Novel Monte Carlo Methods and Uncertainty Quantification

(Lecture II)

Robert Scheichl

Department of Mathematical Sciences
University of Bath

Workshop on PDEs with Random Coefficients
Weierstrass Institute, Berlin, 13–15 November, 2013

Recall from Lecture 1

Numerical Analysis of Elliptic PDEs with Random Coefficients

- **Motivation:** uncertainty/lack of data & stochastic modelling
  Examples of PDEs with random data
- **Model problem:** groundwater flow and radwaste disposal
  Elliptic PDEs with **rough** stochastic coefficients
- What are the **computational/analytical challenges**?
- **Numerical Analysis**
  - Assumptions, existence, uniqueness, regularity
  - FE analysis: Cea Lemma, interpolation error, functionals
  - Variational crimes (truncation error, quadrature)
  - Mixed finite element methods

Outline – Lecture 2

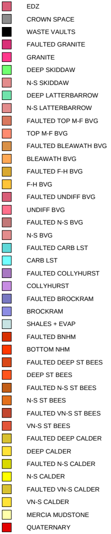
Novel Monte Carlo Methods and Uncertainty Quantification

- Stochastic Uncertainty Quantification (in PDEs)
- *The Curse of Dimensionality* & the **Monte Carlo Method**
- **Multilevel Monte Carlo** methods & Complexity Analysis
- Analysis of multilevel MC for the elliptic model problem
- **Quasi–Monte Carlo** methods
- Analysis of QMC for the elliptic model problem
- Bayesian Inference (stochastic inverse problems):

  Multilevel Markov Chain Monte Carlo

Model Problem: Uncertainty in Groundwater Flow

(applications in risk analysis of radwaste disposal, etc...)



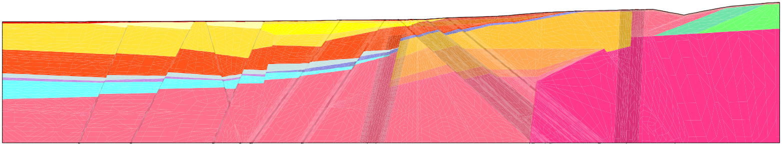
Darcy's Law: $\vec{q} + k(x,\omega)\,\nabla p = \vec{f}(x,\omega)$

Incompressibility: $\nabla \cdot \vec{q} = g(x,\omega)$

$+$ **Boundary Conditions**

Uncertainty in k $\Longrightarrow$ **Uncertainty** in p & $\vec{q}$

Stochastic Modelling!



Geology at Sellafield (former potential UK radwaste site) ©NIREX UK Ltd.

PDEs with Lognormal Random Coefficients

Key Computational Challenges

$$-\nabla\cdot(k(x,\omega)\nabla p(x,\omega)) = f(x,\omega), \quad x \in D \subset \mathbb{R}^d,\ \omega \in \Omega \text{ (prob. space)}$$

- **Sampling** from random field ($\log k(x,\omega)$ Gaussian) :
  - truncated Karhunen-Loève expansion of $\log k$
  - matrix factorisation, e.g. circulant embedding (FFT)
  - via pseudodifferential "precision" operator (PDE solves)
- **High-Dimensional Integration** (especially w.r.t. posterior):
  - stochastic Galerkin/collocation (+sparse)
  - Monte Carlo, QMC & Markov Chain MC
- **Solve** large number of **multiscale** deterministic PDEs:
  - Efficient discretisation & FE error analysis
  - Multigrid Methods, AMG, DD Methods

PDEs with Lognormal Random Coefficients

Key Computational Challenges

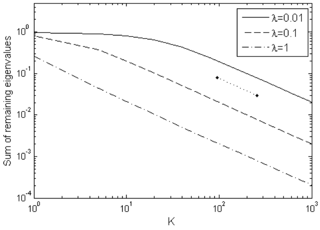
$$-\nabla\cdot(k(x,\omega)\nabla p(x,\omega)) = f(x,\omega), \quad x \in D \subset \mathbb{R}^d,\ \omega \in \Omega \text{ (prob. space)}$$

- **Sampling** from random field ($\log k(x,\omega)$ Gaussian) :
  - truncated Karhunen-Loève expansion of $\log k$
  - matrix factorisation, e.g. circulant embedding (FFT)
  - via pseudodifferential "precision" operator (PDE solves)
- **High-Dimensional Integration** (especially w.r.t. posterior):
  - stochastic Galerkin/collocation (+sparse)
  - Monte Carlo, QMC & Markov Chain MC — Lecture 2
- **Solve** large number of **multiscale** deterministic PDEs:
  - Efficient discretisation & FE error analysis
  - Multigrid Methods, AMG, DD Methods

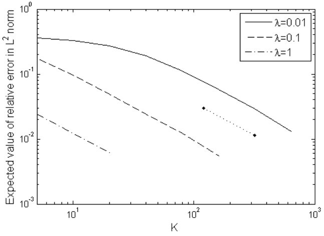
Why is it computationally so challenging?

- Low regularity (global): $k \in C^{0,\eta},\ \eta < 1 \implies$ **fine** mesh $h \ll 1$
- Large σ^2 & exponential $\implies$ **high** contrast $k_{\max}/k_{\min} > 10^6$
- Small $\lambda \implies$ **multiscale** + **high** stochast. dimension $s > 100$

e.g. for truncated KL expansion $\log k(x,\omega) \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x) Y_j(\omega)$



Remainder $\sum_{j>J} \mu_j$ in 1D



Truncation error of $\mathbb{E}[\|p\|_{L_2(0,1)}]$ w.r.t. s

Curse of Dimensionality ($s > 100$)

- **Stochastic Galerkin/collocation methods**
  - in their basic form cost grows very fast with dimension s (faster than exponential) $\rightarrow$ #stochastic DOFs $\mathcal{O}\left(\frac{(s+p)!}{s!p!}\right)$
  - lower # with sparse grids (Smolyak) but still **exponential**!

Curse of Dimensionality ($s > 100$)

- **Stochastic Galerkin/collocation methods**
  - in their basic form cost grows very fast with dimension s (faster than exponential) $\rightarrow$ #stochastic DOFs $\mathcal{O}\left(\frac{(s+p)!}{s!p!}\right)$
  - lower # with sparse grids (Smolyak) but still **exponential**!
  - (A priori) anisotropic sparse grids (Nobile) or (a posteriori) adaptive best N-term approximation (Eigel) can be dimension independent, but needs **smoothness**!

Curse of Dimensionality ($s > 100$)

- **Stochastic Galerkin/collocation methods**
  - in their basic form cost grows very fast with dimension s (faster than exponential) $\rightarrow$ #stochastic DOFs $\mathcal{O}\left(\frac{(s+p)!}{s!p!}\right)$
  - lower # with sparse grids (Smolyak) but still **exponential**!
  - (A priori) anisotropic sparse grids (Nobile) or (a posteriori) adaptive best N-term approximation (Eigel) can be dimension independent, but needs **smoothness**!

- **Monte Carlo type methods**
  - convergence of plain vanilla Monte Carlo is **always** dimension independent (even for rough problems) !
  - **BUT** order of convergence is slow: $\mathcal{O}(N^{-1/2})$!

Curse of Dimensionality ($s > 100$)

- **Stochastic Galerkin/collocation methods**
  - in their basic form cost grows very fast with dimension s (faster than exponential) $\rightarrow$ #stochastic DOFs $\mathcal{O}\left(\frac{(s+p)!}{s!p!}\right)$
  - lower # with sparse grids (Smolyak) but still **exponential**!
  - (A priori) anisotropic sparse grids (Nobile) or (a posteriori) adaptive best N-term approximation (Eigel) can be dimension independent, but needs **smoothness**!

- **Monte Carlo type methods**
  - convergence of plain vanilla Monte Carlo is **always** dimension independent (even for rough problems) !
  - **BUT** order of convergence is slow: $\mathcal{O}(N^{-1/2})$!
  - Quasi-MC also **dimension independent** and faster: $\sim \mathcal{O}(N^{-1})$! **But** requires **(also some) smoothness** !

Nonlinear Parameter Dependence

- Monte Carlo methods **do not** rely on KL-type expansion
  (can use circulant embedding or sparse pseudodifferential operators)
- Stochastic Galerkin matrix $\mathcal{A}$ is **block dense** due to nonlinear parameter dependence $\rightarrow$ even applying $\mathcal{A}$ is expensive!
  (can transform to convection-diffusion problem, but requires more smoothness and is not conservative [Elman, Ullmann, Ernst, 2010])
- best N-term theory by [Cohen, Schwab et al] does not apply!

Nonlinear Parameter Dependence

- Monte Carlo methods **do not** rely on KL-type expansion
  (can use circulant embedding or sparse pseudodifferential operators)
- Stochastic Galerkin matrix $\mathcal{A}$ is **block dense** due to nonlinear parameter dependence $\rightarrow$ even applying $\mathcal{A}$ is expensive!
  (can transform to convection-diffusion problem, but requires more smoothness and is not conservative [Elman, Ullmann, Ernst, 2010])
- best N-term theory by [Cohen, Schwab et al] does not apply!

Monte Carlo methods do not suffer from **curse of dimensionality**, they are **"non-intrusive"** and nonlinear parameter dependence is no problem, **but** the plain vanilla version is **too slow!**

Nonlinear Parameter Dependence

- Monte Carlo methods **do not** rely on KL-type expansion
  (can use circulant embedding or sparse pseudodifferential operators)
- Stochastic Galerkin matrix $\mathcal{A}$ is **block dense** due to nonlinear parameter dependence $\rightarrow$ even applying $\mathcal{A}$ is expensive!
  (can transform to convection-diffusion problem, but requires more smoothness and is not conservative [Elman, Ullmann, Ernst, 2010])
- best N-term theory by [Cohen, Schwab et al] does not apply!

Monte Carlo methods do not suffer from **curse of dimensionality**, they are **"non-intrusive"** and nonlinear parameter dependence is no problem, **but** the plain vanilla version is **too slow!**

Alternatives?

Monte Carlo for large scale problems (plain vanilla)

$$\mathbf{Z}_s(\omega) \in \mathbb{R}^s \xrightarrow{\text{Model}(h)} \mathbf{X}_h(\omega) \in \mathbb{R}^{M_h} \xrightarrow{\text{Output}} Q_{h,s}(\omega) \in \mathbb{R}$$

random input — state vector — quantity of interest

- e.g. $\mathbf{Z}_s$ multivariate Gaussian; $\mathbf{X}_h$ numerical solution of PDE; $Q_{h,s}$ a (non)linear functional of $\mathbf{X}_h$

Monte Carlo for large scale problems (plain vanilla)

$$\mathbf{Z}_s(\omega) \in \mathbb{R}^s \xrightarrow{\text{Model}(h)} \mathbf{X}_h(\omega) \in \mathbb{R}^{M_h} \xrightarrow{\text{Output}} Q_{h,s}(\omega) \in \mathbb{R}$$

random input | state vector | quantity of interest

- e.g. $\mathbf{Z}_s$ multivariate Gaussian; $\mathbf{X}_h$ numerical solution of PDE; $Q_{h,s}$ a (non)linear functional of $\mathbf{X}_h$
- $Q(\omega)$ inaccessible random variable s.t. $\mathbb{E}[Q_{h,s}] \xrightarrow{h\to 0,\, s\to\infty} \mathbb{E}[Q]$ and $|\mathbb{E}[Q_{h,s} - Q]| = \mathcal{O}(h^\alpha) + \mathcal{O}(s^{-\alpha'})$
- **Standard Monte Carlo** estimator for $\mathbb{E}[Q]$:

$$\hat{Q}^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_{h,s}^{(i)}$$

where $\{Q_{h,s}^{(i)}\}_{i=1}^{N}$ are i.i.d. samples computed with Model(h)

Monte Carlo for large scale problems (plain vanilla)

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\Big[\big(\hat{Q}^{\mathrm{MC}} - \mathbb{E}[Q]\big)^2\Big]}_{=:\ \mathrm{MSE}} = \mathbb{V}[\hat{Q}^{\mathrm{MC}}] + \big(\mathbb{E}[\hat{Q}^{\mathrm{MC}}] - \mathbb{E}[Q]\big)^2$$

$$= \underbrace{\frac{\mathbb{V}[Q_{h,s}]}{N}}_{\textbf{sampling error}} + \underbrace{\Big(\mathbb{E}[Q_{h,s} - Q]\Big)^2}_{\textbf{model error (“bias”)}}$$

- Typical (2D): $\alpha = 1 \;\Rightarrow\; \mathrm{MSE} = \mathcal{O}(N^{-1}) + \mathcal{O}(h^{-2}) \leq \mathrm{TOL}$

Monte Carlo for large scale problems (plain vanilla)

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\left[\left(\hat{Q}^{\mathrm{MC}} - \mathbb{E}[Q]\right)^2\right]}_{=:\ \mathrm{MSE}} = \mathbb{V}[\hat{Q}^{\mathrm{MC}}] + \left(\mathbb{E}[\hat{Q}^{\mathrm{MC}}] - \mathbb{E}[Q]\right)^2$$

$$= \underbrace{\frac{\mathbb{V}[Q_{h,s}]}{N}}_{\text{sampling error}} + \underbrace{\left(\mathbb{E}[Q_{h,s} - Q]\right)^2}_{\text{model error (“bias”)}}$$

- Typical (2D): $\alpha = 1 \;\Rightarrow\; \mathrm{MSE} = \mathcal{O}(N^{-1}) + \mathcal{O}(h^{-2}) \leq \mathrm{TOL}$

- Thus $h^{-2} \sim N \sim \mathrm{TOL}^{-2}$ and $\mathrm{Cost} = \mathcal{O}(Nh^{-2}) = \mathcal{O}(\mathrm{TOL}^{-4})$

  (e.g. for $\mathrm{TOL} = 10^{-3}$ we get $h^{-2} \sim N \sim 10^6$ and $\mathrm{Cost} = \mathcal{O}(10^{12})$!!)

Quickly becomes **prohibitively expensive !**

Return to model problem

- (Recall:) **Standard FEs** (cts pw. linear) on $\mathcal{T}^h$:

$$\longrightarrow \quad A(\omega)\,\mathbf{p}(\omega) = \mathbf{b}(\omega) \quad M_h \times M_h \text{ linear system}$$

(similarly for **mixed FEs**)

- **Quantity of interest:** Expected value $\mathbb{E}[Q]$ of $Q := \mathcal{G}(p)$
some (nonlinear) functional of the PDE solution p

Return to model problem

- (Recall:) **Standard FEs** (cts pw. linear) on $\mathcal{T}^h$:

$$\longrightarrow \quad A(\omega)\,\mathbf{p}(\omega) = \mathbf{b}(\omega) \quad M_h \times M_h \text{ linear system}$$

(similarly for **mixed FEs**)

- **Quantity of interest:** Expected value $\mathbb{E}[Q]$ of $Q := \mathcal{G}(p)$
  some (nonlinear) functional of the PDE solution p

- **Standard Monte Carlo** (MC) estimate for **inaccessible** $\mathbb{E}[Q]$:

$$\hat{Q}_h^{\mathrm{MC}} := \frac{1}{N} \sum_{i=1}^{N} Q_h^{(i)}, \qquad Q_h^{(i)} \text{ i.i.d. samples on } \mathcal{T}_h\,.$$

Return to model problem

- (Recall:) **Standard FEs** (cts pw. linear) on $\mathcal{T}^h$:

$$\longrightarrow \quad A(\omega)\,\mathbf{p}(\omega) = \mathbf{b}(\omega) \quad M_h \times M_h \text{ linear system}$$

(similarly for **mixed FEs**)

- **Quantity of interest:** Expected value $\mathbb{E}[Q]$ of $Q := \mathcal{G}(p)$
  some (nonlinear) functional of the PDE solution p

- **Standard Monte Carlo** (MC) estimate for **inaccessible** $\mathbb{E}[Q]$:

$$\hat{Q}_h^{\mathrm{MC}} := \frac{1}{N} \sum_{i=1}^{N} Q_h^{(i)}, \qquad Q_h^{(i)} \text{ i.i.d. samples on } \mathcal{T}_h\,.$$

- **(Quasi-)optimal** sampling & PDE solver (eg. **FFT** & **AMG**):

$$\Rightarrow \mathrm{Cost}(Q_h^{(i)}) = \mathcal{O}(M_h \log(M_h)) = \mathcal{O}(h^{-d}|\log h|)$$

Complexity of Standard Monte Carlo (avoiding log-factors)

Assuming

(A1)	$\left\lvert \mathbb{E}[Q_h - Q] \right\rvert = \mathcal{O}(h^\alpha)$	(mean FE error)
(A2')	$\mathbb{V}[Q_h] < \infty$	
(A3)	$\mathrm{Cost}(Q_h^{(i)}) = \mathcal{O}(h^{-\gamma})$	(deterministic solver)

to obtain **mean square error**

$$\mathbb{E}\left[(\hat{Q}_h^{\mathrm{MC}} - \mathbb{E}[Q])^2\right] \;=\; \mathcal{O}(\varepsilon^2)$$

the **total cost** is

$$\mathrm{Cost}(\hat{Q}_h^{\mathrm{MC}}) \;=\; \mathcal{O}\left(\varepsilon^{-2-\frac{\gamma}{\alpha}}\right)$$

Proof

Since

$$\underbrace{\mathbb{E}\Big[\big(\hat{Q}^{\rm MC} - \mathbb{E}[Q]\big)^2\Big]}_{=:\, e_{MSE}(\hat{Q}^{\rm MC})} \;=\; \frac{\mathbb{V}[Q_h]}{N} \;+\; \Big(\mathbb{E}[Q_h - Q]\Big)^2$$

a sufficient condition for $e_{MSE}(\hat{Q}^{\rm MC}) = \mathcal{O}(\varepsilon^2)$ is

$$N \;=\; \lceil 2\mathbb{V}[Q_h]\,\varepsilon^{-2}\rceil \quad \text{and} \quad h \;=\; c\varepsilon^{1/\alpha}$$

Therefore

$$\mathrm{Cost}(\hat{Q}_h^{\rm MC}) \;=\; N\,\mathrm{Cost}(Q_h^{(i)}) \;=\; \mathcal{O}\Big(\varepsilon^{-2-\frac{\gamma}{\alpha}}\Big)$$

Numerical Example (Standard Monte Carlo)

$D = (0,1)^2$, covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$ and $Q = \| -k\frac{\partial p}{\partial x_1}\|_{L^1(D)}$

using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

Numerical Example (Standard Monte Carlo)

$D = (0,1)^2$, covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$ and $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Numerically observed FE-error: $\approx \mathcal{O}(h^{3/4}) \implies \alpha \approx 3/4$.
- Numerically observed cost/sample: $\approx \mathcal{O}(h^{-2}) \implies \gamma \approx 2$.

Numerical Example (Standard Monte Carlo)

$D = (0,1)^2$, covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$ and $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Numerically observed FE-error: $\approx \mathcal{O}(h^{3/4}) \implies \alpha \approx 3/4$.
- Numerically observed cost/sample: $\approx \mathcal{O}(h^{-2}) \implies \gamma \approx 2$.
- **Total cost** to get RMSE $\mathcal{O}(\varepsilon)$: $\approx \mathcal{O}(\varepsilon^{-14/3})$
  to get error reduction by a factor 2 $\rightarrow$ cost grows by a factor 25!

Case 1: $\lambda = 0.3$, $\sigma^2 = 1$

ε	h^{-1}	N_h	Cost
0.01	129	1.4×10^4	21 min
0.002	1025	3.5×10^5	30 days

Case 2: $\lambda = 0.1$, $\sigma^2 = 3$

ε	h^{-1}	N_h	Cost
0.01	513	8.5×10^3	4 h
0.002	**Prohibitively large!!**		

(actual numbers & CPU times on a 2GHz Intel T7300 processor)

Multilevel Monte Carlo Methods

Multilevel Monte Carlo [Heinrich 2000], [Giles 2007]

Main Idea:

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

where $h_{\ell-1} = 2h_\ell$ and $Q_\ell := \mathcal{G}(p_{h_\ell})$

Multilevel Monte Carlo [Heinrich 2000], [Giles 2007]

Main Idea:

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

where $h_{\ell-1} = 2h_\ell$ and $Q_\ell := \mathcal{G}(p_{h_\ell})$

Key Observation (as in multigrid: easier to find corrections)

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0 \quad \text{as} \quad h_\ell \to 0 \quad \textbf{!}$$

Define following **multilevel MC** estimator for $\mathbb{E}[Q]$:

$$\hat{Q}_L^{\mathrm{ML}} := \sum_{\ell=0}^{L} \widehat{Y}_\ell^{\mathrm{MC}} \text{ where } Y_\ell := Q_\ell - Q_{\ell-1} \;\&\; Q_{-1} = 0$$

Complexity of Multilevel Monte Carlo (avoiding log's)

Assuming

(A1)	$\left\|\mathbb{E}[Q_\ell - Q]\right\| = \mathcal{O}(h_\ell^\alpha)$	(mean FE error)
(A2)	$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(h_\ell^\beta)$	(variance reduction)
(A3)	$\mathrm{Cost}(Q_\ell^{(i)}) = \mathcal{O}(h_\ell^{-\gamma})$	(deterministic solver)

$\exists L$ and $\{N_\ell\}_{\ell=0}^{L}$ such that to obtain **mean square error**

$$\mathbb{E}\left[(\hat{Q}_L^{\mathrm{ML}} - \mathbb{E}[Q])^2\right] = \mathcal{O}(\varepsilon^2)$$

the **total cost** is

$$\mathrm{Cost}(\hat{Q}_L^{\mathrm{ML}}) = \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right)$$

- **Adaptive error estimators** (to estimate L and $\{N_\ell\}$ on the fly):

$$|\widehat{Y}_\ell^{\mathrm{MC}}| \sim \left|\mathbb{E}[Q_{\ell-1} - Q]\right| \quad \text{and} \quad s(\widehat{Y}_\ell^{\mathrm{MC}}) \sim \mathbb{V}[Q_\ell - Q_{\ell-1}]$$

$\uparrow$ sample variance estimator

- **Adaptive error estimators** (to estimate L and $\{N_\ell\}$ on the fly):

$$|\widehat{Y}_\ell^{\mathrm{MC}}| \sim \big|\mathbb{E}[Q_{\ell-1} - Q]\big| \quad \text{and} \quad s(\widehat{Y}_\ell^{\mathrm{MC}}) \sim \mathbb{V}[Q_\ell - Q_{\ell-1}]$$

↑ sample variance estimator

- Assuming optimal AMG solver ($\gamma \approx d$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 0.75$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

- **Adaptive error estimators** (to estimate L and $\{N_\ell\}$ on the fly):

$$|\widehat{Y}_\ell^{\mathrm{MC}}| \sim \big|\mathbb{E}[Q_{\ell-1} - Q]\big| \quad \text{and} \quad s(\widehat{Y}_\ell^{\mathrm{MC}}) \sim \mathbb{V}[Q_\ell - Q_{\ell-1}]$$

$\uparrow$ sample variance estimator

- Assuming optimal AMG solver ($\gamma \approx d$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 0.75$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

d	MC	MLMC	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

Optimality (for $\gamma > \beta = 2\alpha$)

MLMC cost is asymptotically the same as **one deterministic solve** to accuracy ε in 2D & 3D, i.e. $\mathcal{O}(\varepsilon^{-\gamma/\alpha})$ **!!**

- **Adaptive error estimators** (to estimate L and $\{N_\ell\}$ on the fly):

$$|\widehat{Y}_\ell^{\mathrm{MC}}| \sim \left|\mathbb{E}[Q_{\ell-1} - Q]\right| \quad \text{and} \quad s(\widehat{Y}_\ell^{\mathrm{MC}}) \sim \mathbb{V}[Q_\ell - Q_{\ell-1}]$$

↑ sample variance estimator

- Assuming optimal AMG solver ($\gamma \approx d$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 0.75$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

d	MC	MLMC	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

Optimality (for $\gamma > \beta = 2\alpha$)

MLMC cost is asymptotically the same as **one deterministic solve** to accuracy ε in 2D & 3D, i.e. $\mathcal{O}(\varepsilon^{-\gamma/\alpha})$ **!!**

Can we achieve such huge gains in practice?

Numerical Examples (Multilevel MC)

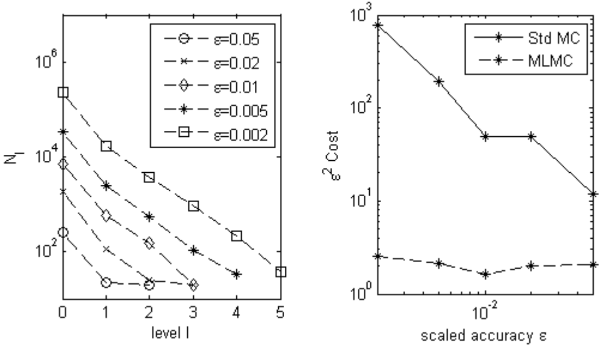
$D = (0,1)^2$; covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$; $Q = \|p\|_{L_2(D)}$

Std. FE discretisation, circulant embedding

Numerical Examples (Multilevel MC)

$D = (0,1)^2$; covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$; $Q = \|p\|_{L_2(D)}$

Std. FE discretisation, circulant embedding

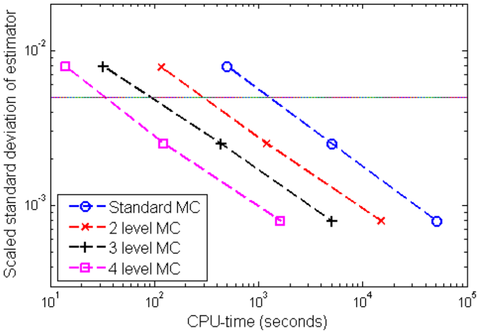


$$\sigma^2 = 1, \quad \lambda = 0.3, \quad h_0 = \frac{1}{8}$$

Numerical Examples (Multilevel MC)

$D = (0,1)^2$; covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$; $Q = \|p\|_{L_2(D)}$

Std. FE discretisation, circulant embedding



$h_L = 1/256$ (solid line is FE-error)

`Matlab` implementation on 3GHz Intel Core 2 Duo E8400 processor, 3.2GByte RAM, with **sparse direct solver**, i.e. $\gamma \approx 2.4$

Proof of Multilevel Complexity Theorem

Because $\widehat{Y}_\ell^{\mathrm{MC}}$ are independent, we get similar to single-level case

$$\mathbb{E}\left[\left(\hat{Q}_L^{\mathrm{ML}} - \mathbb{E}[Q]\right)^2\right] = \sum_{\ell=0}^{L} \frac{\mathbb{V}[Y_\ell]}{N_\ell} + \Big(\mathbb{E}[Q_L - Q]\Big)^2$$

A sufficient condition for the bias to be $\mathcal{O}(\varepsilon^2)$ is again $h_L \eqsim \varepsilon^{1/\alpha}$.

Proof of Multilevel Complexity Theorem

Because $\widehat{Y}_\ell^{\text{MC}}$ are independent, we get similar to single-level case

$$\mathbb{E}\big[\big(\hat{Q}_L^{\text{ML}} - \mathbb{E}[Q]\big)^2\big] \;=\; \sum_{\ell=0}^{L} \frac{\mathbb{V}[Y_\ell]}{N_\ell} \;+\; \Big(\mathbb{E}[Q_L - Q]\Big)^2$$

A sufficient condition for the bias to be $\mathcal{O}(\varepsilon^2)$ is again $h_L \eqsim \varepsilon^{1/\alpha}$.
Then we can minimise (with $V_l = \mathbb{V}[Y_\ell]$ and $C_\ell = \text{Cost}(Q_\ell^{(i)})$)

$$\sum_\ell N_\ell C_\ell \quad \text{subject to} \quad \sum_\ell V_\ell / N_\ell = \varepsilon^2/2$$

w.r.t. $\{N_\ell\}$, to get (for the case $\gamma > \beta$ – the other cases are similar):

$$N_\ell = 2\varepsilon^{-2} \Big(\sum_{\ell'} \sqrt{V_{\ell'} C_{\ell'}}\Big) \sqrt{V_\ell / C_\ell} \eqsim \varepsilon^{-2} \Bigg(\sum_{\ell'} h_{\ell'}^{\frac{\beta-\gamma}{2}}\Bigg) h_\ell^{\frac{\beta+\gamma}{2}} \;.$$

Proof of Multilevel Complexity Theorem

Because $\widehat{Y}_\ell^{\mathrm{MC}}$ are independent, we get similar to single-level case

$$\mathbb{E}\left[\left(\hat{Q}_L^{\mathrm{ML}} - \mathbb{E}[Q]\right)^2\right] \;=\; \sum_{\ell=0}^{L} \frac{\mathbb{V}[Y_\ell]}{N_\ell} \;+\; \Big(\mathbb{E}[Q_L - Q]\Big)^2$$

A sufficient condition for the bias to be $\mathcal{O}(\varepsilon^2)$ is again $h_L \eqsim \varepsilon^{1/\alpha}$.
Then we can minimise (with $V_l = \mathbb{V}[Y_\ell]$ and $C_\ell = \mathrm{Cost}(Q_\ell^{(i)})$)

$$\sum\nolimits_\ell N_\ell C_\ell \quad \text{subject to} \quad \sum\nolimits_\ell V_\ell / N_\ell = \varepsilon^2/2$$

w.r.t. $\{N_\ell\}$, to get (for the case $\gamma > \beta$ – the other cases are similar):

$$N_\ell = 2\varepsilon^{-2}\left(\sum\nolimits_{\ell'} \sqrt{V_{\ell'} C_{\ell'}}\right)\sqrt{V_\ell / C_\ell} \eqsim \varepsilon^{-2}\left(\sum\nolimits_{\ell'} h_{\ell'}^{\frac{\beta-\gamma}{2}}\right) h_\ell^{\frac{\beta+\gamma}{2}} .$$

Since $h_\ell = 2^{L-\ell} h_L \eqsim 2^{L-\ell}\varepsilon^{1/\alpha}$ the bound on $\sum_\ell C_\ell N_\ell$ follows.

Theory: Verifying Assumptions **(A1)** & **(A2)**

Recall from Wednesday's Lecture

- **Assumptions.** $\exists t \in (0,1], \quad q_* \geq 1$ s.t.

$$1/k^{min}(\omega) \in L^q(\Omega), \quad k \in L^q(\Omega, C^{0,t}(\overline{D})), \quad \forall q < \infty$$

$$f \in L^{q_*}(\Omega, H^{t-1}(D)), \quad \Phi \in L^{q_*}(\Omega, H^{t+\frac{1}{2}}(\partial D))$$

and D (convex) Lipschitz polygonal.

Theory: Verifying Assumptions **(A1)** & **(A2)**

Recall from Wednesday's Lecture

- **Assumptions.** $\exists t \in (0,1], \quad q_* \geq 1$ s.t.

$$1/k^{min}(\omega) \in L^q(\Omega), \quad k \in L^q(\Omega, C^{0,t}(\overline{D})), \quad \forall q < \infty$$

$$f \in L^{q_*}(\Omega, H^{t-1}(D)), \quad \Phi \in L^{q_*}(\Omega, H^{t+\frac{1}{2}}(\partial D))$$

  and D (convex) Lipschitz polygonal.

- **Theorem 2.** *$\forall q < q_*$, $s < t$ we have $p \in L^q(\Omega, H^{1+s}(D))$.*

- **Theorem 3.** *$\forall q < q_*$, $s < t$ we have*

$$\|p - p_h\|_{L^q(\Omega, H^1(D))} = \mathcal{O}(h^s) \;\&\; \|p - p_h\|_{L^q(\Omega, L^2(D))} = \mathcal{O}(h^{2s}).$$

- **Theorem 3b.** *If $\mathcal{G}(v) \in L^{q_*}(\Omega, H^{t-1}(D)^*)$ Fréchet diff'ble, then $\forall q < q_*$, $s < t$ we have*

$$\|\mathcal{G}(p) - \mathcal{G}(p_h)\|_{L^q(\Omega)} = \mathcal{O}(h^{2s})$$

- Thus, with $q = 1$ we get

$$|\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p_h)]| \;\leq\; \|\mathcal{G}(p) - \mathcal{G}(p_h)\|_{L^1(\Omega)} \;=\; \mathcal{O}(h^{2s})$$

$\Longrightarrow$ **(A1)** holds for any $\alpha < 2t$ (i.e. $\alpha < 1$ for exponential cov.)

- Thus, with $q = 1$ we get

$$|\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p_h)]| \;\leq\; \|\mathcal{G}(p) - \mathcal{G}(p_h)\|_{L^1(\Omega)} \;=\; \mathcal{O}(h^{2s})$$

$\Longrightarrow$ **(A1)** holds for any $\alpha < 2t$ (i.e. $\alpha < 1$ for exponential cov.)

- And with $q = 2$ we get

$$\mathbb{V}[\mathcal{G}(p_h) - \mathcal{G}(p_{2h})] \;\leq\; \|\mathcal{G}(p_h) - \mathcal{G}(p_{2h})\|^2_{L^2(\Omega)} \;\leq\; \mathcal{O}(h^{4s})$$

$\Longrightarrow$ **(A2)** holds for any $\beta < 4t$ (i.e. $\beta < 2$ for exponential cov.)

- Thus, with $q = 1$ we get

$$|\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p_h)]| \;\leq\; \|\mathcal{G}(p) - \mathcal{G}(p_h)\|_{L^1(\Omega)} \;=\; \mathcal{O}(h^{2s})$$

$\Longrightarrow$ **(A1)** holds for any $\alpha < 2t$ (i.e. $\alpha < 1$ for exponential cov.)

- And with $q = 2$ we get

$$\mathbb{V}[\mathcal{G}(p_h) - \mathcal{G}(p_{2h})] \;\leq\; \|\mathcal{G}(p_h) - \mathcal{G}(p_{2h})\|^2_{L^2(\Omega)} \;\leq\; \mathcal{O}(h^{4s})$$

$\Longrightarrow$ **(A2)** holds for any $\beta < 4t$ (i.e. $\beta < 2$ for exponential cov.)

- Hence (for rough fields, e.g. $t < 1/2$) **Cost** $= \mathcal{O}(\varepsilon^{-\gamma/\alpha})$
(Same as for deterministic solve!)

- Thus, with $q = 1$ we get

$$|\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p_h)]| \;\le\; \|\mathcal{G}(p) - \mathcal{G}(p_h)\|_{L^1(\Omega)} \;=\; \mathcal{O}(h^{2s})$$

$\Longrightarrow$ **(A1)** holds for any $\alpha < 2t$ (i.e. $\alpha < 1$ for exponential cov.)

- And with $q = 2$ we get

$$\mathbb{V}[\mathcal{G}(p_h) - \mathcal{G}(p_{2h})] \;\le\; \|\mathcal{G}(p_h) - \mathcal{G}(p_{2h})\|^2_{L^2(\Omega)} \;\le\; \mathcal{O}(h^{4s})$$

$\Longrightarrow$ **(A2)** holds for any $\beta < 4t$ (i.e. $\beta < 2$ for exponential cov.)

- Hence (for rough fields, e.g. $t < 1/2$) **Cost** $= \mathcal{O}(\varepsilon^{-\gamma/\alpha})$

(Same as for deterministic solve!)

Hence optimal and robust deterministic solver with $\gamma = d$ **crucial**!

This is a whole talk in itself!

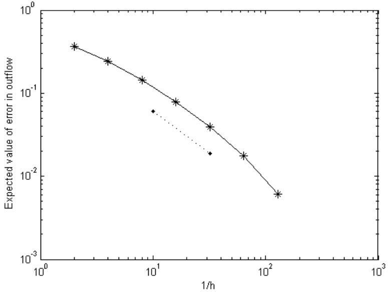
Numerical Confirmation

$D = (0,1)^2$; covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$ with $\lambda = 0.3$ and $\sigma^2 = 1$;

Std. FE discretisation, circulant embedding

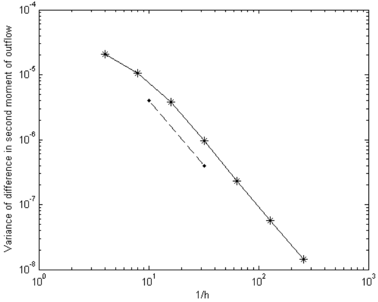
Numerical Confirmation

$D = (0,1)^2$; covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$ with $\lambda = 0.3$ and $\sigma^2 = 1$;
Std. FE discretisation, circulant embedding



$$\left|\mathbb{E}\left[\mathcal{G}^{(1)}(p) - \mathcal{G}^{(1)}(p_h)\right]\right|$$

where $\mathcal{G}^{(1)}(p) := L_\omega(\Psi) - b_\omega(\Psi, v)$
given $\Psi(x) = x$ (outflow on right).



$$\mathbb{V}\left[\mathcal{G}^{(2)}(p_h) - \mathcal{G}^{(2)}(p_{2h})\right]$$

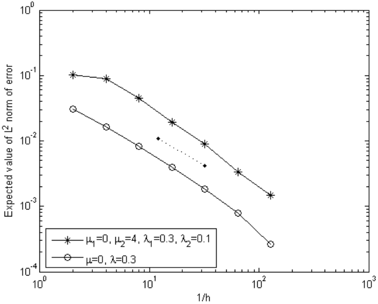
where $\mathcal{G}^{(2)}(p) := \left(\frac{1}{|D^*|}\int_{D^*} p(x)\,\mathrm{d}x\right)^2$
(i.e. 2nd moment of p over small patch)

$$\Longrightarrow \quad \alpha = 1 \text{ and } \beta = 2$$

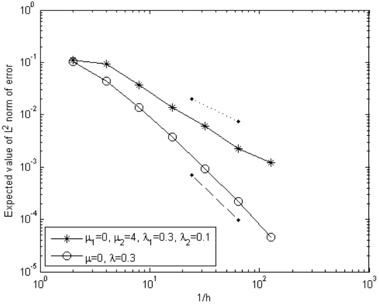
Discontinuous Permeability (piecewise lognormal)

Three layers; functional $\mathcal{G}(p) = \|p\|_{L_2(D)}$.

Exponential covariance



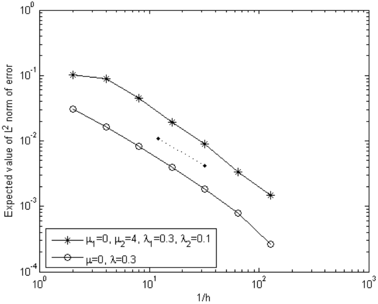
Gaussian covariance



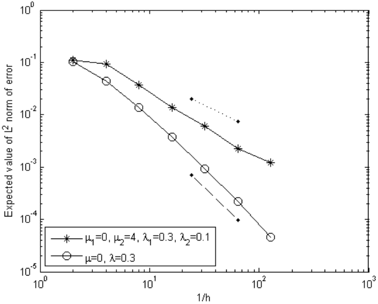
Discontinuous Permeability (piecewise lognormal)

Three layers; functional $\mathcal{G}(p) = \|p\|_{L_2(D)}$.

Exponential covariance



Gaussian covariance



As mentioned on Wednesday we can also analyse this case.

Similarly for the case of random interfaces
(and piecewise correlated random fields).

Point Evaluations and Particle Paths [Teckentrup, 2013]

- If in addition we assume $f \in L^{q_*}(\Omega, L^r(D))$ with $r > d/(1-t)$ then for all $q < q_*$

$$\|p - p_h\|_{L^q(\Omega, L^\infty(D))} = \mathcal{O}(h^{1+t}) \quad \text{and}$$

$$\|p - p_h\|_{L^q(\Omega, W^{1,\infty}(D))} = \mathcal{O}(h^t)$$

Point Evaluations and Particle Paths [Teckentrup, 2013]

- If in addition we assume $f \in L^{q_*}(\Omega, L^r(D))$ with $r > d/(1-t)$ then for all $q < q_*$

$$\|p - p_h\|_{L^q(\Omega, L^\infty(D))} = \mathcal{O}(h^{1+t}) \quad \text{and}$$

$$\|p - p_h\|_{L^q(\Omega, W^{1,\infty}(D))} = \mathcal{O}(h^t)$$

- Therefore **point evaluations** $p(x_*)$ and $\vec{q}(x_*)$ converge with $\mathcal{O}(h^{1+t})$ and $\mathcal{O}(h^t)$, respectively.

Point Evaluations and Particle Paths [Teckentrup, 2013]

- If in addition we assume $f \in L^{q_*}(\Omega, L^r(D))$ with $r > d/(1-t)$ then for all $q < q_*$

$$\|p - p_h\|_{L^q(\Omega, L^\infty(D))} = \mathcal{O}(h^{1+t}) \quad \text{and}$$

$$\|p - p_h\|_{L^q(\Omega, W^{1,\infty}(D))} = \mathcal{O}(h^t)$$

- Therefore **point evaluations** $p(x_*)$ and $\vec{q}(x_*)$ converge with $\mathcal{O}(h^{1+t})$ and $\mathcal{O}(h^t)$, respectively.

- This is of particular interest for **particle paths** (e.g. a plume spreading) computed via the integral

$$\vec{x}(T) = \vec{x}_0 + \int_0^T \vec{q}(\vec{x}(\tau))\, \mathrm{d}\tau$$

If $t = 1$ (current proof needs Lipschitz continuity of $\vec{q}$), then

$$\|\vec{x}(T) - \vec{x}_h(T)\|_{L^q(\Omega)} \lesssim \|p - p_h\|_{L^q(\Omega, W^{1,\infty}(D))} = \mathcal{O}(h).$$

Level-dependent Estimators (important in practice!)

Use $Q_\ell := \mathcal{G}(\widetilde{p}^{\,\ell}_{h_\ell})$ with level-dependent $\widetilde{p}^{\,\ell}_{h_\ell}$ in multilevel splitting

$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] \; + \; \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}],$$

e.g. vary #terms s_ℓ in KL-expansion (**smoother** on coarse grids)

Level-dependent Estimators (important in practice!)

Use $Q_\ell := \mathcal{G}(\widetilde{p}^{\ell}_{h_\ell})$ with level-dependent $\widetilde{p}^{\ell}_{h_\ell}$ in multilevel splitting

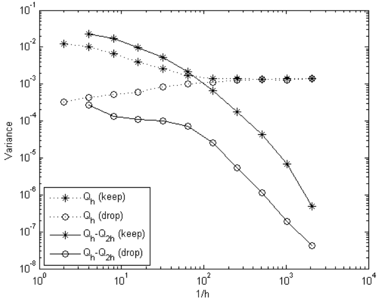
$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}],$$

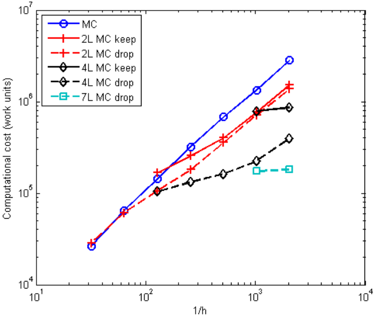
e.g. vary #terms s_ℓ in KL-expansion (**smoother** on coarse grids)

- **Strang Lemma:** Same bounds on α and β if $s_\ell^{-1} = \mathcal{O}(h_\ell)$.
  (using the truncation error analysis I showed on Wednesday)
- **No** gain asymptotically (but also no loss!).
- Helps with the **absolute gain** of the multilevel estimator and makes it **feasible** also on **coarser** grids with $h_\ell > \lambda$.
  (in basic multilevel MC need $h_0 < \lambda$)

Level-dependent Estimators (important in practice!)

1D Example: $\mathcal{G}(p) = p(x^*)$, $\sigma^2 = 1$, $\lambda = 0.01$ and $s_\ell := h_\ell^{-1}$





Other developments in MLMC

- many other PDEs and applications
- similar results for mixed FEs, FVM, ...
- can optimise **all** parameters (not just $\{N_\ell\}$) [Hajiali, Tempone]
- adaptivity [Von Schwerin, Tempone et al]
- variance estimation [Bierig, Chernov]
- optimal estimation of CDFs, PDFs [Giles, Nagapetyan, Ritter]
- antithetic sampling & coarse grid variates [Park, Giles et al]
- hybrid with stochastic collocation [Tesei, Nobile et al]
- generalisation to general multilevel quadrature [Harbrecht et al]
- multilevel QMC [Kuo, Schwab, Sloan] see below

Quasi–Monte Carlo Methods

Reducing # Samples (Quasi–Monte Carlo)

[Graham, Kuo, Nuyens, RS, Sloan '11], [Gra., Kuo, Nichols, RS, Schwab, Slo. '13]

$$\mathbb{E}[\mathcal{G}(p)] \approx \int_{[0,1]^s} \mathcal{G}\Big(p_h^s(\cdot, \mathbf{\Phi}^{-1}(\mathbf{z}))\Big) \mathrm{d}\mathbf{z} \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{G}\Big(p_h^s(\cdot, \mathbf{\Phi}^{-1}(\mathbf{z}^{(i)}))\Big)$$

with $\mathbf{\Phi} : \mathbb{R}^s \to [0,1]^s$ the cumulative normal distribution function.

Reducing # Samples (Quasi–Monte Carlo)

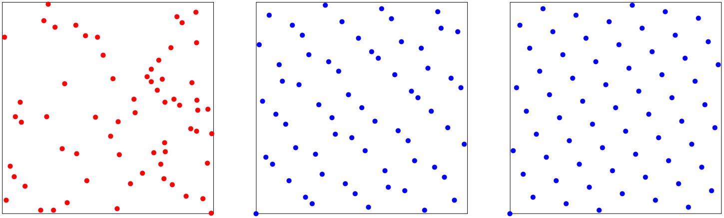
[Graham, Kuo, Nuyens, RS, Sloan '11], [Gra., Kuo, Nichols, RS, Schwab, Slo. '13]

$$\mathbb{E}[\mathcal{G}(p)] \approx \int_{[0,1]^s} \mathcal{G}\Big(p_h^s(\cdot, \mathbf{\Phi}^{-1}(\mathbf{z}))\Big)\mathrm{d}\mathbf{z} \approx \frac{1}{N}\sum_{i=1}^{N} \mathcal{G}\Big(p_h^s(\cdot, \mathbf{\Phi}^{-1}(\mathbf{z}^{(i)}))\Big)$$

with $\mathbf{\Phi} : \mathbb{R}^s \to [0,1]^s$ the cumulative normal distribution function.

Monte Carlo: $\mathbf{z}^{(n)}$ unif. random
$\mathcal{O}(N^{-1/2})$ convergence
order of variables irrelevant

QMC: $\mathbf{z}^{(n)}$ deterministic
close to $\mathcal{O}(N^{-1})$ convergence
order of variables v. important



64 random points   64 Sobol′ points   64 lattice points

Numerical Results

[Graham, Kuo, Nuyens, RS, Sloan, JCP 2011]

Covariance

$$r(\mathbf{x},\mathbf{y}) = \sigma^2 \exp\Big(- \|\mathbf{x}-\mathbf{y}\|_1/\lambda\Big) \qquad (\, \|\cdot\|_2 \text{ similar})$$

	Case 1	Case 2	Case 3	Case 4	Case 5
σ^2	1	1	1	3	3
λ	1	0.3	0.1	1	0.1

Mixed FEM (RT0 + p.w. const): Uniform grid $h = 1/m$ on $(0,1)^2$

Sampling: circulant embedding, dimension $s = \mathcal{O}(m^2)$ (v. large)

("discrete KL-expansion" via FFT)

QMC Method: randomised QMC with N Sobol' points

Algorithm profile

Time (in sec) on modest laptop for $N = 1000$, CASE 1:

(similar for other cases)

m	s	Setup	Φ^{-1}	FFTW	PDE Solve	TOT
33	4.1 (+3)	0.00	1.0	0.22	4.5	5.9
65	1.7 (+4)	0.01	3.9	1.2	16.5	22
129	6.6 (+4)	0.06	15	5.1	67	92
257	2.6 (+5)	0.15	62	31	290	400
513	1.0 (+6)	0.6	258	145	1280	1750
Order	m^2	m^2	m^2	$m^2 \log m$	$\sim m^2$	$\sim m^2$

Algorithm profile

Time (in sec) on modest laptop for $N = 1000$, CASE 1:

(similar for other cases)

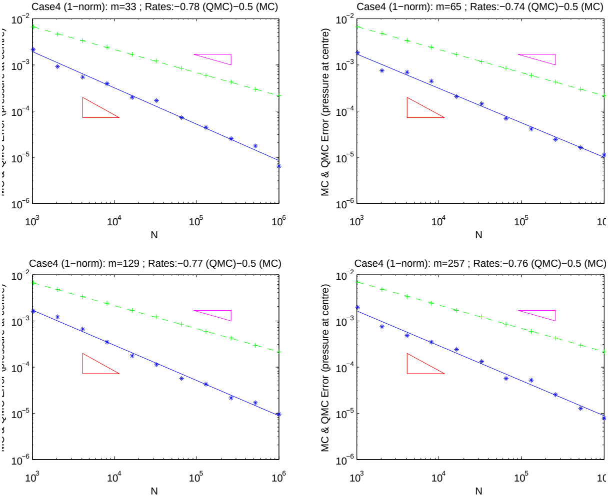
m	s	Setup	Φ^{-1}	FFTW	PDE Solve	TOT
33	4.1 (+3)	0.00	1.0	0.22	4.5	5.9
65	1.7 (+4)	0.01	3.9	1.2	16.5	22
129	6.6 (+4)	0.06	15	5.1	67	92
257	2.6 (+5)	0.15	62	31	290	400
513	1.0 (+6)	0.6	258	145	1280	1750
Order	m^2	m^2	m^2	$m^2 \log m$	$\sim m^2$	$\sim m^2$

Using divergence free reduction to SPD problem and `amg1r5`
[Cliffe, Graham, RS, Stals, 2000]
One mixed FE (saddle point system) solve with $\approx 1.3(+6)$ DOF takes ≈ 1.3s !!

Dimension independence (increasing m and hence s)

Quadrature error for mean pressure at centre (CASE 4)

(no FE error, MC in green, QMC in blue)



Robustness (varying σ^2 and λ)

Expected value of effective permeability (here FE error present)

h needed to obtain a discretization error $< 10^{-3}$

N needed to obtain (Q)MC error $< 0.5 \times 10^{-3}$ (95% confidence)

σ^2	λ	$1/h$	N (QMC)	N (MC)	CPU (QMC)	CPU (MC)
1	1	17	1.2(+5)	1.9(+7)	0.05 h	8 h
1	0.3	129	3.3(+4)	3.9(+6)	0.9 h	110 h
1	0.1	513	1.2(+4)	5.9(+5)	6.5 h	330 h
3	1	33	4.3(+6)	3.6(+8)	9 h	750 h
3	0.1	513	3.0(+4)	5.8(+5)	20 h	390 h

(last line calculated with twice the tolerance!)

Robustness (varying σ^2 and λ)

Expected value of effective permeability (here FE error present)

h needed to obtain a discretization error $< 10^{-3}$

N needed to obtain (Q)MC error $< 0.5 \times 10^{-3}$ (95% confidence)

σ^2	λ	$1/h$	N (QMC)	N (MC)	CPU (QMC)	CPU (MC)
1	1	17	1.2(+5)	1.9(+7)	0.05 h	8 h
1	0.3	129	3.3(+4)	3.9(+6)	0.9 h	110 h
1	0.1	513	1.2(+4)	5.9(+5)	6.5 h	330 h
3	1	33	4.3(+6)	3.6(+8)	9 h	750 h
3	0.1	513	3.0(+4)	5.8(+5)	20 h	390 h

(last line calculated with twice the tolerance!)

Smaller λ needs smaller h but also smaller N (ergodicity).

Strong superiority of QMC in all cases.

Theory [Graham, Kuo, Nicholls, RS, Schwab, Sloan, 2013]

- Truncated Karhunen-Loeve expansion:

$$k(\mathbf{x},\omega) \approx k^s(\mathbf{x},\omega) := k_*(\mathbf{x}) + k_0(\mathbf{x}) \exp\left(\sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(\mathbf{x}) Y_j(\omega) \right)$$

$\mathbf{y} = (Y_j)_{j=1}^s$ i.i.d. $N(0,\sigma^2)$; (μ_j, ϕ_j) orth. eigenpairs of $\int_\Omega R(\mathbf{x},\mathbf{x}')\phi(\mathbf{x}')\mathrm{d}\mathbf{x}'$

Theory [Graham, Kuo, Nicholls, RS, Schwab, Sloan, 2013]

- Truncated Karhunen-Loeve expansion:

$$k(\mathbf{x},\omega) \approx k^s(\mathbf{x},\omega) := k_*(\mathbf{x}) + k_0(\mathbf{x}) \exp\left(\sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(\mathbf{x}) Y_j(\omega)\right)$$

$\mathbf{y} = (Y_j)_{j=1}^s$ i.i.d. $N(0,\sigma^2)$; (μ_j, ϕ_j) orth. eigenpairs of $\int_\Omega R(\mathbf{x},\mathbf{x}')\phi(\mathbf{x}')\mathrm{d}\mathbf{x}'$

- Standard cts. p.w. linear FEs on grid $\mathcal{T}^h$: Find $p_h^s \in V_h$ s.t.

$$\int_D k^s(\mathbf{x},\omega)\nabla p_h^s(\mathbf{x},\omega)\cdot\nabla v_h\,\mathrm{d}\mathbf{x} = \langle f, v_h\rangle \quad \forall v_h \in V_h, \text{ a.s. } \omega \in \Omega$$

Three Sources of Error:

- **Truncation error** (s): $|\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p^s)]|$
- **Discretisation error** (h): $|\mathbb{E}[\mathcal{G}(p^s) - \mathcal{G}(p_h^s)]|$ as above
- **Quadrature error** (N): $\left|\int_{[0,1]^s} \mathcal{G}(p_h^s(\cdot,\boldsymbol{\Phi}^{-1}(\mathbf{z}))\mathrm{d}\mathbf{z} - Q_N^s(\mathcal{G}(p_h^s))\right|$

Truncation Error (recall from Wednesday)

- Uses **Fernique's Thm.** & depends on decay of KL-eigvals μ_j
  - $O(j^{-(d+1)/d})$ for exponential covariance with 2-norm
  - $O(\exp(-c_1 j))$ for Gaussian covariance
  - $O(j^{-(d+2\nu)/d})$ for Matérn class (with parameter $\nu > 1/2$)

and on growth of $\|\nabla \phi_j\|_{L^\infty(D)}$ (**hard to estimate!**)

Truncation Error (recall from Wednesday)

- Uses **Fernique's Thm.** & depends on decay of KL-eigvals μ_j
  - $O(j^{-(d+1)/d})$ for exponential covariance with 2-norm
  - $O(\exp(-c_1 j))$ for Gaussian covariance
  - $O(j^{-(d+2\nu)/d})$ for Matérn class (with parameter $\nu > 1/2$)

and on growth of $\|\nabla\phi_j\|_{L^\infty(D)}$ (**hard to estimate!**)

If $\exists r^* \in (0,1)$ s.t. $\sum_{j\geq 1} j^\sigma \mu_j^2 \|\phi_j\|_{L^\infty(D)}^{2(1-r)} \|\nabla\phi_j\|_{L^\infty(D)}^{2r} < \infty$ then

$$\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p^s)] \;=\; O(s^{-\sigma/2})$$

Truncation Error (recall from Wednesday)

- Uses **Fernique's Thm.** & depends on decay of KL-eigvals μ_j
  - $O(j^{-(d+1)/d})$ for exponential covariance with 2-norm
  - $O(\exp(-c_1 j))$ for Gaussian covariance
  - $O(j^{-(d+2\nu)/d})$ for Matérn class (with parameter $\nu > 1/2$)

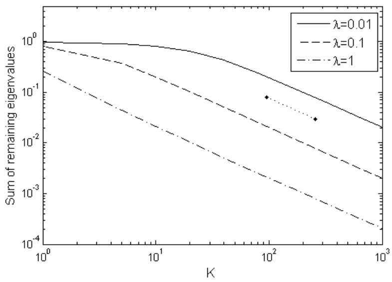
and on growth of $\|\nabla\phi_j\|_{L^\infty(D)}$ (**hard to estimate!**)

If $\exists r^* \in (0,1)$ s.t. $\sum_{j\geq 1} j^\sigma \mu_j^2 \|\phi_j\|_{L^\infty(D)}^{2(1-r)} \|\nabla\phi_j\|_{L^\infty(D)}^{2r} < \infty$ then

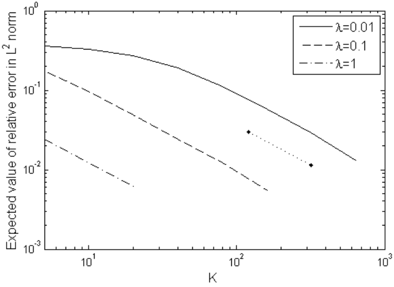
$$\mathbb{E}[\mathcal{G}(p) - \mathcal{G}(p^s)] \;=\; O(s^{-\sigma/2})$$

- Assumption satisfied for 1-norm exponential with $\sigma < 1$
- and for Matérn with $\nu > d/2$ (proof in [Graham et al, 2013])
- For Gaussian covariance one can prove exponential decay

Truncation Error (recall from Wednesday)



Remainder $\sum_{j>s} \mu_j$ in 1D (exponential)



Converg. of $\left|\mathbb{E}[\|p\|_{L_2(0,1)} - \|p^s\|_{L_2(0,1)}]\right|$

Importance of correlation length λ !

Quadrature Error (Standard Monte Carlo)

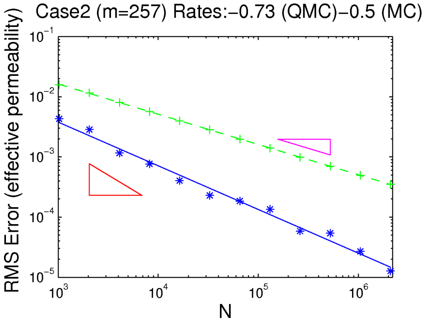
- By *Law of Large Numbers* for **random** points $\mathbf{z}^{(i)} \in [0,1]^s$:

$$\mathrm{RMSE}\left[\mathbb{E}[\mathcal{G}(p_h^s)] - \left(\widehat{\mathcal{G}(p_h^s)}\right)_N^{\mathrm{MC}}\right] = O(N^{-1/2})$$

Quadrature Error (Standard Monte Carlo)

- By *Law of Large Numbers* for **random** points $\mathbf{z}^{(i)} \in [0,1]^s$:

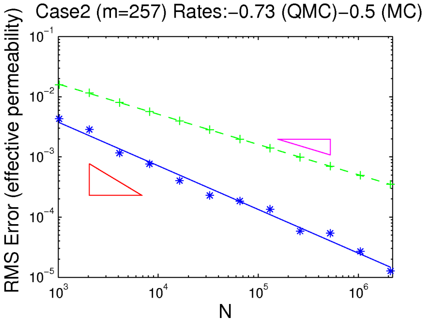
$$\mathrm{RMSE}\left[\mathbb{E}[\mathcal{G}(p_h^s)] - \left(\widehat{\mathcal{G}(p_h^s)}\right)_N^{\mathsf{MC}}\right] = O(N^{-1/2})$$



Quadrature Error (Standard Monte Carlo)

- By *Law of Large Numbers* for **random** points $\mathbf{z}^{(i)} \in [0,1]^s$:

$$\mathrm{RMSE}\left[\mathbb{E}[\mathcal{G}(p_h^s)] - \widehat{\left(\mathcal{G}(p_h^s)\right)}_N^{\mathsf{MC}}\right] = O(N^{-1/2})$$



Can we do better with deterministically chosen points & can we prove it?

Sample Points & Equal Weight Quadrature Rules

Quasi-Monte Carlo: $Q_N^s\big(\mathcal{G}(p_h^s)\big) := \frac{1}{N}\sum_{i=1}^{N}\mathcal{G}\Big(p_h^s\big(\cdot, \mathbf{\Phi}^{-1}(\mathbf{z}^{(i)})\big)\Big)$

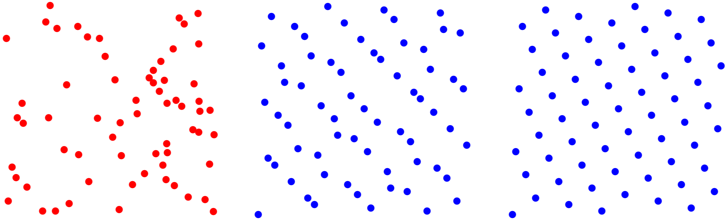
How to choose $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}$?

Sample Points & Equal Weight Quadrature Rules

Quasi-Monte Carlo: $Q_N^s(\mathcal{G}(p_h^s)) := \dfrac{1}{N}\sum_{i=1}^{N} \mathcal{G}\Big(p_h^s(\cdot, \boldsymbol{\Phi}^{-1}(\mathbf{z}^{(i)}))\Big)$

How to choose $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}$?

- **Low discrepancy points**: Sobol (1950s), Faure, Niederreiter (1980s), Dick ...
- **Lattice rules**: Korobov, Hlawka, Hua, Wang (50s), Sloan...



64 random points | 64 Sobol′ points | 64 lattice points

Quasi-Monte Carlo Lattice Rule (of rank 1)

[Sloan & Joe, Lattice Methods for Multiple Integration, OUP, 1994]

Given a *generating vector* $\mathbf{z}_{\mathrm{gen}} \in \{1, \ldots, N-1\}^s$ and a *random shift* $\boldsymbol{\Delta} \sim U\big[(0,1)^s\big]$:

$$\mathbf{z}^{(i)} := \mathrm{frac}\left(\frac{i\,\mathbf{z}_{\mathrm{gen}}}{N} + \boldsymbol{\Delta}\right), \qquad i = 1, \ldots, N$$

Now theory available for functions in weighted tensor product Sobolev spaces.

Quasi-Monte Carlo Lattice Rule (of rank 1)

[Sloan & Joe, Lattice Methods for Multiple Integration, OUP, 1994]

Given a *generating vector* $\mathbf{z}_{\rm gen} \in \{1, \ldots, N-1\}^s$ and a *random shift* $\boldsymbol{\Delta} \sim U\big[(0,1)^s\big]$:

$$\mathbf{z}^{(i)} := \mathrm{frac}\left(\frac{i\,\mathbf{z}_{\rm gen}}{N} + \boldsymbol{\Delta}\right), \qquad i = 1, \ldots, N$$

Now theory available for functions in weighted tensor product Sobolev spaces.

- **Weighted spaces/existence**: Sloan, Woźniakoski, '98 & '01
- **Construction**: Sloan, Reztsov, Kuo, Joe, 2002
  (see also `www.maths.unsw.edu.au/~fkuo`: CBC construction)
- **Infinite dimensions and improper integrals**:
  Kuo, Sloan, Wasilkowski, Waterhouse, 2010;
  Kuo, Nicholls, 2013

Quadrature Error Analysis (non-affine lognormal case)

[Graham, Kuo, Nichols, RS, Schwab, Sloan, 2013]

Dimension-independent bounds if integrand F is in special **weighted tensor-product Sobolev space** $\mathcal{W}_{s,\boldsymbol{\gamma},\psi} := (H^1_{\boldsymbol{\gamma},\psi}(\mathbb{R}))^s$ with norm

$$\|F\|^2_{\mathcal{W}_{s,\boldsymbol{\gamma},\psi}} := \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{\mathbb{R}^{|\mathfrak{u}|}} \left| \frac{\partial^{|\mathfrak{u}|} F}{\partial \mathbf{y}_{\mathfrak{u}}}(\mathbf{y}_{\mathfrak{u}}; \mathbf{0}) \right|^2 \prod_{j\in\mathfrak{u}} \psi^2(y_j)\, \mathrm{d}\mathbf{y}_{\mathfrak{u}} \,.$$

Quadrature Error Analysis (non-affine lognormal case)

[Graham, Kuo, Nichols, RS, Schwab, Sloan, 2013]

Dimension-independent bounds if integrand F is in special **weighted tensor-product Sobolev space** $\mathcal{W}_{s,\gamma,\psi} := (H^1_{\gamma,\psi}(\mathbb{R}))^s$ with norm

$$\|F\|^2_{\mathcal{W}_{s,\gamma,\psi}} := \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{\mathbb{R}^{|\mathfrak{u}|}} \left| \frac{\partial^{|\mathfrak{u}|} F}{\partial \mathbf{y}_{\mathfrak{u}}}(\mathbf{y}_{\mathfrak{u}}; \mathbf{0}) \right|^2 \prod_{j\in\mathfrak{u}} \psi^2(y_j)\, \mathrm{d}\mathbf{y}_{\mathfrak{u}} \,.$$

- Ordering of coordinates **crucial**! Cannot construct rules that are equally good in all dimensions.

Quadrature Error Analysis (non-affine lognormal case)

[Graham, Kuo, Nichols, RS, Schwab, Sloan, 2013]

Dimension-independent bounds if integrand F is in special **weighted tensor-product Sobolev space** $\mathcal{W}_{s,\gamma,\psi} := (H^1_{\gamma,\psi}(\mathbb{R}))^s$ with norm

$$\|F\|^2_{\mathcal{W}_{s,\gamma,\psi}} := \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{\mathbb{R}^{|\mathfrak{u}|}} \left| \frac{\partial^{|\mathfrak{u}|} F}{\partial \mathbf{y}_{\mathfrak{u}}}(\mathbf{y}_{\mathfrak{u}};\mathbf{0}) \right|^2 \prod_{j\in\mathfrak{u}} \psi^2(y_j)\,\mathrm{d}\mathbf{y}_{\mathfrak{u}}\,.$$

- Ordering of coordinates **crucial**! Cannot construct rules that are equally good in all dimensions.
- Weight function ψ^2 controls decay at infinity for improper integrals (either exponential or Gaussian).
- Weights $\gamma_{\mathfrak{u}}$ (for subsets $\mathfrak{u}$ of coordinates) have to decrease sufficiently fast.

Quadrature Error Analysis (non-affine lognormal case)

[Graham, Kuo, Nichols, RS, Schwab, Sloan, 2013]

Dimension-independent bounds if integrand F is in special **weighted tensor-product Sobolev space** $\mathcal{W}_{s,\gamma,\psi} := (H^1_{\gamma,\psi}(\mathbb{R}))^s$ with norm

$$\|F\|^2_{\mathcal{W}_{s,\gamma,\psi}} := \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{\mathbb{R}^{|\mathfrak{u}|}} \left| \frac{\partial^{|\mathfrak{u}|} F}{\partial \mathbf{y}_{\mathfrak{u}}}(\mathbf{y}_{\mathfrak{u}}; \mathbf{0}) \right|^2 \prod_{j\in\mathfrak{u}} \psi^2(y_j)\, \mathrm{d}\mathbf{y}_{\mathfrak{u}} .$$

- Ordering of coordinates **crucial**! Cannot construct rules that are equally good in all dimensions.
- Weight function ψ^2 controls decay at infinity for improper integrals (either exponential or Gaussian).
- Weights $\gamma_{\mathfrak{u}}$ (for subsets $\mathfrak{u}$ of coordinates) have to decrease sufficiently fast.
- Efficient CBC construction available – controlled by weights $\gamma_{\mathfrak{u}}$.

Quadrature Error Analysis (contd.)

- To show $\mathcal{G}(p_h^s) \in \mathcal{W}_{s,\gamma,\psi}$ bound mixed 1st derivatives of p_h^s w.r.t. parameters in a finite subset $\mathfrak{u} \subset \mathbb{N}$:

$$\left| \frac{\partial^{|\mathfrak{u}|} p_h^s}{\partial \mathbf{y}_{\mathfrak{u}}}(\cdot, \mathbf{y}) \right|_{H^1(D)} \leq \frac{\|f\|_{H^{-1}(D)}}{k_{\min}(\mathbf{y})} \frac{|\mathfrak{u}|!}{\ln 2^{|\mathfrak{u}|}} \left(\prod_{j \in \mathfrak{u}} \sqrt{\mu_j} \, \|\phi_j\|_{L^\infty(D)} \right)$$

Quadrature Error Analysis (contd.)

- To show $\mathcal{G}(p_h^s) \in \mathcal{W}_{s,\gamma,\psi}$ bound mixed 1st derivatives of p_h^s w.r.t. parameters in a finite subset $\mathfrak{u} \subset \mathbb{N}$:

$$\left| \frac{\partial^{|\mathfrak{u}|} p_h^s}{\partial \mathbf{y}_{\mathfrak{u}}}(\cdot, \mathbf{y}) \right|_{H^1(D)} \leq \frac{\|f\|_{H^{-1}(D)}}{k_{\min}(\mathbf{y})} \frac{|\mathfrak{u}|!}{\ln 2^{|\mathfrak{u}|}} \left(\prod_{j \in \mathfrak{u}} \sqrt{\mu_j} \, \|\phi_j\|_{L^\infty(D)} \right)$$

- Assume $\mathcal{G}(p_h^s)$ linear. If KL-eigenvalues μ_j decay suff'ly fast we can find weights $\gamma_{\mathfrak{u}}$ s.t. $\mathcal{G}(p_h^s) \in \mathcal{W}_{s,\gamma,\psi}$. In particular, can choose $\gamma_{\mathfrak{u}} = \left(\frac{|\mathfrak{u}|!}{(\ln 2)^{|\mathfrak{u}|}}\right)^{2/(1+\lambda)} \prod_{j \in \mathfrak{u}} \gamma_j(\mu_j, \lambda)$ and λ depends on decay rate of μ_j.

Theorem (hidden constants independent of s!)

$$\mathbb{E}[\mathcal{G}(p_h^s)] - Q_N^s(\mathcal{G}(p_h^s)) = \mathcal{O}(N^{-1/2}) \quad \text{if} \quad \mu_j \|\phi_j\|^2_{L^\infty(D)} = O(j^{-2-\delta})$$

$$\mathbb{E}[\mathcal{G}(p_h^s)] - Q_N^s(\mathcal{G}(p_h^s)) = \mathcal{O}(N^{-1+\delta}) \quad \text{if} \quad \mu_j \|\phi_j\|^2_{L^\infty(D)} = O(j^{-3})$$

Optimal rates (**provable**) for Matérn with $\nu > \frac{3}{2}d$.

Regularity Proof Idea

(also important for the analysis of the stochastic Galerkin/collocation methods)

- For regularity, start with Lax-Milgram $\Rightarrow$

$$\|p_h^s(\cdot, \mathbf{y})\|_a \leq \frac{1}{\sqrt{k_{\min}(\mathbf{y})}} \|f\|_{H^{-1}(D)} \quad \text{for a.a. } \mathbf{y} \in \mathbb{R}^{\mathbb{N}}$$

Regularity Proof Idea

(also important for the analysis of the stochastic Galerkin/collocation methods)

- For regularity, start with Lax-Milgram $\Rightarrow$

$$\|p_h^s(\cdot,\mathbf{y})\|_a \le \frac{1}{\sqrt{k_{\min}(\mathbf{y})}}\|f\|_{H^{-1}(D)} \quad \text{for a.a. } \mathbf{y} \in \mathbb{R}^{\mathbb{N}}$$

- Then show inductively that (with $b_j = \sqrt{\mu_j}\|\phi_j\|_{L^\infty(D)}$)

$$\|\partial^{\mathfrak{u}} p_h^s(\cdot,\mathbf{y})\|_a \le \Lambda_{|\mathfrak{u}|} \prod_{j\ge 1} b_j^{\nu_j} \frac{\|f\|_{H^{-1}(D)}}{\sqrt{k_{\min}(\mathbf{y})}}$$

where $\Lambda_0 = 1$ and $\Lambda_n = \sum_{i=0}^{n-1}\binom{n}{i}\Lambda_i$ using the Leibniz rule and the simple bound $\left\|\frac{\partial^{\mathfrak{u}} k(.,\mathbf{y})}{k(.,\mathbf{y})}\right\| \le \prod_{j\ge1} b_j^{\nu_j}$ (where $\nu_j = \delta_{j\in\mathfrak{u}}$).

Regularity Proof Idea

(also important for the analysis of the stochastic Galerkin/collocation methods)

- For regularity, start with Lax-Milgram $\Rightarrow$

$$\|p_h^s(\cdot,\mathbf{y})\|_a \le \frac{1}{\sqrt{k_{\min}(\mathbf{y})}}\|f\|_{H^{-1}(D)} \quad \text{for a.a. } \mathbf{y} \in \mathbb{R}^{\mathbb{N}}$$

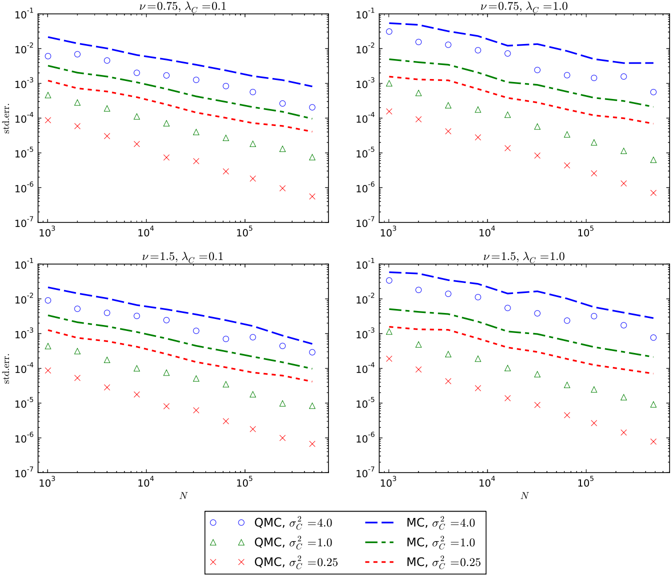
- Then show inductively that (with $b_j = \sqrt{\mu_j}\|\phi_j\|_{L^\infty(D)}$)

$$\|\partial^{\mathfrak{u}} p_h^s(\cdot,\mathbf{y})\|_a \le \Lambda_{|\mathfrak{u}|} \prod_{j\ge 1} b_j^{\nu_j} \frac{\|f\|_{H^{-1}(D)}}{\sqrt{k_{\min}(\mathbf{y})}}$$

where $\Lambda_0 = 1$ and $\Lambda_n = \sum_{i=0}^{n-1} \binom{n}{i} \Lambda_i$ using the Leibniz rule and the simple bound $\left\| \frac{\partial^{\mathfrak{u}} k(.,\mathbf{y})}{k(.,\mathbf{y})} \right\| \le \prod_{j\ge 1} b_j^{\nu_j}$ (where $\nu_j = \delta_{j\in\mathfrak{u}}$).

- Finally prove by induction that $\Lambda_n \le \dfrac{n!}{(\log 2)^n}$

Quadrature Error (1D, Matérn covariance, rank-1 lattice rule)



Quadrature Error (1D, Matérn covariance, rank-1 lattice rule)

Rates

ν	σ^2	$\lambda_C = 0.1$	$\lambda_C = 1.0$
0.75	0.25	0.82	0.89
	1.00	0.64	0.83
	4.00	0.60	0.63
1.5	0.25	0.80	0.86
	1.00	0.66	0.73
	4.00	0.58	0.55

Partial Conclusions & Summary

- MC-type methods currently the only ones that do not suffer from curse of dimensionality (for non-smooth non-affine problems)
- Multilevel MC is **optimal**, i.e. same cost as deterministic solver
- Theory based on careful FE error analysis [recall Wed]
  (level-dependent approximations for better variance reduction)
- Quasi MC acceleration (with new s-independent theory!)
- MLMC and QMC are complementary $\Rightarrow$ MLQMC
  [Giles, Waterhouse, 2009], [Kuo, Schwab, Sloan, 2012],
  [Harbrecht, Peters, Siebenmorgen, 2013], ongoing for lognormal

Partial Conclusions & Summary

- MC-type methods currently the only ones that do not suffer from curse of dimensionality (for non-smooth non-affine problems)
- Multilevel MC is **optimal**, i.e. same cost as deterministic solver
- Theory based on careful FE error analysis [recall Wed] (level-dependent approximations for better variance reduction)
- Quasi MC acceleration (with new s-independent theory!)
- MLMC and QMC are complementary $\Rightarrow$ MLQMC [Giles, Waterhouse, 2009], [Kuo, Schwab, Sloan, 2012], [Harbrecht, Peters, Siebenmorgen, 2013], ongoing for lognormal

$\nu = \frac{1}{2}$	$d = 1$	2	3
MC	ε^{-3}	ε^{-4}	ε^{-5}
QMC	ε^{-3}	ε^{-4}	ε^{-5}
MLMC	ε^{-2}	ε^{-2}	ε^{-3}
MLQMC	ε^{-2}	ε^{-2}	ε^{-3}

ν suff. large	$d = 1$	2	3
MC	$\varepsilon^{-5/2}$	ε^{-3}	$\varepsilon^{-7/2}$
QMC	$\varepsilon^{-3/2}$	ε^{-2}	$\varepsilon^{-5/2}$
MLMC	ε^{-2}	ε^{-2}	ε^{-2}
MLQMC	ε^{-1}	ε^{-1}	$\varepsilon^{-7/4}$

Multilevel Markov Chain Monte Carlo

Inverse Problems – Bayesian Inference

- Model was parametrised by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (the **"prior"**).
  In the subsurface flow application with lognormal coefficients:
  $\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x)Z_j(\omega)$ and $\mathcal{P}(\mathbf{Z}_s) \eqsim (2\pi)^{-s/2} \prod_{j=1}^{s} \exp\left(-\frac{Z_j^2}{2}\right)$
- Usually also some **output data** F_{obs} available (e.g. pressure).
  To reduce uncertainty, incorporate F_{obs} (the **"posterior"**)

Inverse Problems – Bayesian Inference

- Model was parametrised by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (the **"prior"**).
  In the subsurface flow application with lognormal coefficients:
  $\log k \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x) Z_j(\omega)$ and $\mathcal{P}(\mathbf{Z}_s) \eqsim (2\pi)^{-s/2} \prod_{j=1}^{s} \exp\left(-\frac{Z_j^2}{2}\right)$

- Usually also some **output data** F_{obs} available (e.g. pressure).
  To reduce uncertainty, incorporate F_{obs} (the **"posterior"**)

Bayes' Theorem: (RHS computable! Proportionality factor $1/\mathcal{P}(F_{\text{obs}})$ not!)

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

Inverse Problems – Bayesian Inference

- Model was parametrised by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (the **"prior"**).
  In the subsurface flow application with lognormal coefficients:
  $$\log k \approx \sum_{j=1}^s \sqrt{\mu_j}\phi_j(x)Z_j(\omega) \quad \text{and} \quad \mathcal{P}(\mathbf{Z}_s) \eqsim (2\pi)^{-s/2} \prod_{j=1}^s \exp\left(-\frac{Z_j^2}{2}\right)$$

- Usually also some **output data** F_{obs} available (e.g. pressure).
  To reduce uncertainty, incorporate F_{obs} (the **"posterior"**)

Bayes' Theorem: (RHS computable! Proportionality factor $1/\mathcal{P}(F_{\text{obs}})$ not!)

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

- **Likelihood model** (e.g. Gaussian):
  $$\mathcal{L}_h(F_{\text{obs}} \,|\, \mathbf{Z}_s) \eqsim \exp(-\|F_{\text{obs}} - F_h(\mathbf{Z}_s)\|^2/\sigma_{\text{obs}}^2)$$
  $F_h(\mathbf{Z}_s)$ ... model response; σ_{obs} ... fidelity parameter (data error)

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

- Choose $\mathbf{Z}_s^0$.
- At state n generate proposal $\mathbf{Z}_s'$ from distribution $q^{\mathrm{RW}}(\mathbf{Z}_s' \mid \mathbf{Z}_s^n)$ (e.g. random walk or preconditioned random walk [Stuart et al]).
- Accept $\mathbf{Z}_s'$ as a sample with probability

$$\alpha^{h,s} = \min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')\, q^{\mathrm{RW}}(\mathbf{Z}_s^n \mid \mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)\, q^{\mathrm{RW}}(\mathbf{Z}_s' \mid \mathbf{Z}_s^n)}\right) = \overbrace{\min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)}\right)}^{\text{for reversible prop. dist.}}$$

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 $q^{\text{RW}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)$ (e.g. random walk or preconditioned random walk [Stuart et al]).
- Accept $\mathbf{Z}_s'$ as a sample with probability

$$\alpha^{h,s} = \min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')\, q^{\text{RW}}(\mathbf{Z}_s^n \,|\, \mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)\, q^{\text{RW}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)}\right) = \overbrace{\min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)}\right)}^{\text{for reversible prop. dist.}}$$

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

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

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

$$\alpha^{h,s} = \min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')\, q^{\mathrm{RW}}(\mathbf{Z}_s^n \,|\, \mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)\, q^{\mathrm{RW}}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)}\right) = \overbrace{\min\left(1, \frac{\pi^{h,s}(\mathbf{Z}_s')}{\pi^{h,s}(\mathbf{Z}_s^n)}\right)}^{\text{for reversible prop. dist.}}$$

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

Pros:

- Produces a Markov chain $\{\mathbf{Z}_s^n\}_{n\in\mathbb{N}}$, with $\mathbf{Z}_s^n \sim \pi^{h,s}$ as $n \to \infty$.

Cons:

- Evaluation of $\alpha^{h,s} = \alpha^{h,s}(\mathbf{Z}_s' \,|\, \mathbf{Z}_s^n)$ very expensive for small h.
- Acceptance rate $\alpha^{h,s}$ very low for large s $(< 10\%)$.
- ε–Cost $= \mathcal{O}(\varepsilon^{-2-\frac{d}{\gamma}})$ as above, **but** constant depends on $\alpha^{h,s}$ & 'burn-in'

Multilevel Markov Chain Monte Carlo

choose $h_\ell = h_{\ell-1}/2$ and $s_\ell > s_{\ell-1}$, and set $Q_\ell := Q_{h_\ell,s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

Multilevel Markov Chain Monte Carlo

choose $h_\ell = h_{\ell-1}/2$ and $s_\ell > s_{\ell-1}$, and set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

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

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

Multilevel Markov Chain Monte Carlo

choose $h_\ell = h_{\ell-1}/2$ and $s_\ell > s_{\ell-1}$, and set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

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

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

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

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

Multilevel Markov Chain Monte Carlo

choose $h_\ell = h_{\ell-1}/2$ and $s_\ell > s_{\ell-1}$, and set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

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

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

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

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

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

Multilevel Markov Chain Monte Carlo

choose $h_\ell = h_{\ell-1}/2$ and $s_\ell > s_{\ell-1}$, and set $Q_\ell := Q_{h_\ell, s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

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

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

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

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

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

Split $\mathbf{Z}_\ell^n = [\mathbf{Z}_{\ell,\mathrm{C}}^n, \mathbf{Z}_{\ell,\mathrm{F}}^n] = \boxed{Z_{\ell,1}^n, \ldots\text{coarse}\ldots, Z_{\ell,s_{\ell-1}}^n\,,\ Z_{\ell,s_{\ell-1}+1}^n, \ldots\text{fine}\ldots, Z_{\ell,s_\ell}^n}$

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

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

1. On level $\ell - 1$: Generate new state $\mathbf{z}^{n+1}_{\ell-1}$ using Algorithm 1.

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

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

1. On level $\ell - 1$: Generate new state $\mathbf{z}^{n+1}_{\ell-1}$ using Algorithm 1.
2. On level ℓ: Propose $\mathbf{Z}'_\ell = [\mathbf{z}^{n+1}_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$ with $\mathbf{Z}'_{\ell,\mathrm{F}}$ as before
   (e.g. generated via a Crank-Nicholson preconditioned random walk)
   Novel transition prob. q^{ML} depends acceptance prob. $\alpha^{\ell-1}$ on level $\ell - 1$!

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

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

1. On level $\ell - 1$: Generate new state $\mathbf{z}^{n+1}_{\ell-1}$ using Algorithm 1.
2. On level ℓ: Propose $\mathbf{Z}'_\ell = [\mathbf{z}^{n+1}_{\ell-1}, \mathbf{Z}'_{\ell,\mathsf{F}}]$ with $\mathbf{Z}'_{\ell,\mathsf{F}}$ as before (e.g. generated via a Crank-Nicholson preconditioned random walk)

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

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

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

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

1. On level $\ell-1$: Generate new state $\mathbf{z}^{n+1}_{\ell-1}$ using Algorithm 1.
2. On level ℓ: Propose $\mathbf{Z}'_\ell = [\mathbf{z}^{n+1}_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$ with $\mathbf{Z}'_{\ell,\mathrm{F}}$ as before
   (e.g. generated via a Crank-Nicholson preconditioned random walk)
   Novel transition prob. q^{ML} depends acceptance prob. $\alpha^{\ell-1}$ on level $\ell-1$!
3. Accept $\mathbf{Z}'_\ell$ with probability

$$\alpha^\ell_{\mathrm{F}}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) = \min\left(1, \frac{\pi^\ell(\mathbf{Z}'_\ell)\pi^{\ell-1}(\mathbf{Z}^n_{\ell,\mathrm{C}})}{\pi^\ell(\mathbf{Z}^n_\ell)\pi^{\ell-1}(\mathbf{z}^{n+1}_{\ell-1})}\right)$$

   where $\mathbf{Z}^n_{\ell,\mathrm{C}}$ are the coarse modes of $\mathbf{Z}^n_\ell$ (from the chain on level ℓ).

This follows quite easily & both level $\ell-1$ terms have been computed before.

Multilevel MCMC Theory (What can we prove?)

[Ketelsen, RS, Teckentrup, arXiv:1303.7343, March 2013]

- We have genuine **Markov chains** on all levels.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since the two chains $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are independent on each level.
- Multilevel algorithm **converges** for any initial state.

Multilevel MCMC Theory (What can we prove?)

[Ketelsen, RS, Teckentrup, arXiv:1303.7343, March 2013]

- We have genuine **Markov chains** on all levels.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since the two chains $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are independent on each level.
- Multilevel algorithm **converges** for any initial state.
- **But** coarse modes may differ between level ℓ and $\ell - 1$ states:

State $n+1$	Level $\ell - 1$	Level ℓ
accept/accept	$\mathbf{z}'_{\ell-1}$	$[\mathbf{z}'_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$
reject/accept	$\mathbf{z}^n_{\ell-1}$	$[\mathbf{z}^n_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$
accept/reject	$\mathbf{z}'_{\ell-1}$	$[\mathbf{Z}^n_{\ell,\mathrm{C}}, \mathbf{Z}^n_{\ell,\mathrm{F}}]$
reject/reject	$\mathbf{z}^n_{\ell-1}$	$[\mathbf{Z}^n_{\ell,\mathrm{C}}, \mathbf{Z}^n_{\ell,\mathrm{F}}]$

Multilevel MCMC Theory (What can we prove?)

[Ketelsen, RS, Teckentrup, arXiv:1303.7343, March 2013]

- We have genuine **Markov chains** on all levels.
- Multilevel algorithm is **consistent** (= no bias between levels)
  since the two chains $\{\mathbf{Z}_\ell^n\}_{n\geq 1}$ and $\{\mathbf{z}_\ell^n\}_{n\geq 1}$ are independent on each level.
- Multilevel algorithm **converges** for any initial state.
- **But** coarse modes may differ between level ℓ and $\ell - 1$ states:

State $n+1$	Level $\ell - 1$	Level ℓ
accept/accept	$\mathbf{z}'_{\ell-1}$	$[\mathbf{z}'_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$
reject/accept	$\mathbf{z}^n_{\ell-1}$	$[\mathbf{z}^n_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$
accept/reject	$\mathbf{z}'_{\ell-1}$	$[\mathbf{Z}^n_{\ell,\mathrm{C}}, \mathbf{Z}^n_{\ell,\mathrm{F}}]$
reject/reject	$\mathbf{z}^n_{\ell-1}$	$[\mathbf{Z}^n_{\ell,\mathrm{C}}, \mathbf{Z}^n_{\ell,\mathrm{F}}]$

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

Complexity Theorem for Multilevel MCMC

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

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

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

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

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

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

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

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

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

Verifying (M1-M4) for the subsurface flow problem

with standard FEs & Fréchet-diff'ble functionals

Verifying (M1-M4) for the subsurface flow problem

with standard FEs & Fréchet-diff'ble functionals

- First split bias into truncation and discretization error:

$$
\begin{aligned}
|\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]| \;\le\; & |\mathbb{E}_{\pi^\ell}[Q_\ell - Q(\mathbf{Z}_\ell)]| && \textbf{(M1a)} \\
& + \; |\mathbb{E}_{\pi^\ell}[Q(\mathbf{Z}_\ell)] - \mathbb{E}_{\pi^\infty}[Q]| && \textbf{(M1b)}
\end{aligned}
$$

Verifying (M1-M4) for the subsurface flow problem

with standard FEs & Fréchet-diff'ble functionals

- First split bias into truncation and discretization error:

$$
\begin{aligned}
\left|\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]\right| \;&\le\; \left|\mathbb{E}_{\pi^\ell}[Q_\ell - Q(\mathbf{Z}_\ell)]\right| && \textbf{(M1a)} \\
&+\; \left|\mathbb{E}_{\pi^\ell}[Q(\mathbf{Z}_\ell)] - \mathbb{E}_{\pi^\infty}[Q]\right| && \textbf{(M1b)}
\end{aligned}
$$

- For **M1a** use $\mathbb{E}_{\pi^\ell}\left[|X|^q\right] \lesssim \mathbb{E}_{\mathcal{P}_\ell}\left[|X|^q\right]$ (prior bounds posterior) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \;\lesssim\; h_\ell^{2tq-\delta}$ (as before) $\;\Rightarrow\; \alpha < 2t$

Verifying (M1-M4) for the subsurface flow problem

with standard FEs & Fréchet-diff'ble functionals

- First split bias into truncation and discretization error:

$$|\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]| \le |\mathbb{E}_{\pi^\ell}[Q_\ell - Q(\mathbf{Z}_\ell)]| \quad \textbf{(M1a)}$$
$$+ \; |\mathbb{E}_{\pi^\ell}[Q(\mathbf{Z}_\ell)] - \mathbb{E}_{\pi^\infty}[Q]| \quad \textbf{(M1b)}$$

- For **M1a** use $\mathbb{E}_{\pi^\ell}\left[|X|^q\right] \lesssim \mathbb{E}_{\mathcal{P}_\ell}\left[|X|^q\right]$ (prior bounds posterior) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \lesssim h_\ell^{2tq-\delta}$ (as before) $\Rightarrow$ $\alpha < 2t$

- **M1b** & **M2** **not** specific to multilevel MCMC
  - **M1a** requires bound on truncation error in posterior (cf. [Hoang, Schwab, Stuart, '13] for uniformly elliptic/bdd. case)

Verifying (M1-M4) for the subsurface flow problem

with standard FEs & Fréchet-diff'ble functionals

- First split bias into truncation and discretization error:

$$\begin{aligned} |\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]| \;&\le\; |\mathbb{E}_{\pi^\ell}[Q_\ell - Q(\mathbf{Z}_\ell)]| && \textbf{(M1a)} \\ &+\; |\mathbb{E}_{\pi^\ell}[Q(\mathbf{Z}_\ell)] - \mathbb{E}_{\pi^\infty}[Q]| && \textbf{(M1b)} \end{aligned}$$

- For **M1a** use $\mathbb{E}_{\pi^\ell}\left[|X|^q\right] \lesssim \mathbb{E}_{\mathcal{P}_\ell}\left[|X|^q\right]$ (prior bounds posterior) &

$\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \;\lesssim\; h_\ell^{2tq-\delta}$ (as before) $\Rightarrow\; \alpha < 2t$

- **M1b** & **M2** **not** specific to multilevel MCMC
  - **M1a** requires bound on truncation error in posterior
    (cf. [Hoang, Schwab, Stuart, '13] for uniformly elliptic/bdd. case)
  - first steps to prove **M2** are in [Hairer, Stuart, Vollmer, '11]

(but both are still unproved so far!)

Verifying (M1-M4) for the subsurface flow problem

with standard FEs & Fréchet-diff'ble functionals

- First split bias into truncation and discretization error:

$$
\begin{aligned}
|\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]| \;\le\; & |\mathbb{E}_{\pi^\ell}[Q_\ell - Q(\mathbf{Z}_\ell)]| && \textbf{(M1a)} \\
& + |\mathbb{E}_{\pi^\ell}[Q(\mathbf{Z}_\ell)] - \mathbb{E}_{\pi^\infty}[Q]| && \textbf{(M1b)}
\end{aligned}
$$

- For **M1a** use $\mathbb{E}_{\pi^\ell}[|X|^q] \lesssim \mathbb{E}_{\mathcal{P}_\ell}[|X|^q]$ (prior bounds posterior) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \;\lesssim\; h_\ell^{2tq-\delta}$ (as before) $\Rightarrow$ $\alpha < 2t$

- **M1b** & **M2** **not** specific to multilevel MCMC
  - **M1a** requires bound on truncation error in posterior (cf. [Hoang, Schwab, Stuart, '13] for uniformly elliptic/bdd. case)
  - first steps to prove **M2** are in [Hairer, Stuart, Vollmer, '11]

  (but both are still unproved so far!)

- **M4** holds (with suitable multigrid solver – proved only for low contrast)

Key assumption for multilevel MCMC is (M3)

Key Lemma (given only for the 1-norm exponential here)

Assume F^h Fréchet differentiable & sufficiently smooth. Then

$$\lim_{\ell \to \infty} \alpha_{\mathsf{F}}^{\ell}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) = 1, \qquad \text{for } \mathcal{P}_\ell\text{-almost all } \mathbf{Z}'_\ell, \mathbf{Z}^n_\ell,$$

and

$$\mathbb{E}_{\mathcal{P}_\ell}\left[(1 - \alpha_{\mathsf{F}}^{\ell})^q\right]^{1/q} \lesssim h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta} \quad \forall q < \infty\,, \; \delta > 0.$$

Key assumption for multilevel MCMC is (M3)

Key Lemma (given only for the 1-norm exponential here)

Assume F^h Fréchet differentiable & sufficiently smooth. Then

$$\lim_{\ell\to\infty} \alpha_F^\ell(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) = 1, \qquad \text{for } \mathcal{P}_\ell\text{-almost all } \mathbf{Z}'_\ell, \mathbf{Z}^n_\ell,$$

and

$$\mathbb{E}_{\mathcal{P}_\ell}\left[(1-\alpha_F^\ell)^q\right]^{1/q} \lesssim h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta} \quad \forall q < \infty\,,\ \delta > 0.$$

Key assumption for multilevel MCMC is (M3)

Key Lemma (given only for the 1-norm exponential here)

Assume F^h Fréchet differentiable & sufficiently smooth. Then

$$\lim_{\ell \to \infty} \alpha_{\mathrm{F}}^{\ell}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) = 1, \qquad \text{for } \mathcal{P}_\ell\text{-almost all } \mathbf{Z}'_\ell, \mathbf{Z}^n_\ell,$$

and

$$\mathbb{E}_{\mathcal{P}_\ell}\left[(1 - \alpha_{\mathrm{F}}^{\ell})^q\right]^{1/q} \lesssim h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta} \quad \forall q < \infty\,, \ \delta > 0.$$

Lemma (again only for 1-norm exponential)

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

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

and **M3** holds for any $\beta < 1$. ($\beta \neq 2\alpha$ as in "standard" MLMC!)

Numerical Example

$D = (0,1)^2$, exponential covar. with $\sigma^2 = 1$ & $\lambda = 0.5$, $Q = \int_{\Gamma_{\text{out}}} \vec{q} \cdot \vec{n}$, $h_0 = \frac{1}{16}$

Numerical Example

$D = (0,1)^2$, exponential covar. with $\sigma^2 = 1$ & $\lambda = 0.5$, $Q = \int_{\Gamma_{\text{out}}} \vec{q} \cdot \vec{n}$, $h_0 = \frac{1}{16}$

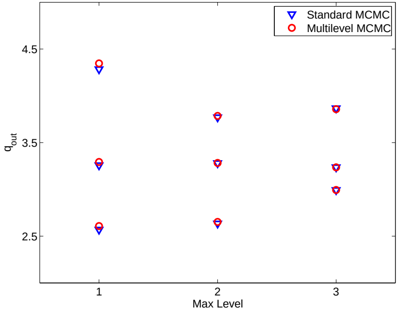
- **"Data"** F_{obs}: Pressure $p(x^*)$ at 9 random points $x^* \in D$.
- # modes: $s_0 = 96$, $s_1 = 121$, $s_2 = 153$ and $s_3 = 169$

Numerical Example

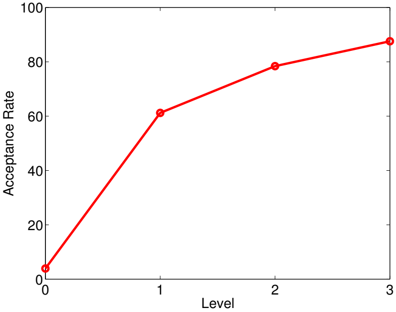
$D = (0,1)^2$, exponential covar. with $\sigma^2 = 1$ & $\lambda = 0.5$, $Q = \int_{\Gamma_{out}} \vec{q} \cdot \vec{n}$, $h_0 = \frac{1}{16}$

- **"Data"** F_{obs}: Pressure $p(x^*)$ at 9 random points $x^* \in D$.
- # modes: $s_0 = 96$, $s_1 = 121$, $s_2 = 153$ and $s_3 = 169$

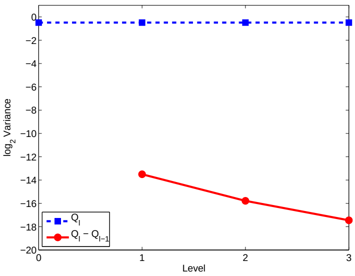
Comparison single- vs. multi-level



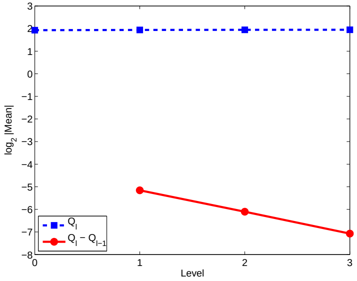
Acceptance rate α_F^ℓ in multilevel estim.



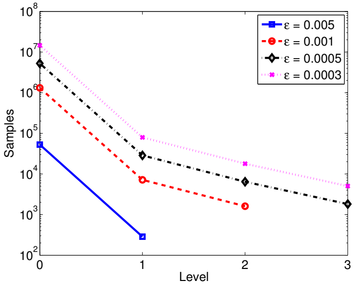
Variance



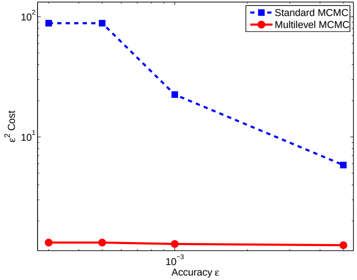
Mean



samples



scaled cost



Additional Comments

- In all tests we got consistent gains of **a factor** $O(10 - 100)$**!**
- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state (and variance estimator suggested by [Gelman & Rubin, 1992])

Additional Comments

- In all tests we got consistent gains of **a factor** $O(10 - 100)$**!**
- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state (and variance estimator suggested by [Gelman & Rubin, 1992])
- Improved **multilevel burn-in** also possible ($\sim 10\times$ cheaper!) (related to two-level work in [Efendiev, Hou, Luo, 2005])

Additional Comments

- In all tests we got consistent gains of **a factor** $O(10 - 100)$**!**
- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state (and variance estimator suggested by [Gelman & Rubin, 1992])
- Improved **multilevel burn-in** also possible ($\sim 10\times$ cheaper!) (related to two-level work in [Efendiev, Hou, Luo, 2005])
- Practically observe $\beta \approx 2\alpha$ (small constant in leading order term)
- Similar results for Matérn covariance and **mixed FEs**

Additional Comments

- In all tests we got consistent gains of **a factor** $O(10 - 100)$**!**
- Using a special "preconditioned" random walk to be dimension independent (Assumption **M2**) from [Cotter, Dashti, Stuart, 2012]
- Using multiple chains to reduce dependence on initial state (and variance estimator suggested by [Gelman & Rubin, 1992])
- Improved **multilevel burn-in** also possible ($\sim 10\times$ cheaper!) (related to two-level work in [Efendiev, Hou, Luo, 2005])
- Practically observe $\beta \approx 2\alpha$ (small constant in leading order term)
- Similar results for Matérn covariance and **mixed FEs**
- Related theoretical work by [Hoang, Schwab, Stuart, 2013] (different multilevel splitting and so far no numerics to compare)

Conclusions on MCMC Part

- "Real" UQ involves incorporating data – Bayesian inference
- Multilevel MC idea extends to Markov chain Monte Carlo
  (with theory for lognormal subsurface model problem)
- As spectacular gains in practice as for standard MLMC!

Conclusions on MCMC Part

- “Real” UQ involves incorporating data – Bayesian inference
- Multilevel MC idea extends to Markov chain Monte Carlo
  (with theory for lognormal subsurface model problem)
- As spectacular gains in practice as for standard MLMC!

Future Work & Open Questions

- More numerical tests and real comparisons with other methods
- 3D, parallelisation, HPC, application to real problems
- Circulant embedding & PDE based sampling instead (+theory)
- Multilevel QMC theory for lognormal case
- Application of multilevel MCMC in other areas (statisticians!)
  other (nonlinear) PDEs, big data applications, molecular dynamics, DA
- Multilevel methods for rare events – “subset simulation”

Thank You!

Most of the material I used is available from my website:

http://people.bath.ac.uk/∼masrs/publications.html