Computational Methods in Uncertainty Quantification

Robert Scheichl

Department of Mathematical Sciences
University of Bath
Romberg Visiting Professor at University of Heidelberg

HGS MathComp Compact Course
IWR, Universität Heidelberg, June 11-17 2015

Part 4

Lecture 4

Bayesian Inverse Problems – Conditioning on Data

- Inverse Problems
- Least Squares Minimisation and Regularisation
- Bayes' Rule and Bayesian Interpretation of Inverse Problems
- Metropolis-Hastings Markov Chain Monte Carlo
- Links to what I have told you so far
- Multilevel Metropolis-Hastings Algorithm
- Some other areas of interest:
  - Data Assimilation and Filtering
  - Rare Event Estimation

Inverse Problems

What is an Inverse Problem?

Inverse problems are concerned with finding an unknown (or uncertain) **parameter vector** (or field) x from a set of typically noisy and incomplete **measurements**

$$y = H(x) + \eta$$

where η describes the noise process and $H(\cdot)$ is the *forward operator* which typically encodes a physical cause-to-consequence mapping. Typically it has a unique solution and depends continuously on data.

Inverse Problems

What is an Inverse Problem?

Inverse problems are concerned with finding an unknown (or uncertain) **parameter vector** (or field) x from a set of typically noisy and incomplete **measurements**

$$y = H(x) + \eta$$

where η describes the noise process and $H(\cdot)$ is the *forward operator* which typically encodes a physical cause-to-consequence mapping. Typically it has a unique solution and depends continuously on data.

The inverse map "H^{-1}" (from y to x) on the other hand is typically (a) **unbounded**, (b) has **multiple** or (c) **no solutions**.

(An ill-posed or ill-conditioned problem in the classical setting; Hadamard 1923.)

Inverse Problems

Examples

- **Deblurring a noisy image**
  y: image; H: blurring operator
- **Seismic**
  y: reflected wave image; H: wave propagation
- **Computer tomography**
  y: radial x-ray attenuation; H: line integral of absorption
- **Weather forecasting**
  y: satellite data, sparse indirect measurem.; H: atmospheric flow
- **Oil reservoir simulation**
  y: well pressure/flow rates, H: subsurface flow
- **Predator-prey model**
  y: state of $u_2(T)$; H: dynamical system

Inverse Problems

Linear Inverse Problems – Least Squares

Let us consider the linear forward operator $H(x) = Ax$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ with $A \in \mathbb{R}^{m \times n}$ ($n > m$, full rank) and assume that $\eta \sim \mathrm{N}(0, \alpha^2 I)$.

Least squares minimisation would seek the "best" solution $\hat{u}$ by minimising the residual norm (or the sum of squares)

$$\mathrm{argmin}_{x \in \mathbb{R}^m} \|y - Ax\|^2$$

Inverse Problems

Linear Inverse Problems – Least Squares

Let us consider the linear forward operator $H(x) = Ax$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ with $A \in \mathbb{R}^{m \times n}$ ($n > m$, full rank) and assume that $\eta \sim \mathrm{N}(0, \alpha^2 I)$.

Least squares minimisation would seek the "best" solution $\hat{u}$ by minimising the residual norm (or the sum of squares)

$$\operatorname{argmin}_{x \in \mathbb{R}^m} \|y - Ax\|^2$$

In the linear case this actually leads to a unique map

$$\hat{x} = (A^T A)^{-1} A^T y$$

which also minimises the mean-square error $\mathbf{E}\left[\|\hat{x} - x\|^2\right]$ and the covariance matrix $\mathbf{E}\left[(\hat{x} - x)(\hat{x} - x)^T\right]$ and satisfies

$$\mathbf{E}\left[\hat{x}\right] = x \quad \text{and} \quad \mathbf{E}\left[(\hat{x} - x)(\hat{x} - x)^T\right] = \alpha^2 (A^T A)^{-1}$$

Inverse Problems

Singular Value Decomposition and Error Amplification

Let $A = U\Sigma V^T$ be the *singular value decomposition* of A with $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ and $U = [u_1, ..., u_m]$, $V = [v_1, ..., v_n]$ unitary. Then we can show (Exercise) that

$$\hat{x} = \sum_{k=1}^{m} \frac{u_k^T y}{\sigma_k} v_k = x + \sum_{k=1}^{m} \frac{u_k^T \eta}{\sigma_k} v_k$$

In typical physical systems $\sigma_k \ll 1$, for $k \gg 1$, and so the "high frequency" error components $u_k^T \eta$ get amplified with $1/\sigma_k$.

Inverse Problems

Singular Value Decomposition and Error Amplification

Let $A = U\Sigma V^T$ be the *singular value decomposition* of A with $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ and $U = [u_1, ..., u_m]$, $V = [v_1, ..., v_n]$ unitary. Then we can show (Exercise) that

$$\hat{x} = \sum_{k=1}^{m} \frac{u_k^T y}{\sigma_k} v_k = x + \sum_{k=1}^{m} \frac{u_k^T \eta}{\sigma_k} v_k$$

In typical physical systems $\sigma_k \ll 1$, for $k \gg 1$, and so the "high frequency" error components $u_k^T \eta$ get amplified with $1/\sigma_k$.

In addition, if $n < m$ or if A is not full rank, then $A^T A$ is not invertible and so $\hat{x}$ is not unique (what is the physically best choice?)

Inverse Problems

Tikhonov Regularisation

A technique that guarantees uniqueness of the least squares minimiser (in the linear case) and prevents amplification of high frequency errors is *regularisation*, i.e solving instead

$$\operatorname*{argmin}_{x \in \mathbb{R}^m} \alpha^{-2} \|y - Ax\|^2 + \delta \|x - x_0\|^2$$

δ is called the *regularisation parameter* and controls how much we trust the data or how much we trust the a priori knowledge about x.

Inverse Problems

Tikhonov Regularisation

A technique that guarantees uniqueness of the least squares minimiser (in the linear case) and prevents amplification of high frequency errors is *regularisation*, i.e solving instead

$$\operatorname*{argmin}_{x \in \mathbb{R}^m} \alpha^{-2} \|y - Ax\|^2 + \delta \|x - x_0\|^2$$

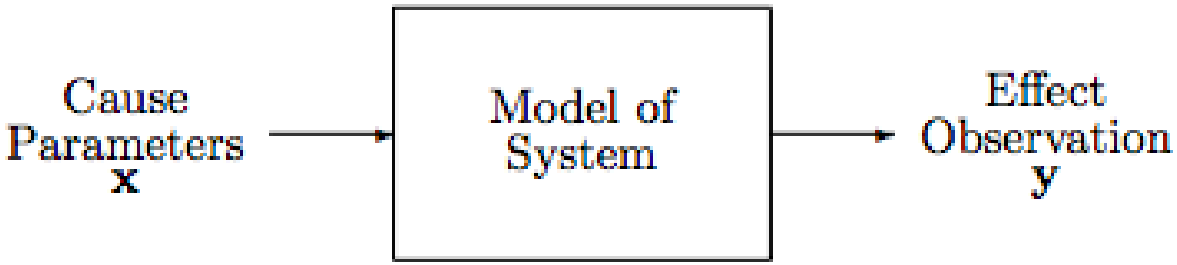
δ is called the *regularisation parameter* and controls how much we trust the data or how much we trust the a priori knowledge about x.

In general, with $\eta \sim \mathrm{N}(0, Q)$ and $H : X \to \mathbb{R}^n$ we solve

$$\operatorname*{argmin}_{x \in X} \|y - H(x)\|^2_{Q^{-1}} + \|x - x_0\|^2_{R^{-1}}$$

Inverse Problems

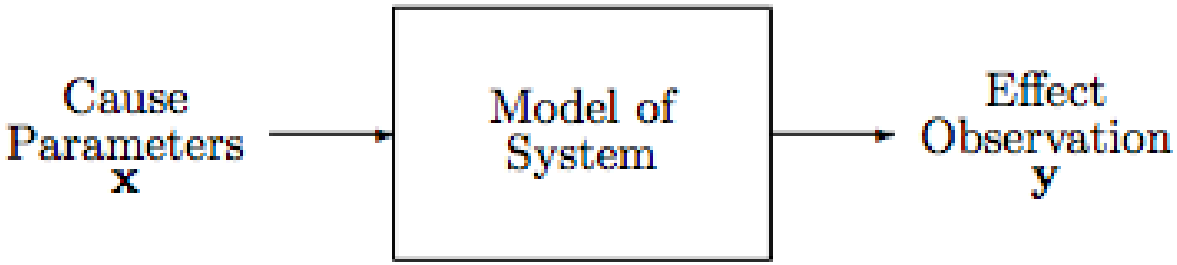
Bayesian interpretation



The (physical) model gives us $\pi(y|x)$, the conditional probability of observing y given x. However, to do UQ, to predict, to control, or to optimise we often are realy interested in $\pi(x|y)$, the conditional probability of possible causes x given the observed data y.

Inverse Problems

Bayesian interpretation



The (physical) model gives us $\pi(y|x)$, the conditional probability of observing y given x. However, to do UQ, to predict, to control, or to optimise we often are realy interested in $\pi(x|y)$, the conditional probability of possible causes x given the observed data y.

A simple consequence of $\mathbf{P}(A,B) = \mathbf{P}(A|B)\mathbf{P}(B) = \mathbf{P}(B|A)\mathbf{P}(A)$ in probability is **Bayes' rule**

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)}$$

Inverse Problems

Bayesian interpretation

In terms of probability densities **Bayes' rule** states

$$\pi(x|y) = \frac{\pi(y|x)\pi(x)}{\pi(y)}$$

- $\pi(x)$ is the **prior density** –
  represents what we know/believe about x prior to observing y
- $\pi(x|y)$ is the **posterior density** –
  represents what we know about x after observing y
- $\pi(y|x)$ is the **likelihood** –
  represents (physical) model; how likely to observe y given x
- $\pi(y)$ is the **marginal** of $\pi(x, y)$ over all possible x
  (a scaling factor that can be determined by normalisation)

Inverse Problems

Link between Bayes' Rule and Tikhonov Regularisation

Hence, the Bayesian interpretation of the least squares solution $\hat{x}$, is to find the *maximum likelihood estimate*.

Inverse Problems

Link between Bayes' Rule and Tikhonov Regularisation

Hence, the Bayesian interpretation of the least squares solution $\hat{x}$, is to find the *maximum likelihood estimate*.

The Bayesian interpretation of the regularisation term is that the prior distribution $\pi(x)$ for x is $\mathrm{N}(x_0, R)$.

Inverse Problems

Link between Bayes' Rule and Tikhonov Regularisation

Hence, the Bayesian interpretation of the least squares solution $\hat{x}$, is to find the *maximum likelihood estimate*.

The Bayesian interpretation of the regularisation term is that the prior distribution $\pi(x)$ for x is $\mathrm{N}(x_0, R)$.

The solution of the regularised least squares problem is called the *maximum a posteriori (MAP) estimator*. In the simple linear case above, it is

$$\hat{x}^{\mathrm{MAP}} = (A^T A + \delta\alpha^2 I)^{-1}(A^T y + \delta\alpha^2 x_0)$$

Inverse Problems

Link between Bayes' Rule and Tikhonov Regularisation

Hence, the Bayesian interpretation of the least squares solution $\hat{x}$, is to find the *maximum likelihood estimate*.

The Bayesian interpretation of the regularisation term is that the prior distribution $\pi(x)$ for x is $\mathrm{N}(x_0, R)$.

The solution of the regularised least squares problem is called the *maximum a posteriori (MAP) estimator*. In the simple linear case above, it is

$$\hat{x}^{\mathrm{MAP}} = (A^T A + \delta\alpha^2 I)^{-1}(A^T y + \delta\alpha^2 x_0)$$

However, in the Bayesian setting, the full posterior contains more information than the MAP estimator alone, e.g. the posterior covariance matrix $P^{-1} = (A^T Q^{-1} A + R^{-1})^{-1}$ reveals those components of x that are relatively more or less certain.

Metropolis-Hastings Markov Chain Monte Carlo

Can we do better than just finding the MAP estimator & the posterior covariance matrix?

Metropolis-Hastings Markov Chain Monte Carlo

Can we do better than just finding the MAP estimator & the posterior covariance matrix?

YES. We can sample from the posterior distribution using ...

ALGORITHM 1 (Metropolis-Hastings Markov Chain Monte Carlo)

- Choose initial state $x^0 \in X$.
- At state n generate proposal $x' \in X$ from distribution $q(x' \,|\, x^n)$
  e.g. via a random walk: $x' \sim \mathrm{N}(x^n, \varepsilon^2 \mathrm{I})$
- Accept x' as a sample with probability

$$\alpha(x'|x^n) \;=\; \min\left(1, \frac{\pi(x'|y)\, q(x^n \,|\, y)}{\pi(x^n|x')\, q(x' \,|\, x^n)}\right)$$

  i.e. $x^{n+1} = x'$ with probability $\alpha(x'|x^n)$; otherwise $x^{n+1} = x^n$.

Metropolis-Hastings Markov Chain Monte Carlo

Theorem (Metropolis et al. 1953, Hastings 1970)

Let $\pi(x|y)$ be a given probability distribution. The Markov chain simulated by the Metropolis-Hastings algorithm is **reversible** *with respect to $\pi(x|y)$. If it is also* **irreducible** *and* **aperiodic**, *then it defines an ergodic Markov chain with unique equilibrium distribution $\pi(x|y)$ (for any initial state x^0).*

The samples $f(x^n)$ of some output function ("statistic") $f(\cdot)$ can be used for inference as usual (even though not i.i.d.):

$$\mathbb{E}_{\pi(x|y)}[f(x)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x^n) := \widehat{f}^{\mathrm{MetH}}$$

Bayesian Uncertainty Quantification

Links to what I have told you so far

- What does this all have to do with UQ and with what I have told you about so far?

Bayesian Uncertainty Quantification

Links to what I have told you so far

- What does this all have to do with UQ and with what I have told you about so far?
- Bayesian statisticians often think of data as the "reality" and use the "prior" only to smooth the problem. We find sentences like
  - "It is better to use an uniformative prior."
  - "Let the data speak."
  - ...

Bayesian Uncertainty Quantification

Links to what I have told you so far

- What does this all have to do with UQ and with what I have told you about so far?
- Bayesian statisticians often think of data as the "reality" and use the "prior" only to smooth the problem. We find sentences like
  - "It is better to use an uniformative prior."
  - "Let the data speak."
  - ...
- *Bayesian Uncertainty Quantification* (in the sense that I am using it) is different in that
  - we **believe** in our physical model, **the prior**, and even require certain consistency between components
  - we usually have extremly limited output data (n v. small) and want to infer information about an ∞–dimensional parameter x.

Bayesian Uncertainty Quantification

Links to what I have told you so far

- In context of what I said so far, we essentially want to "condition" our uncertain models on information about input data (prior) and output data (likelihood).

Bayesian Uncertainty Quantification

Links to what I have told you so far

- In context of what I said so far, we essentially want to "condition" our uncertain models on information about input data (prior) and output data (likelihood).
- In the context of large-scale problems with high-dimensional input spaces, MCMC is even less tractable than standard MC.

Bayesian Uncertainty Quantification

Links to what I have told you so far

- In context of what I said so far, we essentially want to "condition" our uncertain models on information about input data (prior) and output data (likelihood).
- In the context of large-scale problems with high-dimensional input spaces, MCMC is even less tractable than standard MC.
- Again we have to distinguish whether we are interested
  - only in statistics about some Quantity of Interest (**quadrature w.r.t. the posterior** or
  - in the whole posterior distribution of the inputs (and the state)

Bayesian Uncertainty Quantification

Links to what I have told you so far

- In context of what I said so far, we essentially want to "condition" our uncertain models on information about input data (prior) and output data (likelihood).
- In the context of large-scale problems with high-dimensional input spaces, MCMC is even less tractable than standard MC.
- Again we have to distinguish whether we are interested
  - only in statistics about some Quantity of Interest (**quadrature w.r.t. the posterior** or
  - in the whole posterior distribution of the inputs (and the state)
- Often people resort to *"surrogates"/"emulators"* to make it computationally tractable (can use stochastic collocation)

Bayesian Uncertainty Quantification

Links to what I have told you so far

- In context of what I said so far, we essentially want to "condition" our uncertain models on information about input data (prior) and output data (likelihood).
- In the context of large-scale problems with high-dimensional input spaces, MCMC is even less tractable than standard MC.
- Again we have to distinguish whether we are interested
  - only in statistics about some Quantity of Interest (**quadrature w.r.t. the posterior** or
  - in the whole posterior distribution of the inputs (and the state)
- Often people resort to *"surrogates"/"emulators"* to make it computationally tractable (can use stochastic collocation)
- Can be put in ∞-dim'l setting (important for dimension independence)

Bayesian Uncertainty Quantification

Example 1: Predator-Prey Problem

In the predator-prey model, a typical variation on the problem studied so far that leads to a Bayesian UQ problem is:

Bayesian Uncertainty Quantification

Example 1: Predator-Prey Problem

In the predator-prey model, a typical variation on the problem studied so far that leads to a Bayesian UQ problem is:

1. **Prior:** $\mathbf{u}_0 \sim \overline{\mathbf{u}}_0 + \mathrm{U}(-\epsilon, \epsilon)$
2. **Data:** u_2^{obs} at time T with measurement error $\eta \sim \mathrm{N}(0, \alpha^2)$ $\Rightarrow$ likelihood model (w. bias)

$$\pi_M(u_2^{\mathrm{obs}}|\mathbf{u}_0) \;\approx\; \exp\left(\frac{-|u_2^{\mathrm{obs}} - u_{M,2}(\mathbf{u}_0)|}{\alpha^2}\right)$$

3. **Posterior:** $\pi_M(\mathbf{u}_0|u_2^{\mathrm{obs}}) \propto \pi_M(u_2^{\mathrm{obs}}|\mathbf{u}_0)\,\underbrace{\pi(\mathbf{u}_0)}_{=\text{const}}$

Bayesian Uncertainty Quantification

Example 1: Predator-Prey Problem

In the predator-prey model, a typical variation on the problem studied so far that leads to a Bayesian UQ problem is:

1. **Prior:** $\mathbf{u}_0 \sim \overline{\mathbf{u}}_0 + \mathrm{U}(-\epsilon, \epsilon)$
2. **Data:** u_2^{obs} at time T with measurement error $\eta \sim \mathrm{N}(0, \alpha^2)$ $\Rightarrow$ likelihood model (w. bias)

$$\pi_M(u_2^{\text{obs}}|\mathbf{u}_0) \;\approx\; \exp\left(\frac{-|u_2^{\text{obs}} - u_{M,2}(\mathbf{u}_0)|}{\alpha^2}\right)$$

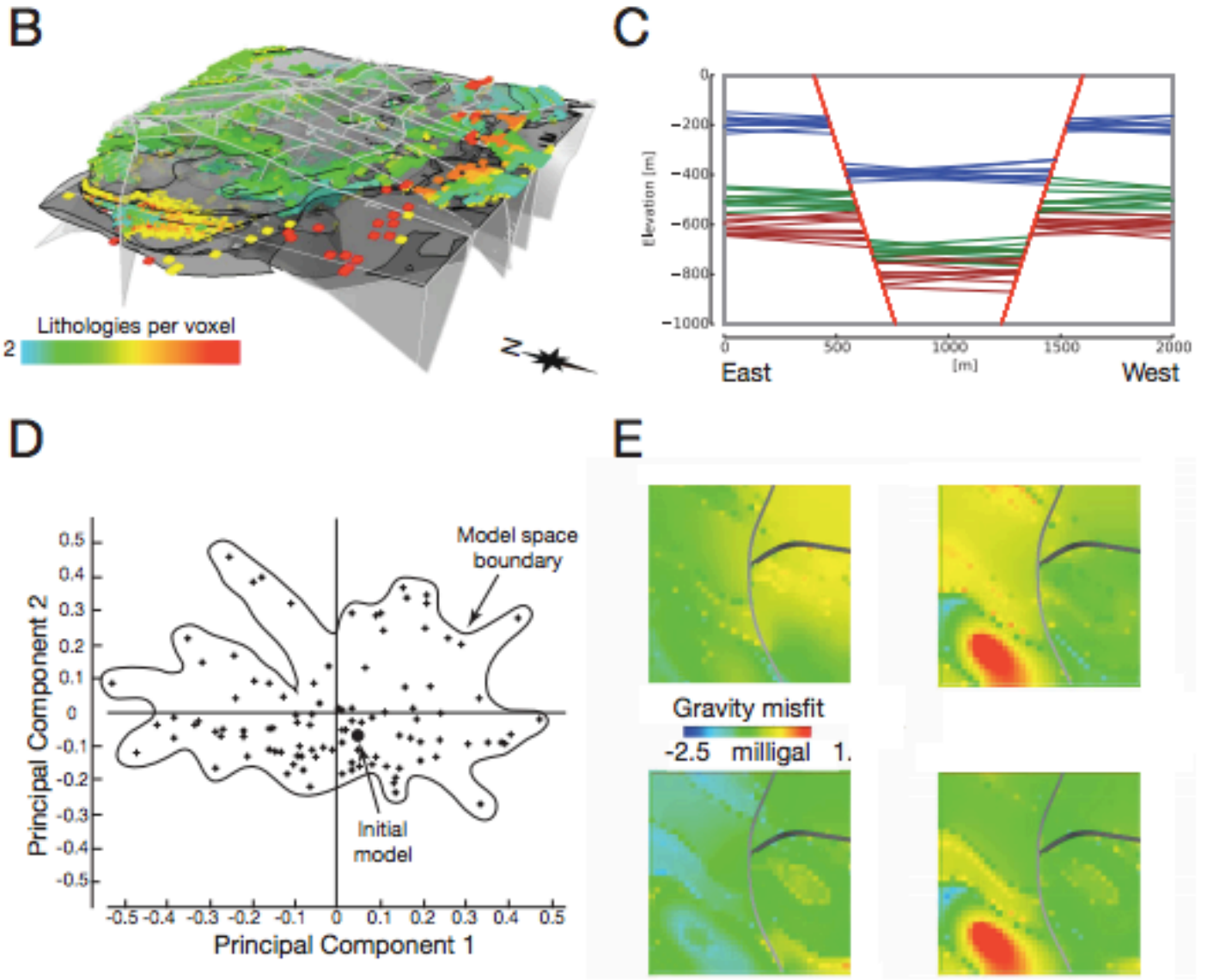
3. **Posterior:** $\pi_M(\mathbf{u}_0|u_2^{\text{obs}}) \approx \pi_M(u_2^{\text{obs}}|\mathbf{u}_0)\underbrace{\pi(\mathbf{u}_0)}_{=\text{const}}$
4. **Statistic:** $\mathbf{E}_{\pi(u_2^{\text{obs}}|\mathbf{u}_0)}[\mathscr{G}_M(\mathbf{u}_0)]$ (expected value under the posterior)

Depending on size of α^2 this leads to a vastly reduced uncertainty in expected value of $u_1(T)$. Can be computed w. Metropolis-Hastings MCMC.

Bayesian Uncertainty Quantification

Example 2: Geostatistics & Imaging

Source: Wellmann et al, 2014



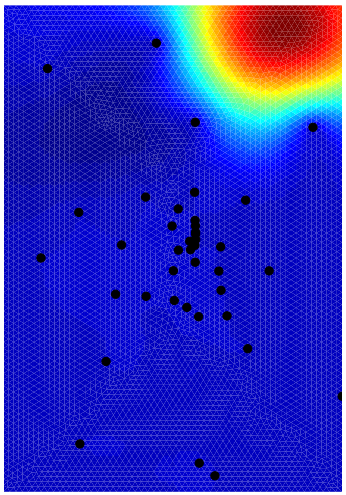
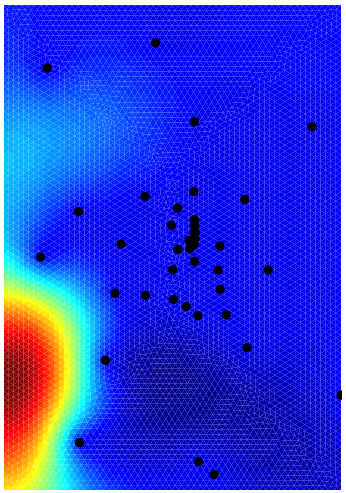
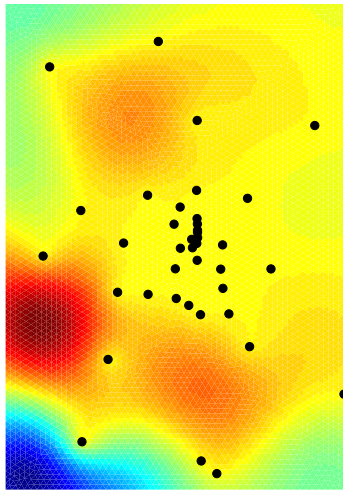
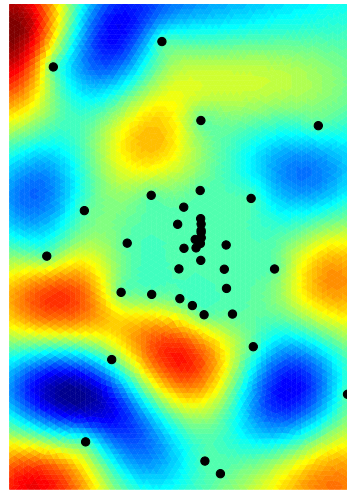
Bayesian Uncertainty Quantification

Example 3: Radioactive Waste Disposal – Prior

Source: Ernst et al, 2014

Prior model: KL modes for WIPP site conditioned on 38 tranmissivity observations (see Lecture 1)

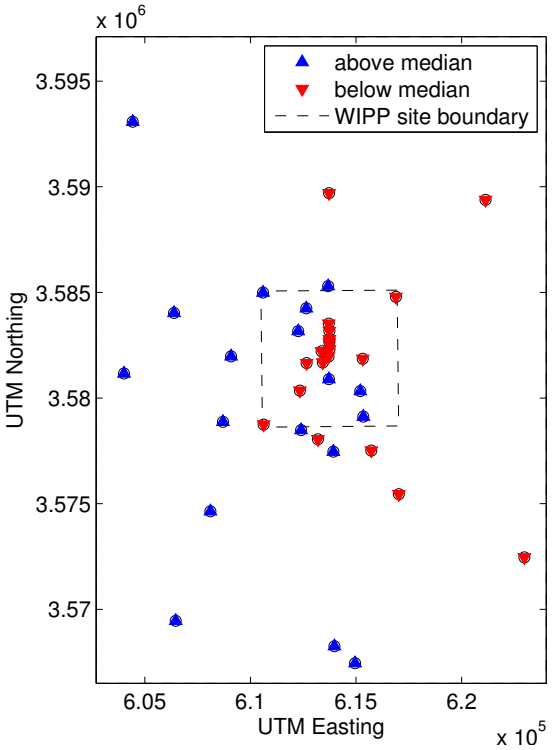
conditioned, $m = 1, 2, 9, 16$

Bayesian Uncertainty Quantification

Example 3: Radioactive Waste Disposal – WIPP Data

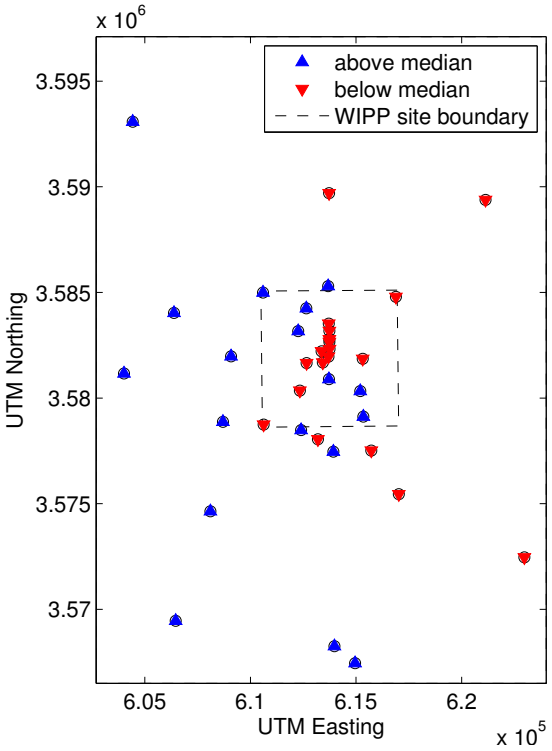
Transmissivity data



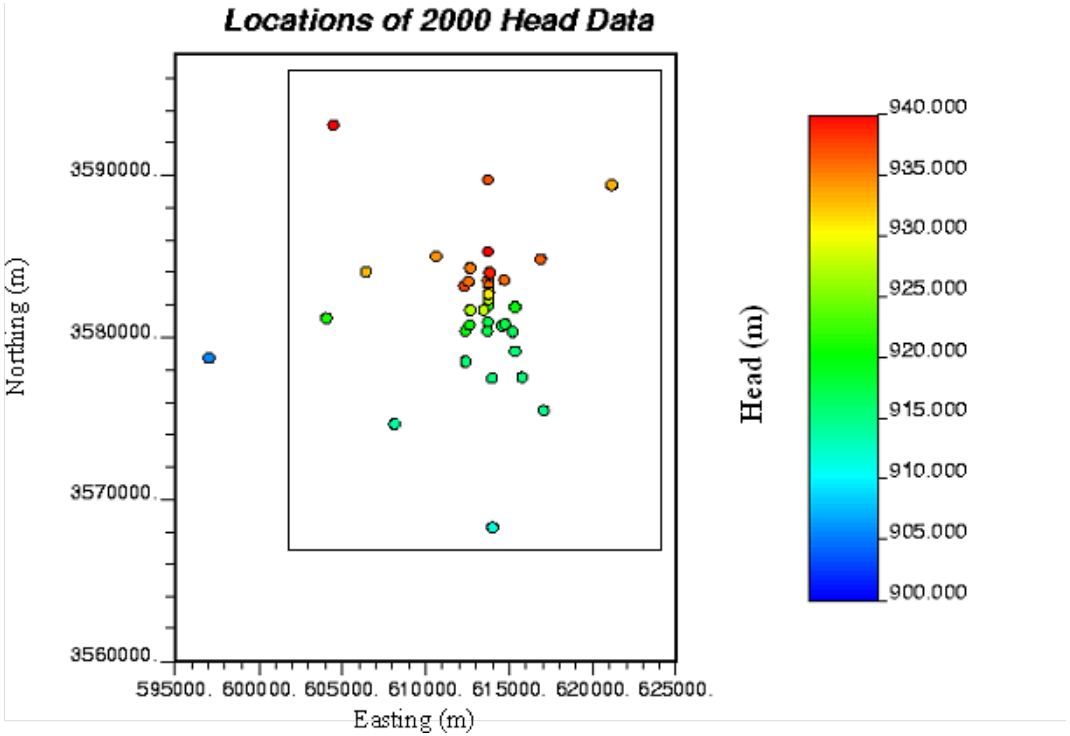
Bayesian Uncertainty Quantification

Example 3: Radioactive Waste Disposal – WIPP Data

Transmissivity data



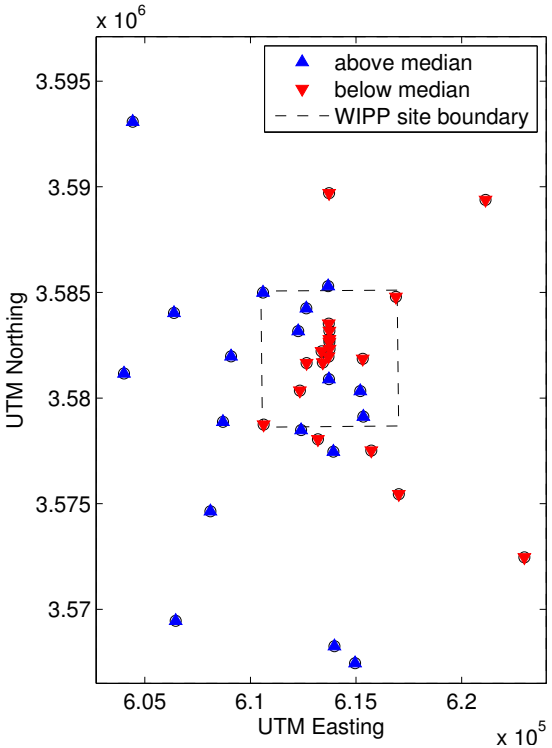
Head (=pressure) data



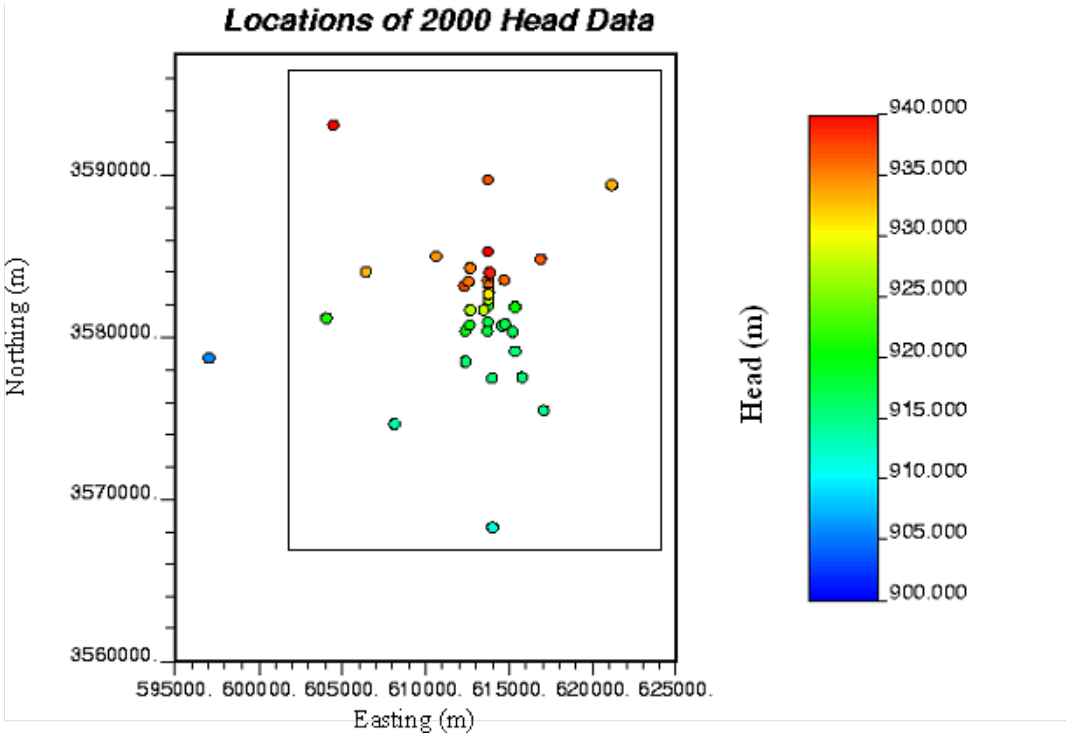
Bayesian Uncertainty Quantification

Example 3: Radioactive Waste Disposal – WIPP Data

Transmissivity data



Head (=pressure) data



Include head data via Bayesian inference using the likelihood model

$$\pi(\mathbf{h}^{\text{obs}} \mid \mathbf{Z}_s) \;\propto\; \exp(-\|\mathbf{h}^{\text{obs}} - F_h(\mathbf{Z}_s)\|^2_{Q^{-1}})$$

$F_h(\mathbf{Z}_s)$ ... model response; Q ... measurement error covariance matrix

Markov Chain Monte Carlo

For large-scale problems with bias error

- In Lecture 1 we parametrised the model by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (**prior**).
  In the subsurface flow application the lognormal coefficient is parametrised
  $$\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x)Z_j(\omega) \quad \text{and} \quad \pi_0^s(\mathbf{Z}_s) \asymp (2\pi)^{-s/2} \prod_{j=1}^{s} \exp\left(-\frac{Z_j^2}{2}\right)$$
- To fit model to **output data** F_{obs} (the **"posterior"**) use
  (e.g. pressure measurements or functionals of pressure: $F_{\text{obs}} = \mathscr{F}(p)$)

Markov Chain Monte Carlo

For large-scale problems with bias error

- In Lecture 1 we parametrised the model by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (**prior**).
  In the subsurface flow application the lognormal coefficient is parametrised
  $$\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x)Z_j(\omega) \quad \text{and} \quad \pi_0^s(\mathbf{Z}_s) \asymp (2\pi)^{-s/2} \prod_{j=1}^{s} \exp\left(-\frac{Z_j^2}{2}\right)$$
- To fit model to **output data** F_{obs} (the **"posterior"**) use
  (e.g. pressure measurements or functionals of pressure: $F_{\text{obs}} = \mathscr{F}(p)$)

Bayes' Theorem: (proportionality factor $1/\pi(F_{\text{obs}})$ expensive to compute!)

$$\underbrace{\pi^{h,s}(\mathbf{Z}_s)}_{\textbf{posterior}} := \pi(\mathbf{Z}_s \,|\, F_{\text{obs}}) \asymp \underbrace{\mathscr{L}_h(F_{\text{obs}} \,|\, \mathbf{Z}_s)}_{\textbf{likelihood}} \underbrace{\pi_0^s(\mathbf{Z}_s)}_{\textbf{prior}}$$

Markov Chain Monte Carlo

For large-scale problems with bias error

- In Lecture 1 we parametrised the model by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (**prior**).
  In the subsurface flow application the lognormal coefficient is parametrised

$$\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x)Z_j(\omega) \quad \text{and} \quad \pi_0^s(\mathbf{Z}_s) \asymp (2\pi)^{-s/2} \prod_{j=1}^{s} \exp\left(-\frac{Z_j^2}{2}\right)$$

- To fit model to **output data** F_{obs} (the **"posterior"**) use
  (e.g. pressure measurements or functionals of pressure: $F_{\text{obs}} = \mathscr{F}(p)$)

Bayes' Theorem: (proportionality factor $1/\pi(F_{\text{obs}})$ expensive to compute!)

$$\underbrace{\pi^{h,s}(\mathbf{Z}_s)}_{\textbf{posterior}} := \pi(\mathbf{Z}_s \,|\, F_{\text{obs}}) \asymp \underbrace{\mathscr{L}_h(F_{\text{obs}} \,|\, \mathbf{Z}_s)}_{\textbf{likelihood}} \underbrace{\pi_0^s(\mathbf{Z}_s)}_{\textbf{prior}}$$

- **Likelihood model** (e.g. Gaussian) also needs to be approximated:

$$\mathscr{L}_h(F_{\text{obs}} \,|\, \mathbf{Z}_s) \asymp \exp(-\|F_{\text{obs}} - F_h(\mathbf{Z}_s)\|^2/\sigma_{\text{fid}}^2)$$

$F_h(\mathbf{Z}_s)$ ... model response; σ_{fid} ... fidelity parameter (data error)

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

- Choose $\mathbf{Z}_s^0$.
- At state n generate proposal $\mathbf{Z}_s'$ from distribution $q^{\text{trans}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)$
  (e.g. preconditioned Crank-Nicholson random walk [Cotter et al, 2012])
- Accept $\mathbf{Z}_s'$ as a sample with probability

$$\alpha^{h,s}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n) \;=\; \min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')\, q^{\text{trans}}(\mathbf{Z}_s^n \,|\, \mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)\; q^{\text{trans}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)}\right)$$

  i.e. $\mathbf{Z}_s^{n+1} = \mathbf{Z}_s'$ with probability $\alpha^{h,s}$; otherwise $\mathbf{Z}_s^{n+1} = \mathbf{Z}_s^n$.

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

- Choose $\mathbf{Z}_s^0$.
- At state n generate proposal $\mathbf{Z}_s'$ from distribution $q^{\text{trans}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)$
  (e.g. preconditioned Crank-Nicholson random walk [Cotter et al, 2012])
- Accept $\mathbf{Z}_s'$ as a sample with probability

$$\alpha^{h,s}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n) \;=\; \min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')\, q^{\text{trans}}(\mathbf{Z}_s^n \,|\, \mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)\; q^{\text{trans}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)}\right)$$

  i.e. $\mathbf{Z}_s^{n+1} = \mathbf{Z}_s'$ with probability $\alpha^{h,s}$; otherwise $\mathbf{Z}_s^{n+1} = \mathbf{Z}_s^n$.

Samples $\mathbf{Z}_s^n$ used as usual for inference (even though not i.i.d.):

$$\mathbb{E}_{\pi^{h,s}}[Q] \;\approx\; \mathbb{E}_{\pi^{h,s}}[Q_{h,s}] \;\approx\; \frac{1}{N}\sum_{i=1}^{N} Q_{h,s}^{(n)} := \widehat{Q}^{\text{MetH}}$$

where $Q_{h,s}^{(n)} = \mathscr{G}\left(\mathbf{X}_h(\mathbf{Z}_s^{(n)})\right)$ is the nth sample of Q using Model(h, s).

Markov Chain Monte Carlo

Comments

Pros:

- Produces a Markov chain $\{\mathbf{Z}_s^n\}_{n\in\mathbb{N}}$, with $\mathbf{Z}_s^n \sim \pi^{h,s}$ as $n \to \infty$.
- Can be made dimension independent (e.g. via pCN sampler).
- Therefore often referred to as **"gold standard"** (Stuart et al)

Markov Chain Monte Carlo

Comments

Pros:

- Produces a Markov chain $\{\mathbf{Z}_s^n\}_{n\in\mathbb{N}}$, with $\mathbf{Z}_s^n \sim \pi^{h,s}$ as $n \to \infty$.
- Can be made dimension independent (e.g. via pCN sampler).
- Therefore often referred to as **"gold standard"** (Stuart et al)

Cons:

- Evaluation of $\alpha^{h,s} = \alpha^{h,s}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)$ **very expensive** for small h.
  (**heterogeneous** deterministic PDE: Cost/sample $\geq \mathcal{O}(M) = \mathcal{O}(h^{-d})$)
- Acceptance rate $\alpha^{h,s}$ can be very low for large s ($< 10\%$).
- Cost $= \mathcal{O}(\varepsilon^{-2-\frac{\gamma}{\alpha}})$, **but** depends on $\alpha^{h,s}$ & burn-in

Markov Chain Monte Carlo

Comments

Pros:

- Produces a Markov chain $\{\mathbf{Z}_s^n\}_{n\in\mathbb{N}}$, with $\mathbf{Z}_s^n \sim \pi^{h,s}$ as $n \to \infty$.
- Can be made dimension independent (e.g. via pCN sampler).
- Therefore often referred to as **"gold standard"** (Stuart et al)

Cons:

- Evaluation of $\alpha^{h,s} = \alpha^{h,s}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)$ **very expensive** for small h.
  (**heterogeneous** deterministic PDE: Cost/sample $\geq \mathcal{O}(M) = \mathcal{O}(h^{-d})$)
- Acceptance rate $\alpha^{h,s}$ can be very low for large s ($< 10\%$).
- Cost $= \mathcal{O}(\varepsilon^{-2-\frac{\gamma}{\alpha}})$, **but** depends on $\alpha^{h,s}$ & burn-in

Prohibitively expensive – significantly more than plain-vanilla MC!

Multilevel Markov Chain Monte Carlo

$h_{\ell-1} = m h_\ell$ and $s_\ell = s_{\ell-1}$ (for simplicity); set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

What are the **key ingredients** of "standard" multilevel Monte Carlo?

Multilevel Markov Chain Monte Carlo

$h_{\ell-1} = m h_\ell$ and $s_\ell = s_{\ell-1}$ (for simplicity); set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

What are the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell] - \mathbb{E}[Q_{\ell-1}]$
- Models with less DOFs on coarser levels **much cheaper** to solve.
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$ $\quad \Rightarrow \quad$ far **less samples** on finer levels

Multilevel Markov Chain Monte Carlo

$h_{\ell-1} = m h_\ell$ and $s_\ell = s_{\ell-1}$ (for simplicity); set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

What are the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell] - \mathbb{E}[Q_{\ell-1}]$
- Models with less DOFs on coarser levels **much cheaper** to solve.
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$ $\quad \Rightarrow \quad$ far **less samples** on finer levels

But Important! In MCMC **target distribution depends on** ℓ:

$$\mathbb{E}_{\pi^L}[Q_L] = \mathbb{E}_{\pi^0}[Q_0] + \sum_\ell \mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^{\ell-1}}[Q_{\ell-1}]$$

Multilevel Markov Chain Monte Carlo

$h_{\ell-1} = m h_\ell$ and $s_\ell = s_{\ell-1}$ (for simplicity); set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

What are the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell] - \mathbb{E}[Q_{\ell-1}]$
- Models with less DOFs on coarser levels **much cheaper** to solve.
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$ $\quad \Rightarrow \quad$ far **less samples** on finer levels

But Important! In MCMC **target distribution depends on** ℓ:

$$\mathbb{E}_{\pi^L}[Q_L] = \underbrace{\mathbb{E}_{\pi^0}[Q_0]}_{\text{standard MCMC}} + \sum_\ell \underbrace{\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^{\ell-1}}[Q_{\ell-1}]}_{\text{2 level MCMC (NEW)}}$$

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0} \sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n) \right)$$

Multilevel Markov Chain Monte Carlo

$h_{\ell-1} = m h_\ell$ and $s_\ell = s_{\ell-1}$ (for simplicity); set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

What are the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell] - \mathbb{E}[Q_{\ell-1}]$
- Models with less DOFs on coarser levels **much cheaper** to solve.
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$ $\quad \Rightarrow \quad$ far **less samples** on finer levels

But Important! In MCMC **target distribution depends on** ℓ:

$$\mathbb{E}_{\pi^L}[Q_L] = \underbrace{\mathbb{E}_{\pi^0}[Q_0]}_{\text{standard MCMC}} + \sum_\ell \underbrace{\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^{\ell-1}}[Q_{\ell-1}]}_{\text{2 level MCMC (NEW)}}$$

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0} \sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n) \right)$$

Can also reduce the number $s_{\ell-1}$ of random parameters on the coarser levels.

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

ALGORITHM 2 (Two-level Metropolis Hastings MCMC for $Q_\ell - Q_{\ell-1}$)

At states $\mathbf{z}_{\ell-1}^n, \mathbf{Z}_\ell^n$ (of two Markov chains on levels $\ell-1$ and ℓ):

1. On level $\ell-1$: Generate an **independent** sample $\mathbf{z}_{\ell-1}^{n+1} \sim \pi^{\ell-1}$ (from coarse posterior)

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

ALGORITHM 2 (Two-level Metropolis Hastings MCMC for $Q_\ell - Q_{\ell-1}$)

At states $\mathbf{z}_{\ell-1}^n, \mathbf{Z}_\ell^n$ (of two Markov chains on levels $\ell-1$ and ℓ):

1. On level $\ell-1$: Generate an **independent** sample $\mathbf{z}_{\ell-1}^{n+1} \sim \pi^{\ell-1}$ (from coarse posterior)

2. On level ℓ: Propose $\mathbf{Z}'_\ell = \mathbf{z}_{\ell-1}^{n+1}$

   Transition prob. q^{ML} depends on posterior on level $\ell-1$!

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

ALGORITHM 2 (Two-level Metropolis Hastings MCMC for $Q_\ell - Q_{\ell-1}$)

At states $\mathbf{z}_{\ell-1}^n, \mathbf{Z}_\ell^n$ (of two Markov chains on levels $\ell-1$ and ℓ):

1. On level $\ell-1$: Generate an **independent** sample $\mathbf{z}_{\ell-1}^{n+1} \sim \pi^{\ell-1}$ (from coarse posterior)
2. On level ℓ: Propose $\mathbf{Z}'_\ell = \mathbf{z}_{\ell-1}^{n+1}$

   Transition prob. q^{ML} depends on posterior on level $\ell-1$!
3. Accept $\mathbf{Z}'_\ell$ with probability

$$\alpha_{\mathrm{F}}^\ell(\mathbf{Z}'_\ell \,|\, \mathbf{Z}_\ell^n) = \min\left(1, \frac{\pi^\ell(\mathbf{Z}'_\ell)\, \mathrm{q}^{\mathrm{ML}}(\mathbf{Z}_\ell^n \,|\, \mathbf{Z}'_\ell)}{\pi^\ell(\mathbf{Z}_\ell^n)\, \mathrm{q}^{\mathrm{ML}}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}_\ell^n)}\right)$$

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

ALGORITHM 2 (Two-level Metropolis Hastings MCMC for $Q_\ell - Q_{\ell-1}$)

At states $\mathbf{z}_{\ell-1}^n, \mathbf{Z}_\ell^n$ (of two Markov chains on levels $\ell-1$ and ℓ):

1. On level $\ell-1$: Generate an **independent** sample $\mathbf{z}_{\ell-1}^{n+1} \sim \pi^{\ell-1}$ (from coarse posterior)

2. On level ℓ: Propose $\mathbf{Z}_\ell' = \mathbf{z}_{\ell-1}^{n+1}$

   Transition prob. q^{ML} depends on posterior on level $\ell-1$!

3. Accept $\mathbf{Z}_\ell'$ with probability

$$\alpha_{\mathrm{F}}^\ell(\mathbf{Z}_\ell' \,|\, \mathbf{Z}_\ell^n) = \min\left(1, \frac{\pi^\ell(\mathbf{Z}_\ell')\pi^{\ell-1}(\mathbf{Z}_\ell^n)}{\pi^\ell(\mathbf{Z}_\ell^n)\pi^{\ell-1}(\mathbf{Z}_\ell')}\right)$$

   where only $\pi^\ell(\mathbf{Z}_\ell')$ has to be computed; all other posterior distributions have already been computed.

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

- $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ is genuine **Markov chain** converging to π^ℓ
  since it is standard Metropolis-Hastings.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since in limit, samples $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are both from posterior π^ℓ.

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

- $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ is genuine **Markov chain** converging to π^ℓ
  since it is standard Metropolis-Hastings.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since in limit, samples $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are both from posterior π^ℓ.
- **But** states may differ between level ℓ and $\ell - 1$:

State $n+1$	Level $\ell - 1$	Level ℓ
accept	$\mathbf{z}_{\ell-1}^{n+1}$	$\mathbf{z}_{\ell-1}^{n+1}$
reject	$\mathbf{z}_{\ell-1}^{n+1}$	$\mathbf{Z}_\ell^n$

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

- $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ is genuine **Markov chain** converging to π^ℓ
  since it is standard Metropolis-Hastings.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since in limit, samples $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are both from posterior π^ℓ.
- **But** states may differ between level ℓ and $\ell - 1$:

State $n+1$	Level $\ell - 1$	Level ℓ
accept	$\mathbf{z}_{\ell-1}^{n+1}$	$\mathbf{z}_{\ell-1}^{n+1}$
reject	$\mathbf{z}_{\ell-1}^{n+1}$	$\mathbf{Z}_\ell^n$

In second case the variance will in general not be small, **but** this does not happen often since **acceptance probability** $\alpha_{\mathsf{F}}^\ell \overset{\ell\to\infty}{\longrightarrow} 1$ (see below).

Multilevel Markov Chain Monte Carlo

[Ketelsen, RS, Teckentrup, arXiv:1303.7343]; revision in preparation (+ Dodwell !)

- $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ is genuine **Markov chain** converging to π^ℓ
  since it is standard Metropolis-Hastings.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since in limit, samples $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are both from posterior π^ℓ.
- **But** states may differ between level ℓ and $\ell - 1$:

State $n+1$	Level $\ell - 1$	Level ℓ
accept	$\mathbf{z}_{\ell-1}^{n+1}$	$\mathbf{z}_{\ell-1}^{n+1}$
reject	$\mathbf{z}_{\ell-1}^{n+1}$	$\mathbf{Z}_\ell^n$

  In second case the variance will in general not be small, **but** this does not happen often since **acceptance probability** $\alpha_{\mathsf{F}}^\ell \overset{\ell\to\infty}{\longrightarrow} 1$ (see below).
- **Practical algorithm:** Use sub-sampling on level $\ell - 1$ to get 'independent' samples (**see below for more details**).

Complexity Theorem for Multilevel MCMC

Let $Y_\ell := Q_\ell - Q_{\ell-1}$ and assume

M1 $|\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]| \lesssim M_\ell^{-\alpha}$ (discretisation and truncation error)

M2 $\mathbb{V}_{\text{alg}}[\widehat{Y}_\ell] + \Big(\mathbb{E}_{\text{alg}}[\widehat{Y}_\ell] - \mathbb{E}_{\pi^\ell,\pi^{\ell-1}}[\widehat{Y}_\ell]\Big)^2 \lesssim \dfrac{\mathbb{V}_{\pi^\ell,\pi^{\ell-1}}[Y_\ell]}{N_\ell}$ (MCMC-err)

M3 $\mathbb{V}_{\pi^\ell,\pi^{\ell-1}}[Y_\ell] \lesssim M_{\ell-1}^{-\beta}$ (multilevel variance decay)

M4 $\text{Cost}(Y_\ell^{(n)}) \lesssim M_\ell^{\gamma}.$ (cost per sample)

Then there exist L, $\{N_\ell\}_{\ell=0}^L$ s.t. $\text{MSE} < \varepsilon^2$ and

$$\varepsilon\text{–Cost}(\widehat{Q}_L^{\text{ML}}) \lesssim \varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)} \quad (+ \text{ some log-factors})$$

(This is totally **abstract** & applies not only to our subsurface model problem!)

Recall: for standard MCMC (under same assumptions) $\text{Cost} \lesssim \varepsilon^{-2-\gamma/\alpha}$.

Verifying Key Assumption M3 for subsurface flow problem

- Proof of Assumptions **M1** and **M4** similar to i.i.d. case.
- **M2** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved so far for lognormal case!)

Verifying Key Assumption M3 for subsurface flow problem

- Proof of Assumptions **M1** and **M4** similar to i.i.d. case.
- **M2** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved so far for lognormal case!)

Key Lemma

Assume $k \in C^{0,\eta}(D)$, $\eta < \frac{1}{2}$ and F^h Fréchet diff'ble and suff'ly smooth. Then

$$\mathbb{E}_{\pi_0^\ell, \pi_0^\ell}\left[1 - \boldsymbol{\alpha}_{\mathrm{F}}^\ell(\cdot|\cdot)\right] \lesssim h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta} \quad \forall \delta > 0.$$

Verifying Key Assumption M3 for subsurface flow problem

- Proof of Assumptions **M1** and **M4** similar to i.i.d. case.
- **M2** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved so far for lognormal case!)

Key Lemma

Assume $k \in C^{0,\eta}(D)$, $\eta < \frac{1}{2}$ and F^h Fréchet diff'ble and suff'ly smooth. Then

$$\mathbb{E}_{\pi_0^\ell, \pi_0^\ell}\left[1 - \boldsymbol{\alpha}_{\mathrm{F}}^\ell(\cdot|\cdot)\right] \lesssim h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta} \quad \forall \delta > 0.$$

Theorem

Let $\mathbf{Z}_\ell^n$ and $\mathbf{z}_{\ell-1}^n$ be from Algorithm 2 and choose $s_\ell \gtrsim h_\ell^{-2}$. Then

$$\mathbb{V}_{\pi^\ell, \pi^{\ell-1}}\left[Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right] \lesssim h_{\ell-1}^{1-\delta}, \quad \text{for any } \delta > 0$$

and **M3** holds for any $\beta < 1$.

Multilevel Markov Chain Monte Carlo

Practical Method

- **Recall:**

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right)$$

Multilevel Markov Chain Monte Carlo

Practical Method

- **Recall:**

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right)$$

- **Practical Algorithm for $Q_1 - Q_0$:**
  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{rN_1}$ on level 0.

Multilevel Markov Chain Monte Carlo

Practical Method

- **Recall:**

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right)$$

- **Practical Algorithm for $Q_1 - Q_0$:**
  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{rN_1}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period r) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$

     Integrated autocorrelation ≈ 86 in example; period $r \approx 50$ sufficient

Multilevel Markov Chain Monte Carlo

Practical Method

- **Recall:**

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right)$$

- **Practical Algorithm for $Q_1 - Q_0$:**
  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{rN_1}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period r) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$

     Integrated autocorrelation ≈ 86 in example; period $r \approx 50$ sufficient
  3. Use Algorithm 2 to get chain $\{\mathbf{Z}_1^n\}_{n=1}^{N_1}$.

Multilevel Markov Chain Monte Carlo

Practical Method

- **Recall:**

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right)$$

- **Practical Algorithm for $Q_1 - Q_0$:**
  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{rN_1}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period r) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$

     Integrated autocorrelation ≈ 86 in example; period $r \approx 50$ sufficient
  3. Use Algorithm 2 to get chain $\{\mathbf{Z}_1^n\}_{n=1}^{N_1}$.

- For higher levels use subsampling and ML-accept test recursively

  Much shorter period, since $\alpha_{\mathrm{F}}^1 \approx 1$ and independent proposals !

Multilevel Markov Chain Monte Carlo

Practical Method

- **Recall:**

$$\widehat{Q}_L^{\mathrm{ML}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_0^n) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n)\right)$$

- **Practical Algorithm for $Q_1 - Q_0$:**
  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{rN_1}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period r) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$

     Integrated autocorrelation ≈ 86 in example; period $r \approx 50$ sufficient
  3. Use Algorithm 2 to get chain $\{\mathbf{Z}_1^n\}_{n=1}^{N_1}$.

- For higher levels use subsampling and ML-accept test recursively

  Much shorter period, since $\alpha_{\mathrm{F}}^1 \approx 1$ and independent proposals !

- Recursive sampling procedure reduces MCMC cost even without multilevel variance reduction (related to [Christen, Fox, 2005]

Autocorrelation, Subsampling & Bias

$D = (0,1)^2$, exponential covariance w. $\sigma^2 = 1$, $\lambda = \frac{1}{2}$, $Q = \int_0^1 k \nabla p \, dx_2$, $h_0 = \frac{1}{9}$

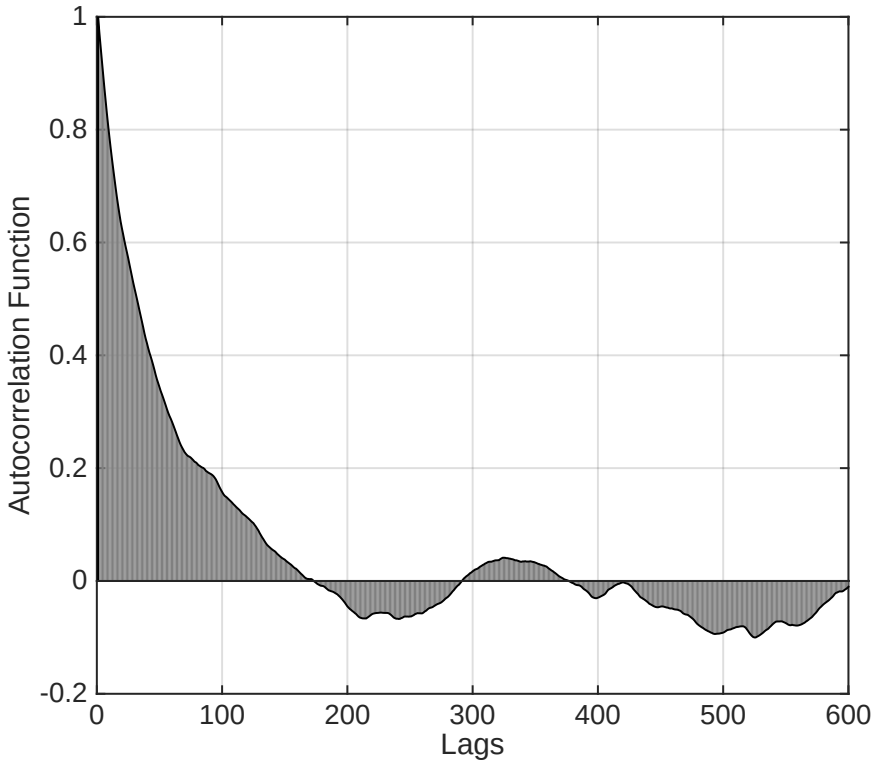
- **"Data"** F_{obs}: pressure at 16 uniformly spaced points $x_j^* \in D$.
- Two-level method with # modes: $s_0 = s_1 = 20$

Autocorrelation, Subsampling & Bias

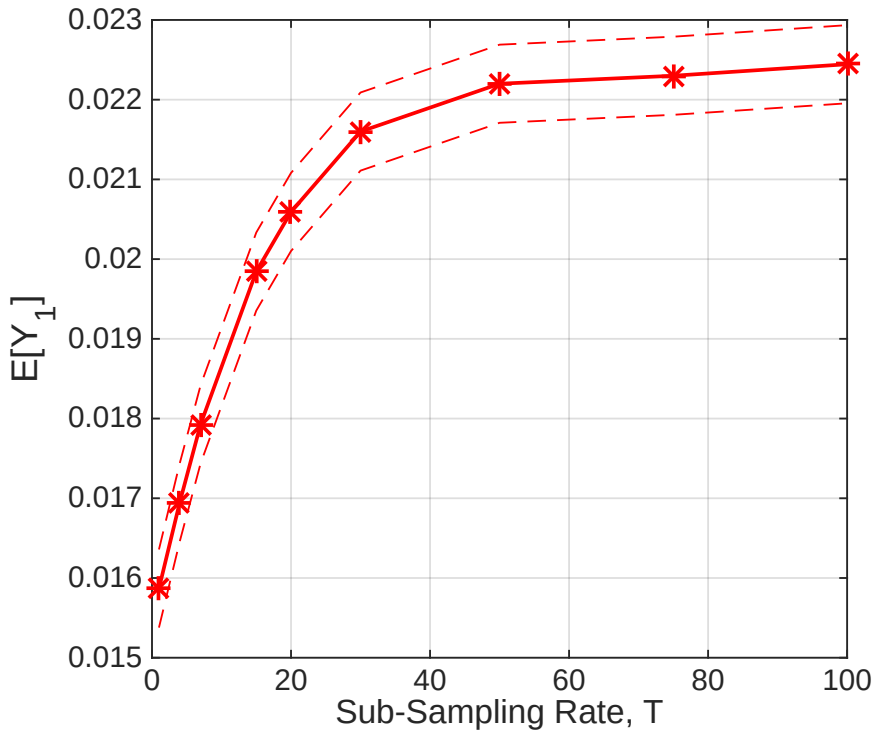
$D = (0,1)^2$, exponential covariance w. $\sigma^2 = 1$, $\lambda = \frac{1}{2}$, $Q = \int_0^1 k \nabla p \, dx_2$, $h_0 = \frac{1}{9}$

- **"Data"** F_{obs}: pressure at 16 uniformly spaced points $x_j^* \in D$.
- Two-level method with # modes: $s_0 = s_1 = 20$

Autocorrelation fct. (a.c. length ≈ 86)



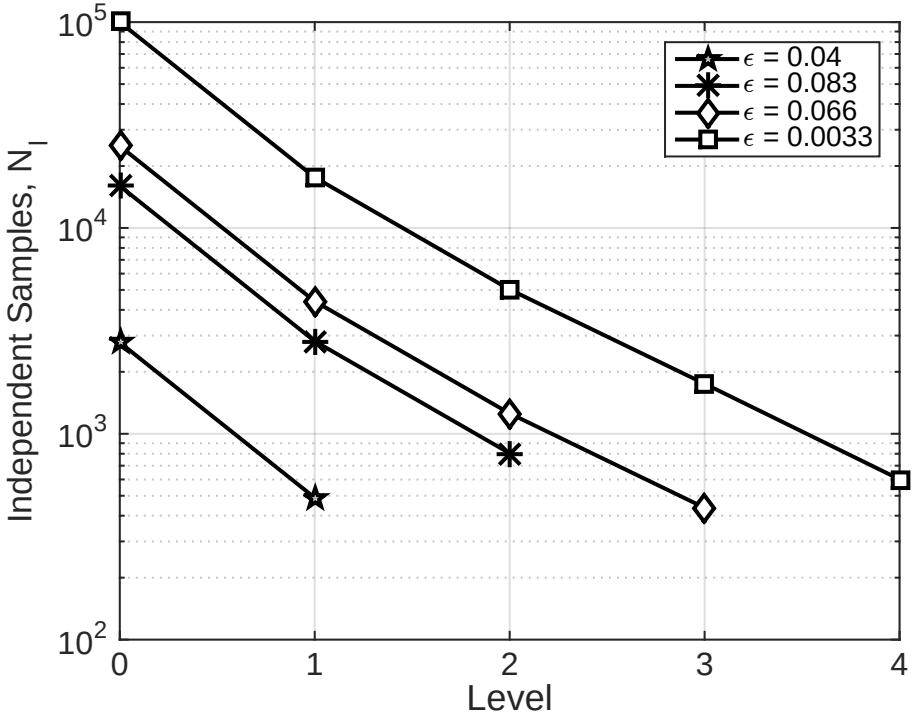
$\mathbb{E}[\widehat{Y}_1]$ w. 95% confidence interval

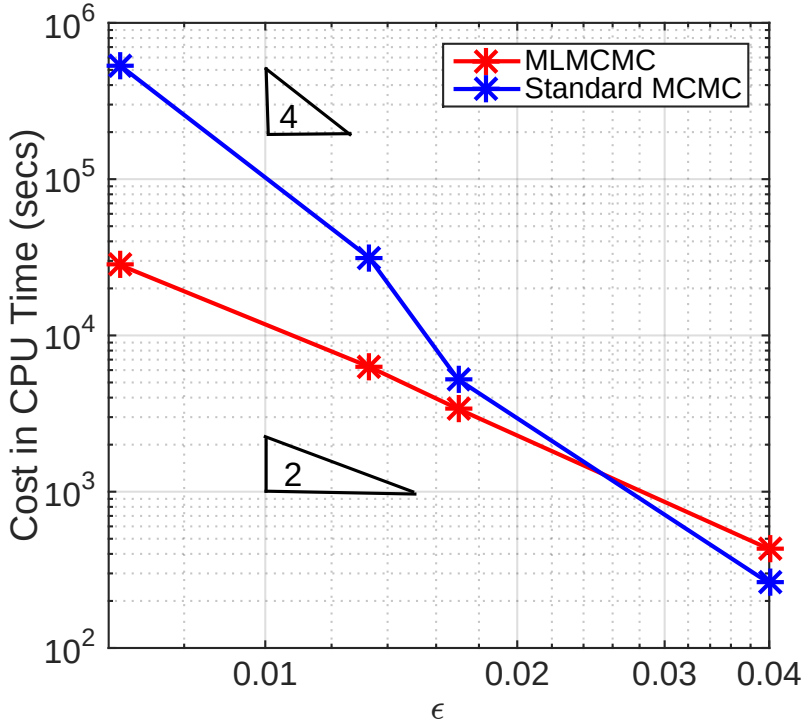


Numbers of Samples & Cost

- **"Data"** as above.
- 6-level method w. #modes increasing from $s_0 = 50$ to $s_5 = 150$

Level	0	1	2	3	4
a.c. time	136.23	3.66	2.93	1.46	1.23





Additional Comments on MLMCMC

- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]

Additional Comments on MLMCMC

- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state

Additional Comments on MLMCMC

- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state
- Reduced autocorrelation related to delayed acceptance method [Christen, Fox, 2005], [Cui, Fox, O'Sullivan, 2011]

Additional Comments on MLMCMC

- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state
- Reduced autocorrelation related to delayed acceptance method [Christen, Fox, 2005], [Cui, Fox, O'Sullivan, 2011]
- Much cheaper **multilevel burn-in** also possible
  (related to two-level work in [Efendiev, Hou, Luo, 2005])

Additional Comments on MLMCMC

- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state
- Reduced autocorrelation related to delayed acceptance method [Christen, Fox, 2005], [Cui, Fox, O'Sullivan, 2011]
- Much cheaper **multilevel burn-in** also possible
  (related to two-level work in [Efendiev, Hou, Luo, 2005])
- Related theoretical work by [Hoang, Schwab, Stuart, 2013]
  (different multilevel splitting and so far no numerics to compare)

Additional Comments on MLMCMC

- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state
- Reduced autocorrelation related to delayed acceptance method [Christen, Fox, 2005], [Cui, Fox, O'Sullivan, 2011]
- Much cheaper **multilevel burn-in** also possible
  (related to two-level work in [Efendiev, Hou, Luo, 2005])
- Related theoretical work by [Hoang, Schwab, Stuart, 2013]
  (different multilevel splitting and so far no numerics to compare)
- pCN random walk not specific; can use other proposals
  (e.g. use posterior Hessian info [Cui, Law, Marzouk, '14])

Some Other Interesting Directions/Open Questions

- Application of multilevel MCMC in other areas (statisticians!)
  other (nonlinear) PDEs, big data, geostatistics, imaging, physics
- Multilevel methods in filtering, data assimiliation, sequential MC
  [Hoel, Law, Tempone, 2015], ...
- Multilevel methods for rare events – "subset simulation"
  [Elfvreson, Hellmann, Malqvist, 2014], [Ullmann, Papaioannou, 2014], [Elfverson, RS, in prep.]

Conclusions

- I hope the course gave you a basic understanding of the questions & challenges in modern uncertainty quantification.
- The focus of the course was on the design of computationally tractable and efficient methods for high-dimensional and large-scale UQ problems in science and engineering.
- Of course it was only possible to give you a snapshot of the available methods and we went over some of them too quickly.
- Finally, I apologise that the course was of course also strongly biased in the direction of my research and my expertise and was probably not doing some other methods enough justice.
- But I hope I managed to interest you in the subject and persuade you of the huge potential of multilevel sampling methods.
- I would be very happy to discuss possible applications and projects on this subject related to your PhD projects with you.