!