Computational Methods in Uncertainty Quantification

Robert Scheichl

Department of Mathematical Sciences
University of Bath

Taught Course Centre Short Course
Department of Mathematical Sciences, University of Bath
Nov 19 - Dec 10 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 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

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)

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

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)

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

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

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

Let $A = U\Sigma V^T$ be the *singular value decomposition* of A with $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m)$ 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 $1/\sigma_k$.

Let $A = U\Sigma V^T$ be the *singular value decomposition* of A with $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m)$ 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?)

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.

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

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

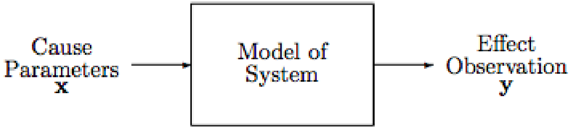
δ is called the *regularisation parameter* and controls how much we trust the data or how much we trust the a priori knowledge about x.

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

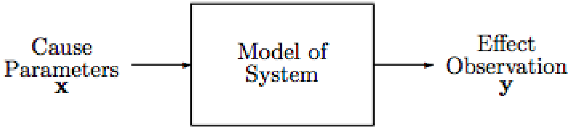
$$\underset{x\in X}{\operatorname{argmin}}\ \|y - H(x)\|^2_{Q^{-1}} + \|x - x_0\|^2_{R^{-1}}$$

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.



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**

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

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)

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

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 $\mathrm{N}(x_0, R)$.

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 $q(x' \mid x^n)$
  e.g. via a random walk: $x' \sim \mathrm{N}(x^n, \varepsilon^2 \mathrm{I})$
- Accept x' as a sample with probability

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

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.):

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

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.

- 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)

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

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)
$$\pi_M(u_2^{\mathrm{obs}}|\mathbf{u}_0) \eqsim \exp\left(\frac{-|u_2^{\mathrm{obs}} - u_{M,2}(\mathbf{u}_0)|}{\alpha^2}\right)$$
3. **Posterior:** $\pi_M(\mathbf{u}_0|u_2^{\mathrm{obs}}) \eqsim \pi_M(u_2^{\mathrm{obs}}|\mathbf{u}_0) \underbrace{\pi(\mathbf{u}_0)}_{=\mathrm{const}}$

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)
$$\pi_M(u_2^{\text{obs}}|\mathbf{u}_0) \eqsim \exp\left(\frac{-|u_2^{\text{obs}} - u_{M,2}(\mathbf{u}_0)|}{\alpha^2}\right)$$
3. **Posterior:** $\pi_M(\mathbf{u}_0|u_2^{\text{obs}}) \eqsim \pi_M(u_2^{\text{obs}}|\mathbf{u}_0)\underbrace{\pi(\mathbf{u}_0)}_{=\text{const}}$
4. **Statistic:** $\mathbf{E}_{\pi(u_2^{\text{obs}}|\mathbf{u}_0)}\left[\mathscr{G}_M(\mathbf{u}_0)\right]$ (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.

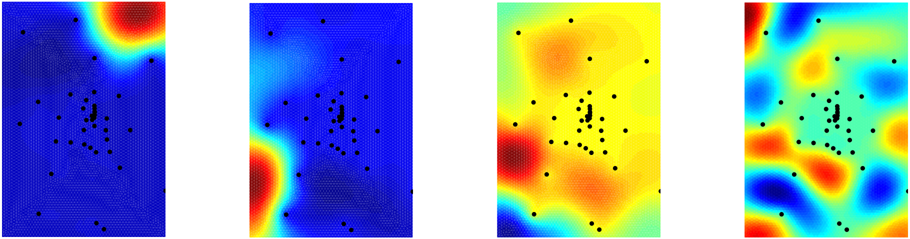
Data for Radioactive Waste Example (WIPP)

Prior and Likelihood Model [Ernst et al, 2014]

$$\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\, \phi_j^{\text{cond}}(x) Z_j(\omega) \text{ with i.i.d. } Z_j \sim \mathrm{N}(0,1)$$

KL modes ($j = 1, 2, 9, 16$) conditioned on 38 permeability observations

(low-rank change to covariance operator)



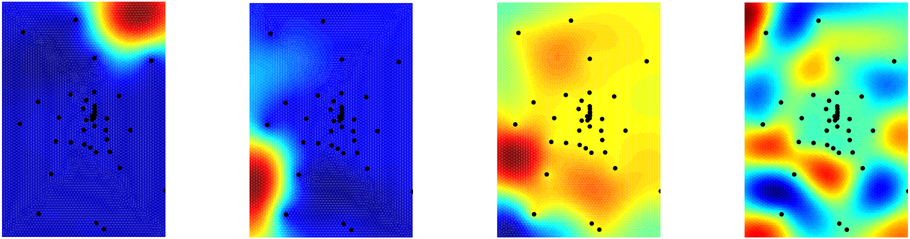
Data for Radioactive Waste Example (WIPP)

Prior and Likelihood Model [Ernst et al, 2014]

$$\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\, \phi_j^{\text{cond}}(x) Z_j(\omega) \text{ with i.i.d. } Z_j \sim \mathrm{N}(0,1)$$

KL modes ($j = 1, 2, 9, 16$) conditioned on 38 permeability observations

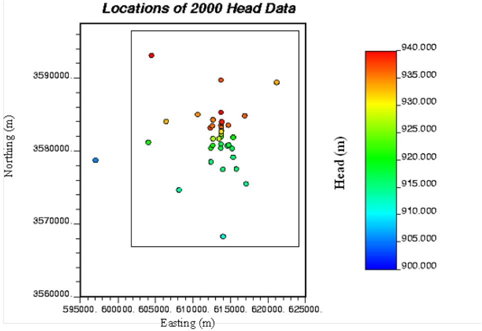
(low-rank change to covariance operator)



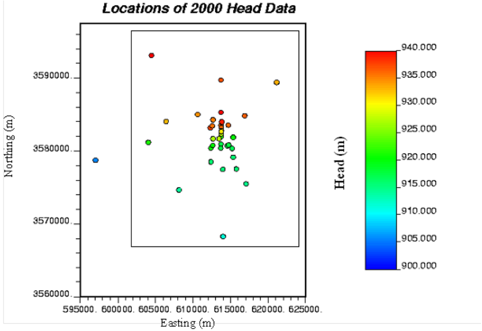
Prior model: $\pi_0^s(\mathbf{Z})$ is the multivariate Gaussian density.

Data for Radioactive Waste Example (WIPP)

Prior and Likelihood Model [Ernst et al, 2014]



- $\mathbf{y}^{\text{obs}}$ are pressure measurements.
- $F_h(\mathbf{Z})$ is the model response.



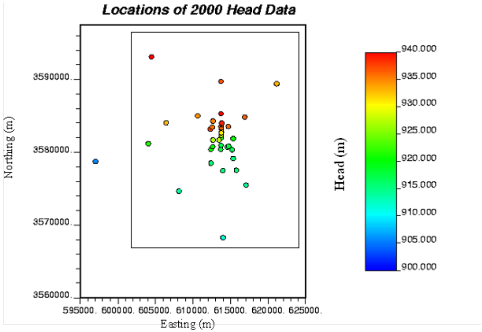
- $\mathbf{y}^{\text{obs}}$ are pressure measurements.
- $F_h(\mathbf{Z})$ is the model response.

Likelihood model: assuming Gaussian errors with covariance Σ^{obs}

$$\pi^{h,s}(\mathbf{y}^{\text{obs}}|\mathbf{Z}) \approx \exp(-\|\mathbf{y}^{\text{obs}} - F_h(\mathbf{Z})\|^2_{\Sigma^{\text{obs}}})$$

Data for Radioactive Waste Example (WIPP)

Prior and Likelihood Model [Ernst et al, 2014]



- $\mathbf{y}^{\text{obs}}$ are pressure measurements.
- $F_h(\mathbf{Z})$ is the model response.

Likelihood model: assuming Gaussian errors with covariance Σ^{obs}

$$\pi^{h,s}(\mathbf{y}^{\text{obs}}|\mathbf{Z}) \approx \exp(-\|\mathbf{y}^{\text{obs}} - F_h(\mathbf{Z})\|^2_{\Sigma^{\text{obs}}})$$

Bayes' rule: $\pi^{h,s}(\mathbf{Z} \,|\, \mathbf{y}_{\text{obs}}) \approx \pi^{h,s}(\mathbf{y}^{\text{obs}} \,|\, \mathbf{Z})\, \pi_0^s(\mathbf{Z})$

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 $Q_{h,s}^{(n)} = \mathscr{G}\big(\mathbf{X}_h(\mathbf{Z}_s^{(n)})\big)$ 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)

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

For simplicity $s_\ell = s_{\ell-1}$.

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

Multilevel Markov Chain Monte Carlo

For simplicity $s_\ell = s_{\ell-1}$.

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

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \xrightarrow{\ell\to\infty} \to 0$ as $\Longrightarrow$ much **fewer samples** on finer levels.

Multilevel Markov Chain Monte Carlo

For simplicity $s_\ell = s_{\ell-1}$.

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

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \xrightarrow{\ell\to\infty} 0$ as $\implies$ much **fewer samples** on finer levels.

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

(on level ℓ let us denote the posterior by $\pi^\ell := \pi^{h_\ell, s_\ell}(\cdot|\mathbf{y}^{\text{obs}})$)

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

Multilevel Markov Chain Monte Carlo

For simplicity $s_\ell = s_{\ell-1}$.

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

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \xrightarrow{\ell\to\infty} \to 0$ as $\Longrightarrow$ much **fewer samples** on finer levels.

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

(on level ℓ let us denote the posterior by $\pi^\ell := \pi^{h_\ell, s_\ell}(\cdot|\mathbf{y}^{\text{obs}})$)

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

$$\widehat{Q}_{h,s}^{\text{MLMetH}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_{0,0}^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{n=1}^{N_\ell} \left(Q_\ell(\mathbf{Z}_{\ell,\ell}^n) - Q_{\ell-1}(\mathbf{Z}_{\ell,\ell-1}^n)\right)$$

Multilevel Markov Chain Monte Carlo

For simplicity $s_\ell = s_{\ell-1}$.

What were the **key ingredients** of “standard” multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \overset{\ell\to\infty}{\longrightarrow} \to 0$ as $\implies$ much **fewer samples** on finer levels.

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

(on level ℓ let us denote the posterior by $\pi^\ell := \pi^{h_\ell, s_\ell}(\cdot|\mathbf{y}^{\text{obs}})$)

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

$$\widehat{Q}_{h,s}^{\text{MLMetH}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\mathbf{Z}_{0,0}^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Q_\ell(\mathbf{Z}_{\ell,\ell}^n) - Q_{\ell-1}(\mathbf{Z}_{\ell,\ell-1}^n)\right)$$

In reality, we also reduce number $s_{\ell-1}$ of random parameters on coarser levels.

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 ... 2015

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

At states $\mathbf{Z}^n_{\ell,0}, \ldots, \mathbf{Z}^n_{\ell,\ell}$ of $\ell + 1$ Markov chains on levels $0, \ldots, \ell$:

1. $k = 0$: Set $\mathbf{z}^0_0 := \mathbf{Z}^n_{\ell,0}$ and generate $T_0 := \prod_{j=0}^{\ell-1} t_j$ samples $\mathbf{z}^i_0 \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler. Choice of t_ℓ ?

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 ... 2015

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

At states $\mathbf{Z}^n_{\ell,0}, \ldots, \mathbf{Z}^n_{\ell,\ell}$ of $\ell + 1$ Markov chains on levels $0, \ldots, \ell$:

1. $k = 0$: Set $\mathbf{z}^0_0 := \mathbf{Z}^n_{\ell,0}$ and generate $T_0 := \prod_{j=0}^{\ell-1} t_j$ samples $\mathbf{z}^i_0 \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler. Choice of t_ℓ ?
2. $k > 0$: Set $\mathbf{z}^0_k := \mathbf{Z}^n_{\ell,k}$ and generate $T_k := \prod_{j=k}^{\ell-1} t_j$ samples $\mathbf{z}^i_k \sim \pi^k$:

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 . . . 2015

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

At states $\mathbf{Z}^n_{\ell,0}, \ldots, \mathbf{Z}^n_{\ell,\ell}$ of $\ell + 1$ Markov chains on levels $0, \ldots, \ell$:

1. $k = 0$: Set $\mathbf{z}^0_0 := \mathbf{Z}^n_{\ell,0}$ and generate $T_0 := \prod_{j=0}^{\ell-1} t_j$ samples $\mathbf{z}^i_0 \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler. Choice of t_ℓ ?
2. $k > 0$: Set $\mathbf{z}^0_k := \mathbf{Z}^n_{\ell,k}$ and generate $T_k := \prod_{j=k}^{\ell-1} t_j$ samples $\mathbf{z}^i_k \sim \pi^k$:
   (a) Propose $\mathbf{z}'_k = \mathbf{z}^{(i+1)t_{k-1}}_{k-1}$ with $\mathrm{q}^{\mathrm{ML}}_k(\mathbf{z}'_k | \mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$ Subsample!

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 ... 2015

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

At states $\mathbf{Z}^n_{\ell,0}, \ldots, \mathbf{Z}^n_{\ell,\ell}$ of $\ell+1$ Markov chains on levels $0, \ldots, \ell$:

1. $k = 0$: Set $\mathbf{z}^0_0 := \mathbf{Z}^n_{\ell,0}$ and generate $T_0 := \prod_{j=0}^{\ell-1} t_j$ samples $\mathbf{z}^i_0 \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler. Choice of t_ℓ ?
2. $k > 0$: Set $\mathbf{z}^0_k := \mathbf{Z}^n_{\ell,k}$ and generate $T_k := \prod_{j=k}^{\ell-1} t_j$ samples $\mathbf{z}^i_k \sim \pi^k$:
   (a) Propose $\mathbf{z}'_k = \mathbf{z}^{(i+1)t_{k-1}}_{k-1}$ with $\mathrm{q}^{\mathrm{ML}}_k(\mathbf{z}'_k|\mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$ Subsample!
   (b) Accept $\mathbf{z}'_k$ with probability

$$\alpha^{\mathrm{ML}}_\ell(\mathbf{z}'_k|\mathbf{z}^i_k) = \min\left(1, \frac{\pi^k(\mathbf{z}'_k)\,\mathrm{q}^{\mathrm{ML}}_k(\mathbf{z}^n_k|\mathbf{z}'_k)}{\pi^k(\mathbf{z}^n_k)\,\mathrm{q}^{\mathrm{ML}}(\mathbf{z}'_k\,|\,\mathbf{z}^n_k)}\right)$$

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 . . . 2015

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

At states $\mathbf{Z}^n_{\ell,0}, \dots, \mathbf{Z}^n_{\ell,\ell}$ of $\ell + 1$ Markov chains on levels $0, \dots, \ell$:

1. $k = 0$: Set $\mathbf{z}^0_0 := \mathbf{Z}^n_{\ell,0}$ and generate $T_0 := \prod_{j=0}^{\ell-1} t_j$ samples $\mathbf{z}^i_0 \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler. Choice of t_ℓ ?
2. $k > 0$: Set $\mathbf{z}^0_k := \mathbf{Z}^n_{\ell,k}$ and generate $T_k := \prod_{j=k}^{\ell-1} t_j$ samples $\mathbf{z}^i_k \sim \pi^k$:
   (a) Propose $\mathbf{z}'_k = \mathbf{z}^{(i+1)t_{k-1}}_{k-1}$ with $\mathrm{q}^{\mathrm{ML}}_k(\mathbf{z}'_k|\mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$ Subsample!

   (b) Accept $\mathbf{z}'_k$ with probability

$$\alpha^{\mathrm{ML}}_\ell(\mathbf{z}'_k|\mathbf{z}^i_k) = \min\left(1, \frac{\pi^k(\mathbf{z}'_k)\pi^{k-1}(\mathbf{z}^n_k)}{\pi^k(\mathbf{z}^n_k)\pi^{k-1}(\mathbf{z}'_k)}\right)$$

   i.e. set $\mathbf{z}^{i+1}_k = \mathbf{z}'_k$ with prob. $\alpha^{\mathrm{ML}}_\ell(\mathbf{z}'_k|\mathbf{z}^i_k)$; otherwise $\mathbf{z}^{i+1}_k = \mathbf{z}^i_k$

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 ... 2015

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

At states $\mathbf{Z}^n_{\ell,0}, \ldots, \mathbf{Z}^n_{\ell,\ell}$ of $\ell + 1$ Markov chains on levels $0, \ldots, \ell$:

1. $k = 0$: Set $\mathbf{z}^0_0 := \mathbf{Z}^n_{\ell,0}$ and generate $T_0 := \prod_{j=0}^{\ell-1} t_j$ samples $\mathbf{z}^i_0 \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler. Choice of t_ℓ ?
2. $k > 0$: Set $\mathbf{z}^0_k := \mathbf{Z}^n_{\ell,k}$ and generate $T_k := \prod_{j=k}^{\ell-1} t_j$ samples $\mathbf{z}^i_k \sim \pi^k$:
   (a) Propose $\mathbf{z}'_k = \mathbf{z}_{k-1}^{(i+1)t_{k-1}}$ with $\mathrm{q}^{\mathrm{ML}}_k(\mathbf{z}'_k|\mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$ Subsample!
   (b) Accept $\mathbf{z}'_k$ with probability
   $$\alpha^{\mathrm{ML}}_\ell(\mathbf{z}'_k|\mathbf{z}^i_k) = \min\left(1, \frac{\pi^k(\mathbf{z}'_k)\pi^{k-1}(\mathbf{z}^n_k)}{\pi^k(\mathbf{z}^n_k)\pi^{k-1}(\mathbf{z}'_k)}\right)$$
   i.e. set $\mathbf{z}^{i+1}_k = \mathbf{z}'_k$ with prob. $\alpha^{\mathrm{ML}}_\ell(\mathbf{z}'_k|\mathbf{z}^i_k)$; otherwise $\mathbf{z}^{i+1}_k = \mathbf{z}^i_k$
3. Set $\mathbf{Z}^{n+1}_{\ell,k} := \mathbf{z}^{T_k}_k$, for all $k = 0, \ldots, \ell$.

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 . . . 2015

- For sufficiently big subsampling rates t_{k-1}, we have (for $n \to \infty$) an *independence sampler* from π^{k-1}, i.e. $\mathbf{z}'_k \sim \pi^{k-1}$ independent of $\mathbf{z}^i_k$.

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 ... 2015

- For sufficiently big subsampling rates t_{k-1}, we have (for $n \to \infty$) an *independence sampler* from π^{k-1}, i.e. $\mathbf{z}'_k \sim \pi^{k-1}$ independent of $\mathbf{z}^i_k$.
- Hence, $\{\mathbf{Z}^n_{\ell,k}\}_{n \geq 1}$ is a **Markov chain** converging to π^k, $k = 0, \ldots, \ell$ (since it is just standard Metropolis-Hastings)

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 . . . 2015

- For sufficiently big subsampling rates t_{k-1}, we have (for $n \to \infty$) an *independence sampler* from π^{k-1}, i.e. $\mathbf{z}'_k \sim \pi^{k-1}$ independent of $\mathbf{z}^i_k$.
- Hence, $\{\mathbf{Z}^n_{\ell,k}\}_{n\geq 1}$ is a **Markov chain** converging to π^k, $k = 0, \ldots, \ell$
  (since it is just standard Metropolis-Hastings)
- The multilevel algorithm is **consistent** (= no bias between levels)
  since both $\{\mathbf{Z}^n_{\ell,\ell}\}_{n\geq 1}$ and $\{\mathbf{Z}^n_{\ell+1,\ell}\}_{n\geq 1}$ are samples from π^ℓ in the limit.

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, 2013 . . . 2015

- For sufficiently big subsampling rates t_{k-1}, we have (for $n \to \infty$) an *independence sampler* from π^{k-1}, i.e. $\mathbf{z}'_k \sim \pi^{k-1}$ independent of $\mathbf{z}^i_k$.
- Hence, $\{\mathbf{Z}^n_{\ell,k}\}_{n\geq 1}$ is a **Markov chain** converging to π^k, $k = 0, \ldots, \ell$ (since it is just standard Metropolis-Hastings)
- The multilevel algorithm is **consistent** (= no bias between levels) since both $\{\mathbf{Z}^n_{\ell,\ell}\}_{n\geq 1}$ and $\{\mathbf{Z}^n_{\ell+1,\ell}\}_{n\geq 1}$ are samples from π^ℓ in the limit.
- **But** states may differ between level ℓ and $\ell - 1$:

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

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

Complexity Theorem for Multilevel MCMC

Suppose there are constants $\alpha, \beta, \gamma, \eta > 0$ such that, for all $\ell = 0, \ldots, L$,

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

M2a $\mathbb{V}_{\mathrm{alg}}[\widehat{Y}_\ell] + \left(\mathbb{E}_{\mathrm{alg}}[\widehat{Y}_\ell] - \mathbb{E}_{\pi^\ell,\pi^{\ell-1}}[\widehat{Y}_\ell]\right)^2 = \mathbb{V}_{\pi^\ell,\pi^{\ell-1}}[Y_\ell]\, \mathscr{O}(N_\ell^{-1})$ (MCMC-error)

M2b $\mathbb{V}_{\pi^\ell,\pi^{\ell-1}}[Y_\ell] = \mathscr{O}(M_\ell^{-\beta})$ (multilevel variance decay)

M3 $\mathrm{Cost}(\widehat{Y}_\ell^{\mathrm{MC}}) = \mathscr{O}(N_\ell\, M_\ell^{\gamma})$. (cost per sample)

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

$$\mathscr{C}_\varepsilon(\widehat{Q}_{h,s}^{\mathrm{MLMetH}}) = \varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)} \quad (+\text{ log-factor when } \beta = \gamma)$$

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

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

FE Analysis – Verifying Assumptions M1-M3

2D lognormal diffusion problem & linear FEs

- Proof of Assumptions **M1** and **M3** similar to i.i.d. case.
- **M2a** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved for lognormal case!)

FE Analysis – Verifying Assumptions M1-M3

2D lognormal diffusion problem & linear FEs

- Proof of Assumptions **M1** and **M3** similar to i.i.d. case.
- **M2a** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved for lognormal case!)

Key Lemma for M2b (Dodwell, Ketelsen, RS, Teckentrup)

Let $\nu = 0.5$ and assume that F^h is Fréchet diff'ble and suff'ly smooth. Then

$$\mathbb{E}_{\pi^\ell,\pi^\ell}\left[1 - \alpha_\ell^{\mathrm{ML}}(\cdot|\cdot)\right] = \mathcal{O}\big(h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta}\big) \quad \forall \delta > 0.$$

FE Analysis – Verifying Assumptions M1-M3

2D lognormal diffusion problem & linear FEs

- Proof of Assumptions **M1** and **M3** similar to i.i.d. case.
- **M2a** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved for lognormal case!)

Key Lemma for M2b (Dodwell, Ketelsen, RS, Teckentrup)

Let $\nu = 0.5$ and assume that F^h is Fréchet diff'ble and suff'ly smooth. Then

$$\mathbb{E}_{\pi^\ell,\pi^\ell}\left[1 - \alpha_\ell^{\mathrm{ML}}(\cdot|\cdot)\right] = \mathcal{O}\left(h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta}\right) \quad \forall \delta > 0.$$

Theorem (Dodwell, Ketelsen, RS, Teckentrup)

Let $\{\mathbf{Z}_{\ell,\ell}^n\}_{n\geq 0}$ and $\{\mathbf{Z}_{\ell,\ell-1}^n\}_{n\geq 0}$ be from **Algorithm 2** and choose $s_\ell \gtrsim h_\ell^{-2}$. Then

$$\mathbb{V}_{\pi^\ell,\pi^{\ell-1}}\left[Q_\ell(\mathbf{Z}_{\ell,\ell}^n) - Q_{\ell-1}(\mathbf{Z}_{\ell,\ell-1}^n)\right] = \mathcal{O}\left(h_\ell^{1-\delta}\right) \quad \forall \delta > 0$$

and **M2b** holds for any $\beta < 1$. (unfortunately $\beta = \alpha$ not 2α)

Numerical Example

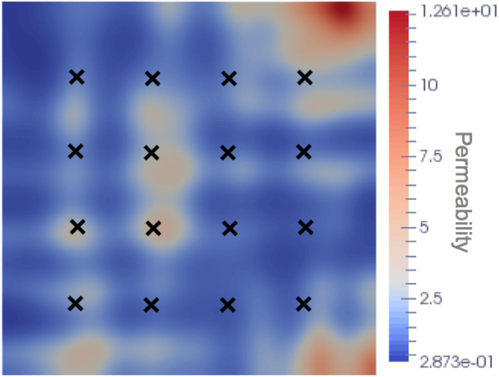
2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

- **Prior:** Separable exponential covariance with $\sigma^2 = 1$, $\lambda = 0.5$.

Numerical Example

2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

- **Prior:** Separable exponential covariance with $\sigma^2 = 1$, $\lambda = 0.5$.
- **"Data"** $\mathbf{y}^{\text{obs}}$: Pressure at 16 points $x_j^* \in D$ and $\Sigma^{\text{obs}} = 10^{-4} I$.



Numerical Example

2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

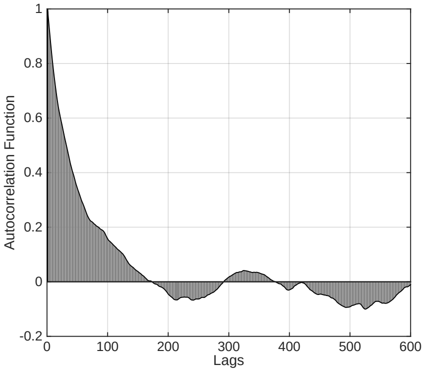
- Quantity of interest: $Q = \int_0^1 k \nabla p \, dx_2$; coarsest mesh size: $h_0 = \frac{1}{9}$
- Two-level method with #modes: $s_0 = s_1 = 20$

Numerical Example

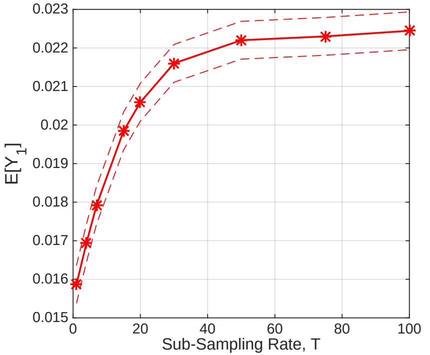
2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

- Quantity of interest: $Q = \int_0^1 k \nabla p \, dx_2$; coarsest mesh size: $h_0 = \frac{1}{9}$
- Two-level method with #modes: $s_0 = s_1 = 20$

Autocorrelation fct. (a.c. time ≈ 86)



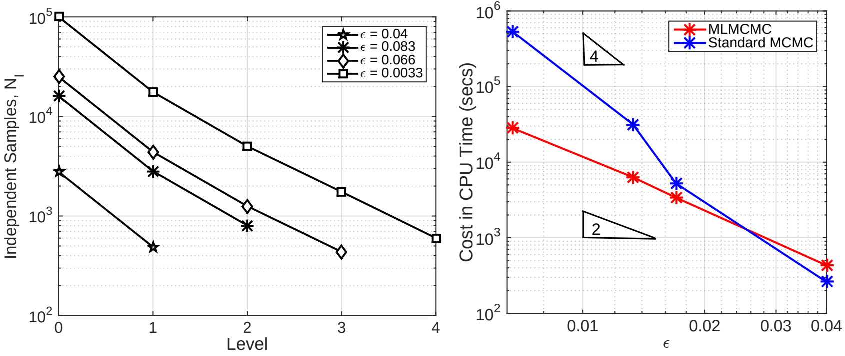
$\mathbb{E}[\widehat{Y}_1]$ w. 95% confidence interval



Numerical Example

2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

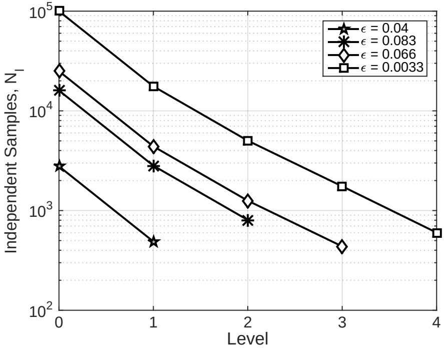
- 5-level method w. #modes increasing from $s_0 = 50$ to $s_4 = 150$

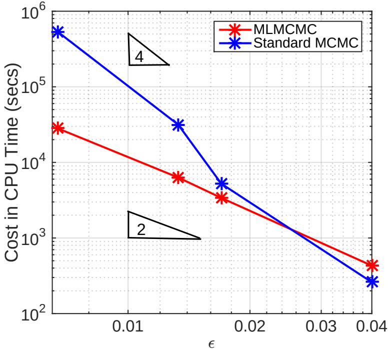


Numerical Example

2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

- 5-level method w. #modes increasing from $s_0 = 50$ to $s_4 = 150$





Level ℓ	0	1	2	3	4
a.c. time $= t_\ell$	136.23	3.66	2.93	1.46	1.23

Additional Comments on MLMCMC

- We use **multiple chains** to reduce dependence on initial state

Additional Comments on MLMCMC

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

Additional Comments on MLMCMC

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

Additional Comments on MLMCMC

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

Additional Comments on MLMCMC

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

- We use **multiple chains** to reduce dependence on initial state
- Using a special **"preconditioned" random walk** to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Reduced autocorrelation related to **delayed acceptance** method [Christen, Fox, 2005], [Cui, Fox, O'Sullivan, 2011]
- **Multilevel burn-in** also much cheaper (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 Hessian info about posterior [Cui, Law, Marzouk, '14])

Some Other Interesting Directions/Open Questions

- Application in **other areas** (especially for multilevel MCMC): other (nonlinear) PDEs, big data, geostatistics, imaging, physics [Elsakout, Christie, Lord, '15]

Some Other Interesting Directions/Open Questions

- Application in **other areas** (especially for multilevel MCMC):
  other (nonlinear) PDEs, big data, geostatistics, imaging, physics
  [Elsakout, Christie, Lord, '15]

- Multilevel **filtering**, **data assimiliation**, **sequential MC**
  [Hoel, Law, Tempone, '15], [Beskos, Jasra, Law, Tempone, Zhou, '15],
  [Gregory, Cotter, Reich, '15], [Jasra, Kamatani, Law, Zhou, '15]

Some Other Interesting Directions/Open Questions

- Application in **other areas** (especially for multilevel MCMC):
  other (nonlinear) PDEs, big data, geostatistics, imaging, physics
  [Elsakout, Christie, Lord, '15]

- Multilevel **filtering**, **data assimiliation**, **sequential MC**
  [Hoel, Law, Tempone, '15], [Beskos, Jasra, Law, Tempone, Zhou, '15], [Gregory, Cotter, Reich, '15], [Jasra, Kamatani, Law, Zhou, '15]

- Multilevel methods for **rare events** – *"subset simulation"*
  [Elfverson et al, '14], [Ullmann, Papaioannou, '14], [Elfverson, RS, in prep]

Some Other Interesting Directions/Open Questions

- Application in **other areas** (especially for multilevel MCMC):
  other (nonlinear) PDEs, big data, geostatistics, imaging, physics
  [Elsakout, Christie, Lord, '15]

- Multilevel **filtering**, **data assimiliation**, **sequential MC**
  [Hoel, Law, Tempone, '15], [Beskos, Jasra, Law, Tempone, Zhou, '15],
  [Gregory, Cotter, Reich, '15], [Jasra, Kamatani, Law, Zhou, '15]

- Multilevel methods for **rare events** – *"subset simulation"*
  [Elfverson et al, '14], [Ullmann, Papaioannou, '14], [Elfverson, RS, in prep]

- Multilevel **stochastic simulation** in systems biology, chemistry..
  [Anderson, Higham '12], [Lester, Yates, Giles, Baker '15], [Moraes et al '15]

Some Other Interesting Directions/Open Questions

- Application in **other areas** (especially for multilevel MCMC): other (nonlinear) PDEs, big data, geostatistics, imaging, physics [Elsakout, Christie, Lord, '15]
- Multilevel **filtering**, **data assimiliation**, **sequential MC** [Hoel, Law, Tempone, '15], [Beskos, Jasra, Law, Tempone, Zhou, '15], [Gregory, Cotter, Reich, '15], [Jasra, Kamatani, Law, Zhou, '15]
- Multilevel methods for **rare events** – *"subset simulation"* [Elfverson et al, '14], [Ullmann, Papaioannou, '14], [Elfverson, RS, in prep]
- Multilevel **stochastic simulation** in systems biology, chemistry.. [Anderson, Higham '12], [Lester, Yates, Giles, Baker '15], [Moraes et al '15]
- Multilevel **high-order QMC** & **adaptive** stochastic collocation

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.