Multilevel Sampling Techniques for Bayesian Inference
(Multilevel Markov Chain Monte Carlo Methods)

Rob Scheichl

Department of Mathematical Sciences

UNIVERSITY OF BATH

joint work with **T.J. Dodwell** (Exeter), **C. Ketelsen** (Boulder), **A. Stuart** & **A.L. Teckentrup** (both Warwick)

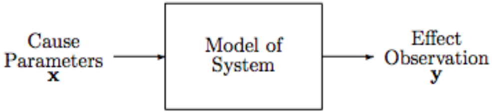
International Conference on Monte Carlo Techniques

Paris, July 8, 2016

Outline

- Bayesian inference in infinite dimensions
- Model problem: Darcy flow with uncertain data
- Multilevel Approach I: Ratio estimator (large noise case)
- Multilevel Approach II: Multilevel MCMC (small noise case)

Bayesian Interpretation of an Inverse Problem

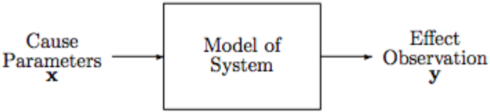


- Physical model gives $\pi(y|x)$, the conditional probability of observing y given x (**"likelihood"**), e.g. assuming additive Gaussian noise:

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

where $H : X \to \mathbb{R}^m$ is the *forward operator* & $\eta \sim \mathrm{N}(\mathbf{0}, \Sigma)$ the *noise*.

Bayesian Interpretation of an Inverse Problem



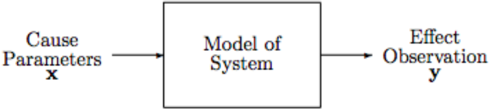
- Physical model gives $\pi(y|x)$, the conditional probability of observing y given x (**"likelihood"**), e.g. assuming additive Gaussian noise:

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

  where $H : X \to \mathbb{R}^m$ is the *forward operator* & $\eta \sim \mathrm{N}(\mathbf{0}, \Sigma)$ the *noise*.

- But often the real interest is in $\pi(x|y)$, i.e. the conditional probability of possible causes x given the observed data y (**"posterior"** density).

Bayesian Interpretation of an Inverse Problem



- Physical model gives $\pi(y|x)$, the conditional probability of observing y given x (**"likelihood"**), e.g. assuming additive Gaussian noise:

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

  where $H : X \to \mathbb{R}^m$ is the *forward operator* & $\eta \sim \mathrm{N}(\mathbf{0}, \Sigma)$ the *noise*.

- But often the real interest is in $\pi(x|y)$, i.e. the conditional probability of possible causes x given the observed data y (**"posterior"** density).

- A simple result about conditional probabilities states

$$\pi(x|y) = \frac{\pi(y|x)\pi(x)}{\pi(y)} \qquad \textbf{(Bayes' rule)}$$

  where $\pi(x) =$ **prior density** – our knowledge/belief about x

  (the scaling factor $\pi(y)$ is the **marginal** of $\pi(x, y)$ over all possible x).

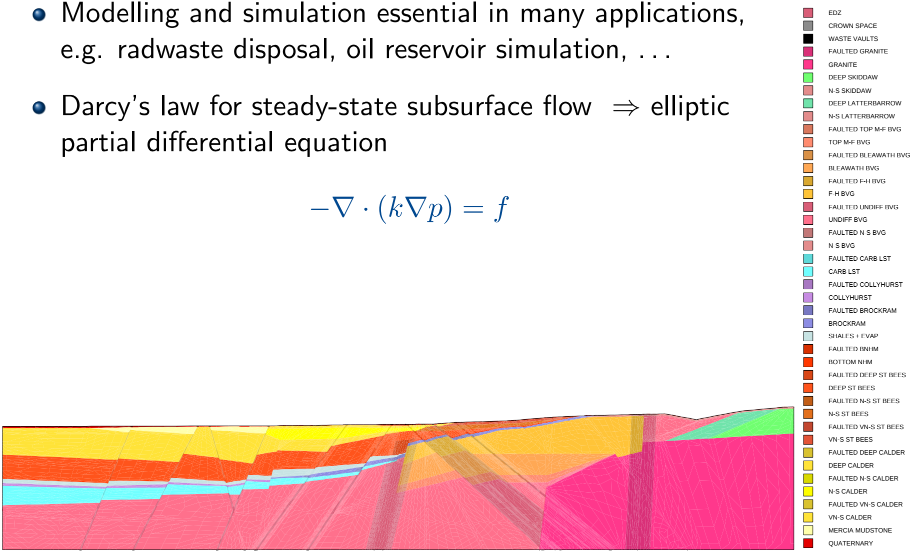
Objective

Computationally tractable and efficient algorithms for **Bayesian inference**, i.e. for computing statistics (moments, CDFs, PDFs) of certain quantities of interest with respect to the posterior measure μ^y : high (or infinite) dimensional quadrature $\mathbb{E}_{\mu^y}[\phi(x)]$ (with rigorous theoretical support).

Model Problem

- Modelling and simulation essential in many applications, e.g. radwaste disposal, oil reservoir simulation, . . .
- Darcy's law for steady-state subsurface flow $\Rightarrow$ elliptic partial differential equation

$$-\nabla \cdot (k \nabla p) = f$$

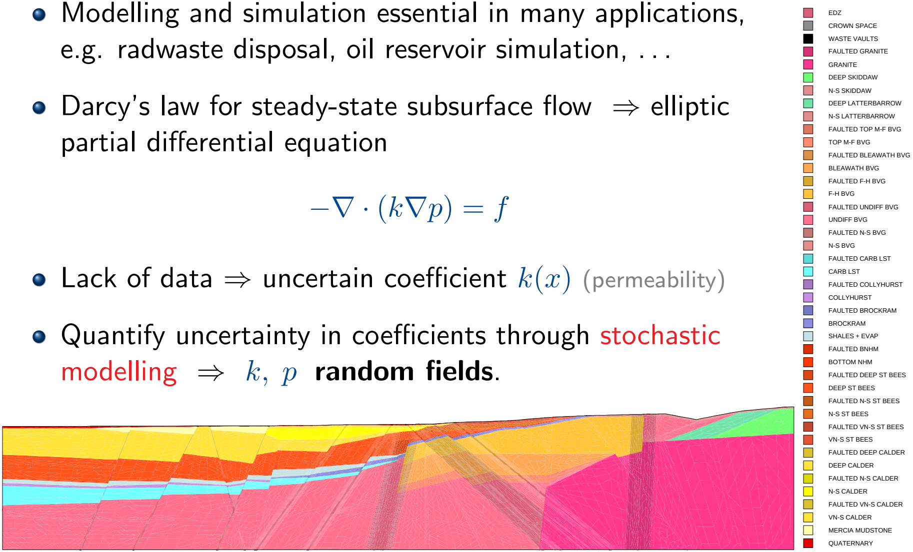


Model Problem

- Modelling and simulation essential in many applications, e.g. radwaste disposal, oil reservoir simulation, ...
- Darcy's law for steady-state subsurface flow $\Rightarrow$ elliptic partial differential equation

$$-\nabla \cdot (k \nabla p) = f$$

- Lack of data $\Rightarrow$ uncertain coefficient $k(x)$ (permeability)
- Quantify uncertainty in coefficients through stochastic modelling $\Rightarrow$ $k,\ p$ **random fields**.



Model problem

- Typical **prior model** for k is a log–normal random field:
  $k = k_* + \exp[g]$, with g a scalar, isotropic Gaussian field and

  $$\mathbb{E}[g(x)] = 0, \quad \mathbb{E}[g(x)g(y)] = \sigma^2 \exp\left(-\frac{|x-y|}{\lambda}\right) \quad \text{or} \quad \sigma^2 \exp\left(-\frac{|x-y|^2}{2\lambda^2}\right)$$

Model problem

- Typical **prior model** for k is a log–normal random field:
  $k = k_* + \exp[g]$, with g a scalar, isotropic Gaussian field and

  $$\mathbb{E}[g(x)] = 0, \quad \mathbb{E}[g(x)g(y)] = \sigma^2 \exp\left(-\frac{|x-y|}{\lambda}\right) \quad \text{or} \quad \sigma^2 \exp\left(-\frac{|x-y|^2}{2\lambda^2}\right)$$

Model problem

- Typical **prior model** for k is a log–normal random field:
  $k = k_* + \exp[g]$, with g a scalar, isotropic Gaussian field and

  $$\mathbb{E}[g(x)] = 0, \quad \mathbb{E}[g(x)g(y)] = \sigma^2 \exp\left(-\frac{|x-y|}{\lambda}\right) \text{ or } \sigma^2 \exp\left(-\frac{|x-y|^2}{2\lambda^2}\right)$$

- The quantity of interest (QoI) $\mathcal{Q}(k,p)$ and the observation operator $H(k,p)$ are some **(nonlinear) functionals** of p and k:
  - point values or local averages of the pressure p
  - point values or local averages of the Darcy flow $\vec{q} = -k\nabla p$
  - travel times of contaminant particles

Model problem

- Typical **prior model** for k is a log–normal random field:
  $k = k_* + \exp[g]$, with g a scalar, isotropic Gaussian field and

$$\mathbb{E}[g(x)] = 0, \quad \mathbb{E}[g(x)g(y)] = \sigma^2 \exp\left(-\frac{|x-y|}{\lambda}\right) \text{ or } \sigma^2 \exp\left(-\frac{|x-y|^2}{2\lambda^2}\right)$$

- The quantity of interest (QoI) $\mathcal{Q}(k,p)$ and the observation operator $H(k,p)$ are some **(nonlinear) functionals** of p and k:
  - point values or local averages of the pressure p
  - point values or local averages of the Darcy flow $\vec{q} = -k\nabla p$
  - travel times of contaminant particles

- Subsurface flow problems are typically characterised by:
  - Low spatial regularity of permeability k and resulting pressure field p
  - Unboundedness of the log–normal distribution
  - High dimensionality of the stochastic space (possibly ∞-dimensional)

Bayesian inference in infinite dimensions

Let $y \in \mathbb{R}^m$, denote by μ_0 the prior log–normal measure on k, and assume

$$y = H(p) + \eta, \quad \text{with} \quad \eta \sim \mathcal{N}(0, \sigma_\eta^2 I_m).$$

Bayesian inference in infinite dimensions

Let $y \in \mathbb{R}^m$, denote by μ_0 the prior log–normal measure on k, and assume

$$y = H(p) + \eta, \quad \text{with} \quad \eta \sim \mathcal{N}(0, \sigma_\eta^2 I_m).$$

Bayes' Theorem (e.g. [Stuart, '10])

$$\frac{d\mu^y}{d\mu_0}(k) = \frac{1}{Z} \exp\left(-\frac{|y - H(p)|^2}{2\sigma_\eta^2} \right) =: \frac{1}{Z} \exp(-\Phi(p))$$

where

$$Z := \mathbb{E}_{\mu_0}[\exp(-\Phi(p))].$$

Bayesian inference in infinite dimensions

Let $y \in \mathbb{R}^m$, denote by μ_0 the prior log–normal measure on k, and assume

$$y = H(p) + \eta, \quad \text{with } \eta \sim \mathcal{N}(0, \sigma_\eta^2 I_m).$$

Bayes' Theorem (e.g. [Stuart, '10])

$$\frac{d\mu^y}{d\mu_0}(k) = \frac{1}{Z} \exp\left(-\frac{|y - H(p)|^2}{2\sigma_\eta^2} \right) =: \frac{1}{Z} \exp(-\Phi(p))$$

where

$$Z := \mathbb{E}_{\mu_0}[\exp(-\Phi(p))].$$

We can write the **posterior expectation** of our QoI as

$$\mathbb{E}_{\mu^y}[\mathcal{Q}(p)] = \mathbb{E}_{\mu_0}\left[\frac{1}{Z} \exp[-\Phi(p)]\, \mathcal{Q}(p) \right] = \frac{\mathbb{E}_{\mu_0}[\mathcal{Q}(p) \exp[-\Phi(p)]]}{\mathbb{E}_{\mu_0}[\exp[-\Phi(p)]]},$$

i.e. the **ratio of two prior expectations**.

Ratio Estimator

Let $\psi(p) := \mathcal{Q}(p) \exp\big(-\Phi(p)\big)$. Then we can now **approximate**

$$\mathbb{E}_{\mu^y}[\mathcal{Q}(p)] \;\approx\; \frac{\widehat{Q}}{\widehat{Z}}\,,$$

where $\widehat{Q}$ is an estimator of $Q := \mathbb{E}_{\mu_0}[\psi(p)]$ and $\widehat{Z}$ is an estimator of Z.

Ratio Estimator

Let $\psi(p) := \mathcal{Q}(p) \exp\big(-\Phi(p)\big)$. Then we can now **approximate**

$$\mathbb{E}_{\mu^y}[\mathcal{Q}(p)] \approx \frac{\widehat{Q}}{\widehat{Z}},$$

where $\widehat{Q}$ is an estimator of $Q := \mathbb{E}_{\mu_0}[\psi(p)]$ and $\widehat{Z}$ is an estimator of Z.

Remark: If m is very large or σ_η^2 is very small, the two prior expectations will be difficult to evaluate. The question is how small/large?

Ratio Estimator

Let $\psi(p) := \mathcal{Q}(p) \exp\big(-\Phi(p)\big)$. Then we can now **approximate**

$$\mathbb{E}_{\mu^y}[\mathcal{Q}(p)] \approx \frac{\widehat{Q}}{\widehat{Z}},$$

where $\widehat{Q}$ is an estimator of $Q := \mathbb{E}_{\mu_0}[\psi(p)]$ and $\widehat{Z}$ is an estimator of Z.

Remark: If m is very large or σ_η^2 is very small, the two prior expectations will be difficult to evaluate. The question is how small/large?

Alternatives:

- Markov Chain Monte Carlo: Gibbs sampler, Metropolis-Hastings, ...
- **Multilevel Metropolis-Hastings** $\longrightarrow$ **Part II below**

Classical Monte Carlo (to estimate Q and Z)

- The **classical (finite element) Monte Carlo (MC)** estimator

$$\widehat{Q}_{h,N}^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N}\psi(p_h^{(i)})$$

is an equal weighted average of N i.i.d. samples $\psi(p_h^{(i)})$, where p_h denotes a finite element discretisation of p with mesh width h.

(sampling from prior k via truncated KL-expansion or circulant embedding)

Classical Monte Carlo (to estimate Q and Z)

- The **classical (finite element) Monte Carlo (MC)** estimator

$$\widehat{Q}_{h,N}^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} \psi(p_h^{(i)})$$

is an equal weighted average of N i.i.d. samples $\psi(p_h^{(i)})$, where p_h denotes a finite element discretisation of p with mesh width h.

(sampling from prior k via truncated KL-expansion or circulant embedding)

- The mean square error satisfies

$$e(\widehat{Q}_{h,N}^{\mathrm{MC}})^2 := \mathbb{E}\big[\big(\widehat{Q}_{h,N}^{\mathrm{MC}} - Q\big)^2\big] = \underbrace{\mathbb{V}[\widehat{Q}_{h,N}^{\mathrm{MC}}]}_{\text{sampling error}} + \underbrace{\big(\mathbb{E}[\widehat{Q}_{h,N}^{\mathrm{MC}}] - Q\big)^2}_{\text{discretisation error}}$$

$$\leq \mathbb{V}[\psi(p_h)]N^{-1} + Ch^s,$$

where the rate $s \in (0, 4]$ is problem dependent (ignoring sampling errors).

Quasi-Monte Carlo [Graham, Kuo, Nicholls, RS, Schwab, Sloan, 2014]

- The **Quasi-Monte Carlo (QMC)** estimator

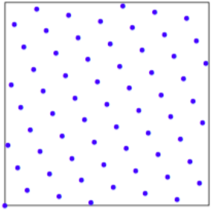
$$\widehat{Q}_{h,N}^{\mathrm{QMC}} = \frac{1}{N} \sum_{j=1}^{N} \psi(p_h^{(j)})$$

is an equal-weighted average of N deterministically chosen samples $\psi(p_h^{(j)})$, with FE soln. p_h as before, e.g. **randomised lattice points:**

- For **linear** functionals $\psi(\cdot)$ and for suff'ly smooth RFs ($\nu > d$ in Matérn), the mean square error satisfies

$$e(\widehat{Q}_{h,N}^{\mathrm{QMC}})^2 \leq C(N^{-2+\delta} + h^s), \quad \text{for any } \delta > 0,$$

where the rate $s \in (0, 4]$ is as before and C is independent of of dimension!



- Can be extended to **analytic** functionals $\psi(\cdot)$

Proof for analytic function of linear functional in [RS, Stuart, Teckentrup, 2016]

Multilevel Monte Carlo [Giles, '07], [Cliffe, Giles, RS, Teckentrup, '11]

The multilevel method works on a **hierarchy of levels**, s.t. $h_\ell = \frac{1}{2} h_{\ell-1}$, $\ell = 0, 1, \ldots, L$. The finest mesh width is $h_L = h$ (as above).

Multilevel Monte Carlo [Giles, '07], [Cliffe, Giles, RS, Teckentrup, '11]

The multilevel method works on a **hierarchy of levels**, s.t. $h_\ell = \frac{1}{2} h_{\ell-1}$, $\ell = 0, 1, \ldots, L$. The finest mesh width is $h_L = h$ (as above). Now, using

$$\mathbb{E}_{\mu_0}\left[\psi(p_{h_L})\right] = \mathbb{E}_{\mu_0}\left[\psi(p_{h_0})\right] + \sum_{\ell=1}^{L} \mathbb{E}_{\mu_0}\left[\psi(p_{h_\ell}) - \psi(p_{h_{\ell-1}})\right],$$

a **multilevel Monte Carlo (MLMC)** estimator can be defined as

$$\widehat{Q}^{\mathrm{ML}}_{\{h_\ell, N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} \psi(p^{(i)}_{h_0}) \; + \; \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \psi(p^{(i)}_{h_\ell}) - \psi(p^{(i)}_{h_{\ell-1}}),$$

i.e. a sum of $L+1$ independent MC estimators.

Multilevel Monte Carlo [Giles, '07], [Cliffe, Giles, RS, Teckentrup, '11]

The multilevel method works on a **hierarchy of levels**, s.t. $h_\ell = \frac{1}{2} h_{\ell-1}$, $\ell = 0, 1, \ldots, L$. The finest mesh width is $h_L = h$ (as above). Now, using

$$\mathbb{E}_{\mu_0}\left[\psi(p_{h_L})\right] = \mathbb{E}_{\mu_0}\left[\psi(p_{h_0})\right] + \sum_{\ell=1}^{L} \mathbb{E}_{\mu_0}\left[\psi(p_{h_\ell}) - \psi(p_{h_{\ell-1}})\right],$$

a **multilevel Monte Carlo (MLMC)** estimator can be defined as

$$\widehat{Q}^{\mathrm{ML}}_{\{h_\ell, N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} \psi(p^{(i)}_{h_0}) \; + \; \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \psi(p^{(i)}_{h_\ell}) - \psi(p^{(i)}_{h_{\ell-1}}),$$

i.e. a sum of $L + 1$ independent MC estimators.

The mean square error of the the multilevel estimator satisfies

$$e(\widehat{Q}^{\mathrm{ML}}_{\{h_\ell, N_\ell\}})^2 \le \frac{\mathbb{V}[\psi(p_{h_0})]}{N_0} \; + \; \sum_{\ell=1}^{L} \frac{\mathbb{V}[\psi(p_{h_\ell}) - \psi(p_{h_{\ell-1}})]}{N_\ell} \; + \; C h^s.$$

MLMC (lognormal) [Charrier, RS, Teckentrup, '13], [Teckentrup et al '14]

Analysis of FE error gives

$$e(\widehat{Q}^{\mathrm{ML}}_{\{h_\ell,N_\ell\}})^2 \le \frac{\mathbb{V}[\psi(p_{h_0})]}{N_0} + \sum_{\ell=1}^{L} \frac{h_\ell^s}{N_\ell} + C h^s,$$

where the rate $s \in (0,4]$ is as before.

MLMC (lognormal) [Charrier, RS, Teckentrup, '13], [Teckentrup et al '14]

Analysis of FE error gives

$$e(\widehat{Q}^{\mathrm{ML}}_{\{h_\ell,N_\ell\}})^2 \le \frac{\mathbb{V}[\psi(p_{h_0})]}{N_0} + \sum_{\ell=1}^{L} \frac{h_\ell^s}{N_\ell} + Ch^s,$$

where the rate $s \in (0,4]$ is as before.

But we can reduce the number N_ℓ of samples on the costly, fine levels:

- N_0 still needs to be large ($\approx N$ in classical MC above), **but** samples are 2^{Ld} times cheaper to obtain on coarse grid

MLMC (lognormal) [Charrier, RS, Teckentrup, '13], [Teckentrup et al '14]

Analysis of FE error gives

$$e(\widehat{Q}^{\mathrm{ML}}_{\{h_\ell,N_\ell\}})^2 \leq \frac{\mathbb{V}[\psi(p_{h_0})]}{N_0} + \sum_{\ell=1}^{L} \frac{h_\ell^s}{N_\ell} + Ch^s,$$

where the rate $s \in (0,4]$ is as before.

But we can reduce the number N_ℓ of samples on the costly, fine levels:

- N_0 still needs to be large ($\approx N$ in classical MC above), **but** samples are 2^{Ld} times cheaper to obtain on coarse grid
- $N_\ell \eqsim \left(\frac{h_\ell}{h_L}\right)^s$, for $\ell > 0$, which is much smaller than N and **in particular** $N_L = \mathcal{O}(1)$ **!**

MLMC (lognormal) [Charrier, RS, Teckentrup, '13], [Teckentrup et al '14]

Analysis of FE error gives

$$e(\widehat{Q}^{\mathrm{ML}}_{\{h_\ell,N_\ell\}})^2 \le \frac{\mathbb{V}[\psi(p_{h_0})]}{N_0} + \sum_{\ell=1}^{L} \frac{h_\ell^s}{N_\ell} + Ch^s,$$

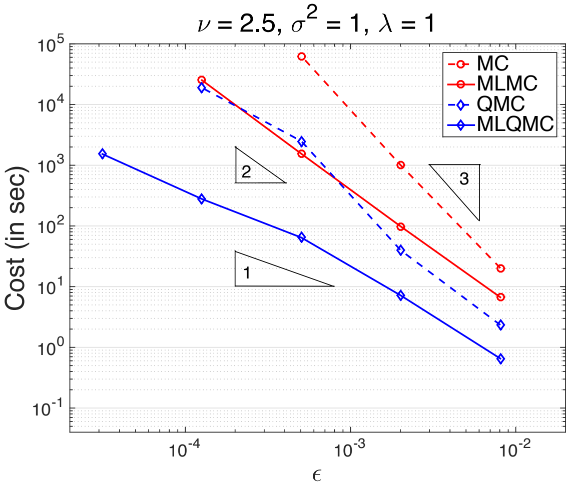
where the rate $s \in (0,4]$ is as before.

But we can reduce the number N_ℓ of samples on the costly, fine levels:

- N_0 still needs to be large ($\approx N$ in classical MC above), **but** samples are 2^{Ld} times cheaper to obtain on coarse grid
- $N_\ell \eqsim \left(\frac{h_\ell}{h_L}\right)^s$, for $\ell > 0$, which is much smaller than N and **in particular** $N_L = \mathcal{O}(1)$ **!**

Gains are complementary: Multilevel Quasi-Monte Carlo
[Kuo, RS, Schwab, Sloan, Ullmann, '15]

Numerical Comparison for lognormal problem (prior expectation)



$D = (0,1)^2$; linear FEs; $\psi(p) := \frac{1}{|D^*|}\int_{D^*} p \, dx$; truncated KLE (w. $s \sim h^{-2/\nu}$);
using a randomised lattice rule with product weights $\gamma_j = 1/j^2$.

Comments on Theory

FE error analysis and MLMC analysis

- PDE **not** uniformly elliptic or bounded.
- For $\nu < 1$ (in Matérn), **no** full regularity (typical in applications).
- Our analysis covers nonlinear functionals, point evaluations, tensors, anisotropic covariance fcts., piecewise lognormal, piecewise constant coeffs on random partitionings, . . .

Comments on Theory

FE error analysis and MLMC analysis

- PDE **not** uniformly elliptic or bounded.
- For $\nu < 1$ (in Matérn), **no** full regularity (typical in applications).
- Our analysis covers nonlinear functionals, point evaluations, tensors, anisotropic covariance fcts., piecewise lognormal, piecewise constant coeffs on random partitionings, . . .

QMC quadrature error analysis

- Bounding mixed first derivatives in stochastic parameters in weighted H^1-norm (spatial H^1-norm for QMC, L^2-norm of Laplacian for MLQMC).
- Fast CBC construction of tailored lattice rules [Kuo, Nuyens, Cools],. . .

Comments on Theory

FE error analysis and MLMC analysis

- PDE **not** uniformly elliptic or bounded.
- For $\nu < 1$ (in Matérn), **no** full regularity (typical in applications).
- Our analysis covers nonlinear functionals, point evaluations, tensors, anisotropic covariance fcts., piecewise lognormal, piecewise constant coeffs on random partitionings, ...

QMC quadrature error analysis

- Bounding mixed first derivatives in stochastic parameters in weighted H^1-norm (spatial H^1-norm for QMC, L^2-norm of Laplacian for MLQMC).
- Fast CBC construction of tailored lattice rules [Kuo, Nuyens, Cools],...
- Original analysis for linear functionals $\mathcal{G}(p)$; needs to be extended here to analytic functions $\psi : \mathbb{R} \to \mathbb{R}$ of $\mathcal{G}(p)$ (crucial for Bayesian inference).

Back to the Inverse Problem and the Ratio Estimator

- To bound the mean square error, we use

$$e\left(\frac{\widehat{Q}}{\widehat{Z}}\right)^2 = \mathbb{E}\left[\left(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\right)^2\right]$$
$$\leq \frac{2}{Z^2}\left(\mathbb{E}\big[(\widehat{Q} - Q)^2\big] + \mathbb{E}\left[\left(\frac{\widehat{Q}}{\widehat{Z}}\right)^2 (\widehat{Z} - Z)^2\right]\right).$$

Back to the Inverse Problem and the Ratio Estimator

- To bound the mean square error, we use

$$e\left(\frac{\widehat{Q}}{\widehat{Z}}\right)^2 = \mathbb{E}\left[\left(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\right)^2\right]$$

$$\leq \frac{2}{Z^2}\left(\mathbb{E}\big[(\widehat{Q} - Q)^2\big] + \mathbb{E}\left[\left(\frac{\widehat{Q}}{\widehat{Z}}\right)^2 (\widehat{Z} - Z)^2\right]\right).$$

- Further analysis depends on integrability of $\widehat{Q}/\widehat{Z}$.

Back to the Inverse Problem and the Ratio Estimator

- To bound the mean square error, we use

$$e\left(\frac{\widehat{Q}}{\widehat{Z}}\right)^2 = \mathbb{E}\left[\left(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\right)^2\right]$$
$$\leq \frac{2}{Z^2}\left(\mathbb{E}\big[(\widehat{Q}-Q)^2\big] + \mathbb{E}\left[\left(\frac{\widehat{Q}}{\widehat{Z}}\right)^2 (\widehat{Z}-Z)^2\right]\right).$$

- Further analysis depends on integrability of $\widehat{Q}/\widehat{Z}$.

- For QMC & MLMC analysis, currently require uniform ellipticity, i.e.
  - uniform prior $k(x) := k_0(x) + \sum_{j=1}^{\infty} u_j k_j(x)$ or
  - "regularised" lognormal prior $k(x) := k_*(x) + \exp(g(x))$ (with $k_* > 0$)

$$\Rightarrow \ \widehat{Q}/\widehat{Z} \in L^{\infty}_{\mathbb{P}}$$

(in the MLMC case under the additional assumption that h_0 is sufficiently small)

Convergence Rates

Theorem: [RS, Stuart, Teckentrup, 2016]

Under a uniform or a "regularised" lognormal prior ($k = k_* + \exp[g]$ with $k_* > 0$) and under suitable assumptions on H and $\mathcal{Q}$, we have

$$e\left(\frac{\widehat{Q}^{\mathrm{MC}}_{h,N}}{\widehat{Z}^{\mathrm{MC}}_{h,N}}\right)^2 \le C_{\mathrm{MC}}\left(N^{-1} + h^s\right),$$

$$e\left(\frac{\widehat{Q}^{\mathrm{ML}}_{\{h_\ell,N_\ell\}}}{\widehat{Z}^{\mathrm{ML}}_{\{h_\ell,N_\ell\}}}\right)^2 \le C_{\mathrm{ML}}\left(\sum_{\ell=0}^{L} \frac{h_\ell^s}{N_\ell} + h^s\right), \quad \text{with} \quad h = h_L$$

$$e\left(\frac{\widehat{Q}^{\mathrm{QMC}}_{h,N}}{\widehat{Z}^{\mathrm{QMC}}_{h,N}}\right)^2 \le C_{\mathrm{QMC}}\left(N^{-2+\delta} + h^s\right), \quad \text{for any } \delta > 0.$$

where the convergence rate $s \in (0, 4]$ is problem dependent.

(in the MLMC case we additional require that h_0 is sufficiently small)

Same convergence rates as for the individual estimators $\widehat{Q}$ and $\widehat{Z}$!

Complexity

The ε-cost is the number of FLOPS required to achieve a MSE of $\mathcal{O}(\varepsilon^2)$.

For the lognormal problem in d dimensions with optimal linear solver, the ε-cost converges like $\mathcal{O}(\varepsilon^{-r})$ with r given in the following tables:

d	**MLMC**	**QMC**	**MC**
1	2	?	3
2	2	?	4
3	3	?	5

exponential covariance ($s = 2$)

d	**MLMC**	**QMC**	**MC**
1	2	1.5	2.5
2	2	2	3
3	2	2.5	3.5

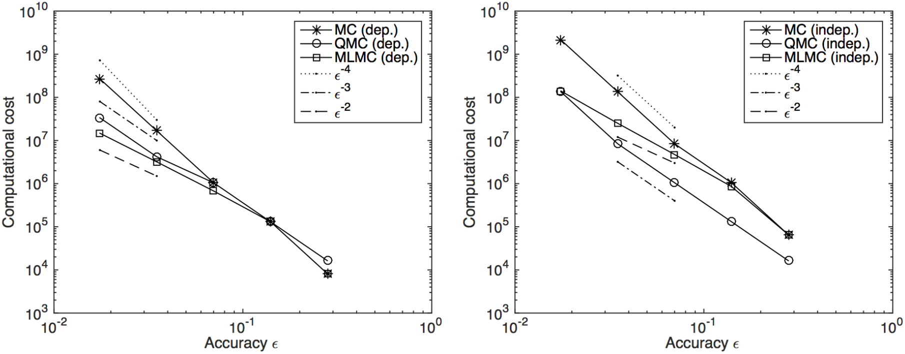
Matérn covariance w. ν suff. big ($s = 4$)

Numerical Results: Specification and Cost Comparison

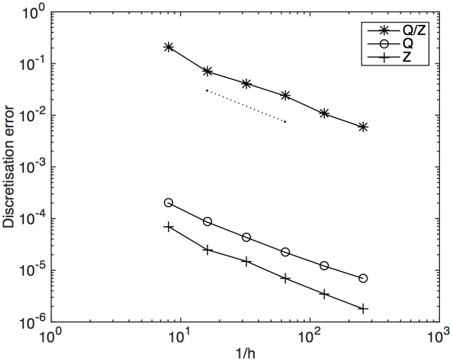
- 2D flow cell model problem on $(0,1)^2$
- prior k is lognormal with exponential covariance, $\lambda = 0.3$ and $\sigma^2 = 1$
- **Synthetic data:** local averages of pressure p at 9 points, $\sigma_\eta^2 = 0.09$
- QoI $\mathcal{Q}(p)$ is outflow over right boundary.

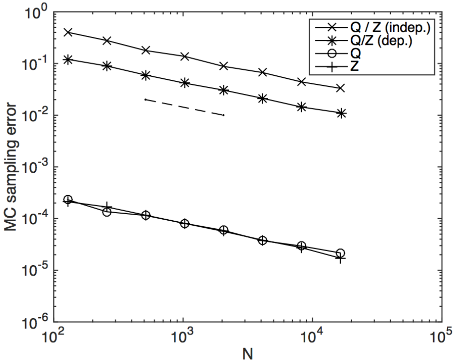
Numerical Results: Specification and Cost Comparison

- 2D flow cell model problem on $(0,1)^2$
- prior k is lognormal with exponential covariance, $\lambda = 0.3$ and $\sigma^2 = 1$
- **Synthetic data:** local averages of pressure p at 9 points, $\sigma_\eta^2 = 0.09$
- QoI $\mathcal{Q}(p)$ is outflow over right boundary.



Numerical Results: Discretisation and Sampling Errors





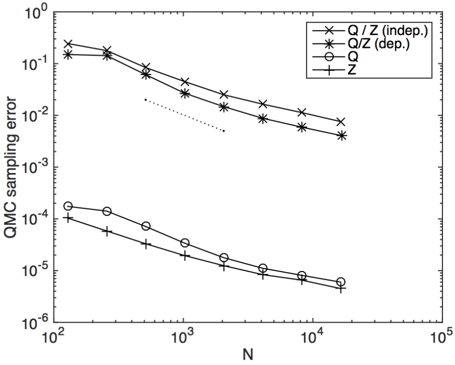
Discretisation error

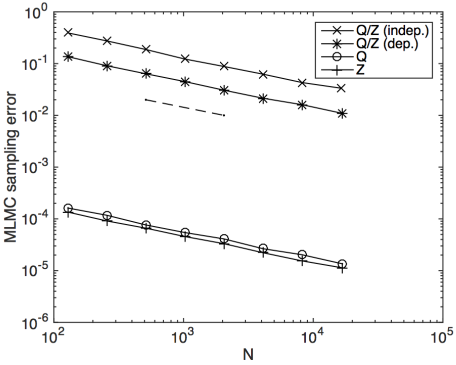
Reference slope h

Sampling error MC

Reference slope $N^{-1/2}$

Numerical Results: Discretisation and Sampling Errors





Sampling error QMC

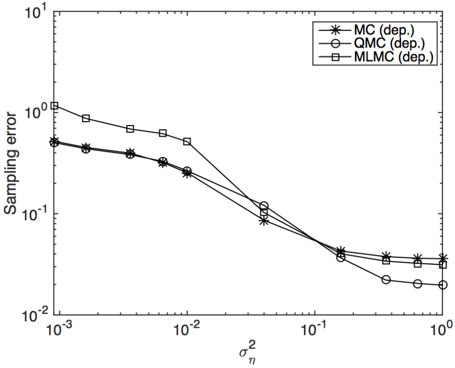
Reference slope N^{-1}

Sampling error MLMC

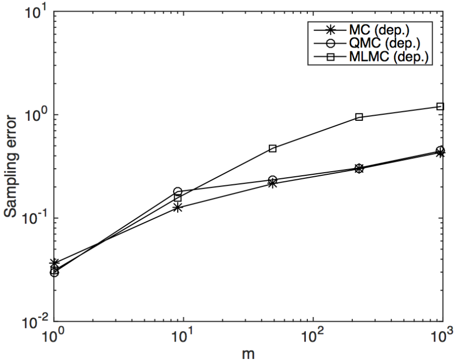
Reference slope $N^{-1/2}$

Numerical Results: Dependence on σ_η^2 and m

- Fixing $h = 1/16$ and $N = 256$.



Varying σ_η^2 (with $m = 9$)



Varying m (with $\sigma_\eta^2 = 1.0$)

Conclusions (Ratio Estimator)

- Posterior expectations can be written as ratio of prior expectations, and in this way approximated using QMC and MLMC methods.
- A convergence and complexity analysis of the resulting estimators showed that the complexity is the same as for prior expectations.
- Numerical tests confirm the effectiveness of the ratio estimator witk QMC and MLMC for a typical, simple model problem in subsurface flow (for a range of values of σ_η^2 and m) (even beyond the theory)

Conclusions (Ratio Estimator)

- Posterior expectations can be written as ratio of prior expectations, and in this way approximated using QMC and MLMC methods.
- A convergence and complexity analysis of the resulting estimators showed that the complexity is the same as for prior expectations.
- Numerical tests confirm the effectiveness of the ratio estimator witk QMC and MLMC for a typical, simple model problem in subsurface flow (for a range of values of σ_η^2 and m) (even beyond the theory)
- See also the recent paper by [Dick, Gantner, Le Gia, Schwab, '16].
- **TODO:** Comparison to other approaches, in particular MCMC and multilevel MCMC

Part II – Multilevel Metropolis-Hastings

(small noise case; avoiding calculation of normalising constant Z)

Metropolis-Hastings Markov Chain Monte Carlo

Recall (no need to know normalising constant $\pi(y)$)

$$\pi(x|y) \eqsim \pi(y|x)\pi(x) \qquad \textbf{(Bayes' rule)}$$

ALGORITHM 1 (Metropolis-Hastings Markov Chain Monte Carlo)

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

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

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

Metropolis-Hastings Markov Chain Monte Carlo

Recall (no need to know normalising constant $\pi(y)$)

$$\pi(x|y) \eqsim \pi(y|x)\pi(x) \qquad \textbf{(Bayes' rule)}$$

ALGORITHM 1 (Metropolis-Hastings Markov Chain Monte Carlo)

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

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

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

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.):

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

Application to the Model Problem

Using **truncated KL-expansion**

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

and assuming **additive Gaussian data noise** with covariance Σ^{obs}

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

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

Apply **Metropolis-Hastings MCMC** (Algorithm 1):

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

Application to the Model Problem

Using **truncated KL-expansion**

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

and assuming **additive Gaussian data noise** with covariance Σ^{obs}

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

Likelihood model: $\pi^{h,s}(\mathbf{y}^{\mathrm{obs}}|\mathbf{Z}) \eqsim \exp(-\|\mathbf{y}^{\mathrm{obs}} - F_h(\mathbf{Z})\|^2_{\Sigma^{\mathsf{obs}}})$

Apply **Metropolis-Hastings MCMC** (Algorithm 1):

$$\widehat{Q}^{\mathsf{MetH}}_{h,s} := \frac{1}{N} \sum_{n=1}^{N} \mathcal{Q}_{h,s}(\mathbf{Z}^n) \approx \mathbb{E}_{\pi^{h,s}}\left[\mathcal{Q}_{h,s}\right] \approx \mathbb{E}_{\pi^\infty}\left[\mathcal{Q}\right]$$

Pros:

- Markov chain $\mathbf{Z}^n \sim \pi^{h,s}$ as $n \to \infty$
  $\Rightarrow$ **"gold standard"** [Stuart et al]
- s-independent, e.g. via pCN sampler
  [Cotter, Dashti, Stuart, 2012]

Application to the Model Problem

Using **truncated KL-expansion**

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

and assuming **additive Gaussian data noise** with covariance Σ^{obs}

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

Likelihood model: $\pi^{h,s}(\mathbf{y}^{\text{obs}}|\mathbf{Z}) \eqsim \exp(-\|\mathbf{y}^{\text{obs}} - F_h(\mathbf{Z})\|^2_{\Sigma^{\mathsf{obs}}})$

Apply **Metropolis-Hastings MCMC** (Algorithm 1):

$$\widehat{Q}_{h,s}^{\mathsf{MetH}} := \frac{1}{N}\sum_{n=1}^{N} \mathcal{Q}_{h,s}(\mathbf{Z}^n) \approx \mathbb{E}_{\pi^{h,s}}[\mathcal{Q}_{h,s}] \approx \mathbb{E}_{\pi^\infty}[\mathcal{Q}]$$

Pros:

- Markov chain $\mathbf{Z}^n \sim \pi^{h,s}$ as $n \to \infty$
  $\Rightarrow$ **"gold standard"** [Stuart et al]
- s-independent, e.g. via pCN sampler [Cotter, Dashti, Stuart, 2012]

Cons:

- $\boldsymbol{\alpha}_{h,s}$ v. expensive for $h \ll 1$.
- $\boldsymbol{\alpha}_{h,s} < 10\%$ for large s.
- Same rate for ε-cost as standard MC **but much bigger** constant **!**

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:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \xrightarrow{\ell\to\infty} 0$ as $\implies$ 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:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \xrightarrow{\ell\to\infty} 0$ 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 $\pi^\ell := \pi^{h_\ell, s_\ell}(\cdot|\mathbf{y}^{\text{obs}})$)

$$\mathbb{E}_{\pi^L}[\mathcal{Q}_L] = \mathbb{E}_{\pi^0}[\mathcal{Q}_0] + \sum_\ell \mathbb{E}_{\pi^\ell}[\mathcal{Q}_\ell] - \mathbb{E}_{\pi^{\ell-1}}[\mathcal{Q}_{\ell-1}]$$

Multilevel Markov Chain Monte Carlo

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

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

- **Telescoping sum:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \xrightarrow{\ell\to\infty} 0$ 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 $\pi^\ell := \pi^{h_\ell, s_\ell}(\cdot|\mathbf{y}^{\text{obs}})$)

$$\mathbb{E}_{\pi^L}[\mathcal{Q}_L] = \underbrace{\mathbb{E}_{\pi^0}[\mathcal{Q}_0]}_{\text{standard MCMC}} + \sum_\ell \underbrace{\mathbb{E}_{\pi^\ell}[\mathcal{Q}_\ell] - \mathbb{E}_{\pi^{\ell-1}}[\mathcal{Q}_{\ell-1}]}_{\text{multilevel MCMC (NEW)}}$$

$$\widehat{Q}_{h,s}^{\text{MLMetH}} := \frac{1}{N_0}\sum_{n=1}^{N_0} \mathcal{Q}_0(\mathbf{Z}_{0,0}^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(\mathcal{Q}_\ell(\mathbf{Z}_{\ell,\ell}^n) - \mathcal{Q}_{\ell-1}(\mathbf{Z}_{\ell,\ell-1}^n)\right)$$

Multilevel Markov Chain Monte Carlo

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

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

- **Telescoping sum:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \xrightarrow{\ell\to\infty} 0$ 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 $\pi^\ell := \pi^{h_\ell, s_\ell}(\cdot|\mathbf{y}^{\text{obs}})$)

$$\mathbb{E}_{\pi^L}[\mathcal{Q}_L] = \underbrace{\mathbb{E}_{\pi^0}[\mathcal{Q}_0]}_{\text{standard MCMC}} + \sum_\ell \underbrace{\mathbb{E}_{\pi^\ell}[\mathcal{Q}_\ell] - \mathbb{E}_{\pi^{\ell-1}}[\mathcal{Q}_{\ell-1}]}_{\text{multilevel MCMC (NEW)}}$$

$$\widehat{Q}_{h,s}^{\text{MLMetH}} := \frac{1}{N_0}\sum_{n=1}^{N_0} \mathcal{Q}_0(\mathbf{Z}_{0,0}^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{n=1}^{N_\ell} \left(\mathcal{Q}_\ell(\mathbf{Z}_{\ell,\ell}^n) - \mathcal{Q}_{\ell-1}(\mathbf{Z}_{\ell,\ell-1}^n)\right)$$

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

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

- **Basic idea:**

$\longrightarrow$ Sketch on blackboard!

- Starting from $\ell = 0$, use i.i.d. samples from posterior π^ℓ as **proposals** for Metropolis-Hastings on level $\ell + 1$.

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

- **Basic idea:**

  $\longrightarrow$ Sketch on blackboard!

- Starting from $\ell = 0$, use i.i.d. samples from posterior π^ℓ as **proposals** for Metropolis-Hastings on level $\ell + 1$.
- How to get i.i.d. samples from π^ℓ? Use **MCMC** and subsampling.

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

- **Basic idea:**

  $\longrightarrow$ Sketch on blackboard!

- Starting from $\ell = 0$, use i.i.d. samples from posterior π^ℓ as **proposals** for Metropolis-Hastings on level $\ell + 1$.
- How to get i.i.d. samples from π^ℓ? Use **MCMC** and subsampling.
- Acceptance $\boldsymbol{\alpha}_{\ell+1}^{\mathsf{ML}} \stackrel{\ell \to \infty}{\longrightarrow} 1$ (since $\pi_\ell \approx \pi_{\ell+1}$):
  - Less and less subsampling necessary on higher levels.
  - **Strong "coupling"** between proposed sample from π^ℓ and next state on level $\ell + 1$ $\Rightarrow$ multilevel variance reduction!

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

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

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, JUQ 2015

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

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 $\mathrm{q}^{\mathsf{ML}}_k(\mathbf{z}'_k | \mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$

Subsample!

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

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

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 $\mathrm{q}^{\mathsf{ML}}_k(\mathbf{z}'_k|\mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$

   Subsample!

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

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

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

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

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

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 $\mathbf{z}_0^i \sim \pi^0$ (coarsest posterior) via **Algorithm 1** with pCN sampler.

   Choice of t_ℓ ?

2. $k>0$: Set $\mathbf{z}_k^0 := \mathbf{Z}^n_{\ell,k}$ and generate $T_k := \prod_{j=k}^{\ell-1} t_j$ samples $\mathbf{z}_k^i \sim \pi^k$:
   (a) Propose $\mathbf{z}_k' = \mathbf{z}_{k-1}^{(i+1)t_{k-1}}$ with $\mathrm{q}_k^{\mathrm{ML}}(\mathbf{z}_k'|\mathbf{z}_k^i) = \pi^{k-1}(\mathbf{z}_k')$

   Subsample!

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

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

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

Multilevel Markov Chain Monte Carlo

Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

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

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 $\mathrm{q}^{\mathsf{ML}}_k(\mathbf{z}'_k|\mathbf{z}^i_k) = \pi^{k-1}(\mathbf{z}'_k)$

     Subsample!

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

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

   i.e. set $\mathbf{z}^{i+1}_k = \mathbf{z}'_k$ with prob. $\boldsymbol{\alpha}^{\mathsf{ML}}_\ell(\mathbf{z}'_k|\mathbf{z}^i_k)$; 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, JUQ 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, JUQ 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, JUQ 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, JUQ 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** $\alpha^{\text{ML}}_\ell \xrightarrow{\ell \to \infty} 1$ (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 $|\mathbb{E}_{\pi^\ell}[\mathcal{Q}_\ell] - \mathbb{E}_{\pi^\infty}[\mathcal{Q}]| = \mathcal{O}(M_\ell^{-\alpha})$ (discretisation and truncation error)

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

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

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

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

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

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

For standard MCMC (under the 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.
  (but crucially, **two** bias errors from posterior and functional approximation!)
- **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. (but crucially, **two** bias errors from posterior and functional approximation!)
- **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

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

FE Analysis – Verifying Assumptions M1-M3

2D lognormal diffusion problem & linear FEs

- Proof of Assumptions **M1** and **M3** similar to i.i.d. case. (but crucially, **two** bias errors from posterior and functional approximation!)
- **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

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

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

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

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

Numerical Example

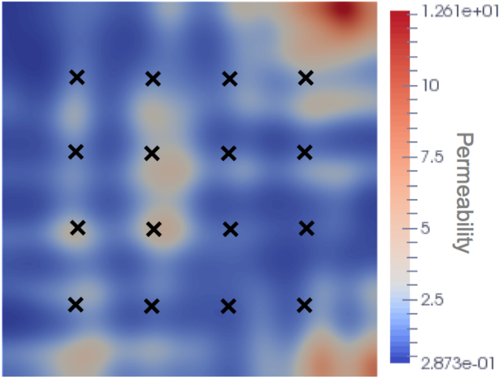
2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

- **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

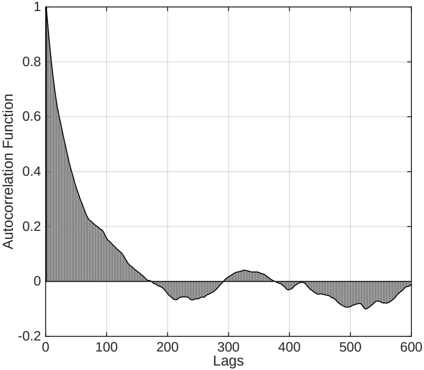
- Quantity of interest: $\mathcal{Q} = \int_0^1 k\nabla p \, dx_2$; coarsest mesh size: $h_0 = \frac{1}{9}$
- Two-level method with #modes: $s_0 = s_1 = 20$

Numerical Example

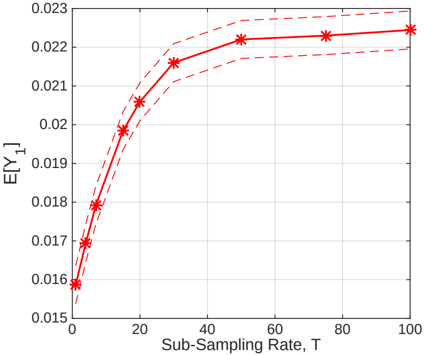
2D lognormal diffusion problem on $D = (0,1)^2$ with linear FEs

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

Autocorrelation fct. (a.c. time ≈ 86)



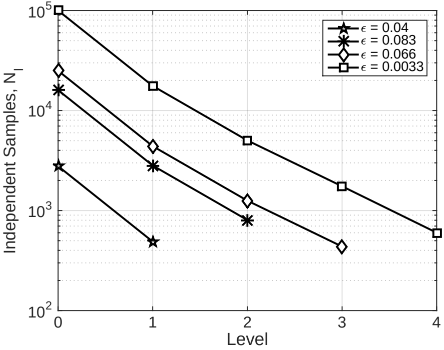
$\mathbb{E}[\widehat{Y}_1]$ w. 95% confidence interval

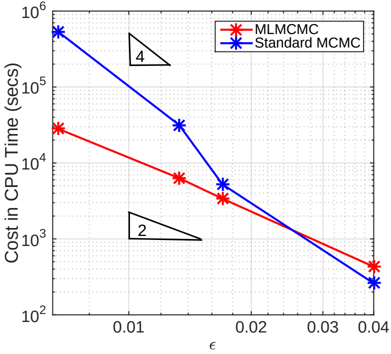


Numerical Example

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

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

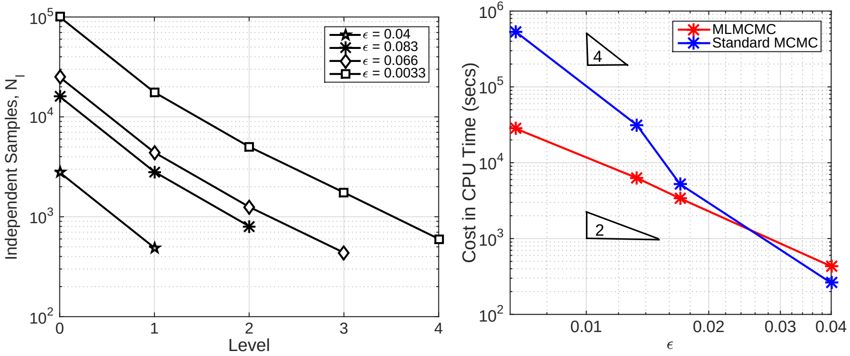




Numerical Example

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
- 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], [Ernst, Sprungk])

Conclusions, Other Directions & Open Questions

Conclusions: In my opinion multilevel methods have **huge potential** for Bayesian Inference, especially for **large-scale PDE-constrained** problems. **Many** interesting open questions (theoretical and practical)!

Conclusions, Other Directions & Open Questions

Conclusions: In my opinion multilevel methods have **huge potential** for Bayesian Inference, especially for **large-scale PDE-constrained** problems. **Many** interesting open questions (theoretical and practical)!

- Application in **other areas** (especially for multilevel MCMC):
  other (nonlinear) PDEs, big data, geostatistics, imaging, physics
  [Elsakout, Christie, Lord, '15]
- Multilevel methods in **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, Thur 16:40]
- Apply **information geometry** ideas (gradient, Hessian) for better proposals
  e.g. [Girolami, Calderhead, '11], [Cui, Law, Marzouk, '16], [Rudolf, Sprungk, '15]
- (Multilevel) **high-order QMC** [Dick, Gantner, Le Gia, Schwab, '16], ...