Multilevel Markov Chain Monte Carlo
with Applications in Subsurface Flow

Robert Scheichl

Department of Mathematical Sciences

UNIVERSITY OF BATH

Collaborators:

AL Teckentrup (Warwick) & **C Ketelsen** (Boulder)

Thanks also to my Bath colleagues F. Lindgren (Stats) & R. Jack (Physics)

Workshop on "Stochastic and Multiscale Inverse Problems"

October 2nd-3rd 2014, Ecole des Ponts Paristech, Paris

Introduction

- Many problems involve **PDEs** with spatially varying **data** which is subject to **uncertainty**.

  Example: **groundwater flow** in rock underground.

- **Uncertainty** enters PDE via its coefficients **(random fields)**. The **quantity of interest**: is a random number or field derived from the PDE solution.

  Examples: **effective permeability** or
  breakthrough time of a pollution plume

- **Typical Computational Goal: expected value of quantity of interest**. (**Uncertainty quantification**)

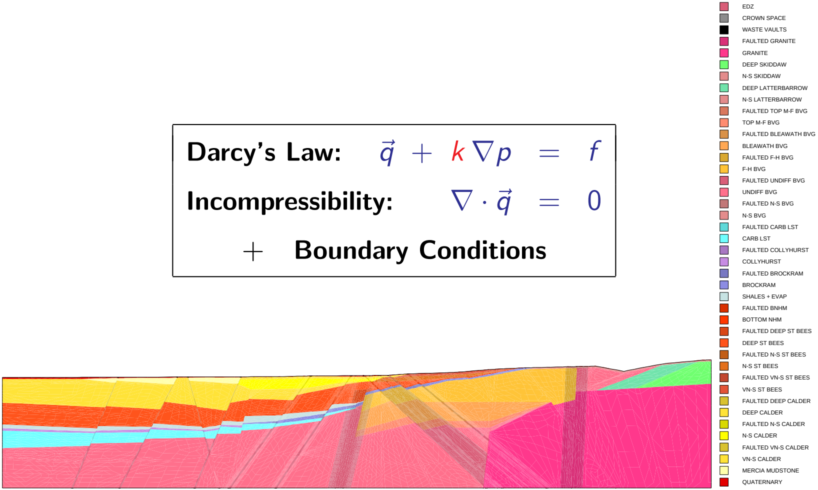
Uncertainty Propagation

The Forward Problem

Example: Uncertainty in Subsurface Flow

(eg. risk analysis of radwaste disposal or optimisation of oil recovery)

Darcy's Law:	$\vec{q}$	$+$	$k\,\nabla p$	$=$	f
Incompressibility:			$\nabla \cdot \vec{q}$	$=$	0
$+$ **Boundary Conditions**					



Rock strata at Sellafield (potential UK radwaste site in 90s) ©NIREX UK Ltd

Example: Uncertainty in Subsurface Flow

(eg. risk analysis of radwaste disposal or optimisation of oil recovery)

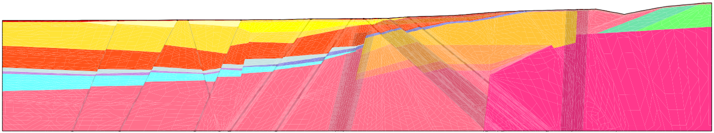
uncertain k $\rightarrow$

Darcy's Law: $\vec{q} + k\nabla p = f$

Incompressibility: $\nabla \cdot \vec{q} = 0$

$+$ **Boundary Conditions**

$\rightarrow$ uncertain p, $\vec{q}$



Rock strata at Sellafield (potential UK radwaste site in 90s) ©NIREX UK Ltd

Stochastic Modelling of Uncertainty:

Model uncertain conductivity tensor k as a **lognormal** random field

Typical simplified model (prior):

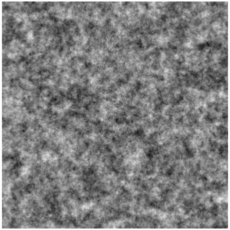
- $\log k(x,\omega)$ isotropic, scalar, **Gaussian**

  e.g. meanfree with exponential covariance

  $$R(x,y) := \sigma^2 \exp\left(-\|x-y\|/\lambda\right)$$

- e.g. truncated Karhunen-Loève expansion

  $$\log k(x,\omega) \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x)Z_j(\omega), \quad Z_j(\omega) \text{ iid } N(0,\sigma^2)$$



typical realisation

$(\lambda = \frac{1}{64}, \sigma^2 = 8)$

Stochastic Modelling of Uncertainty:

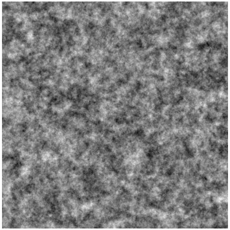
Model uncertain conductivity tensor k as a **lognormal** random field

Typical simplified model (prior):

- $\log k(x,\omega)$ isotropic, scalar, **Gaussian**

  e.g. meanfree with exponential covariance

$$R(x,y) := \sigma^2 \exp\left(-\|x-y\|/\lambda\right)$$



typical realisation

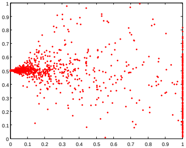
$(\lambda = \frac{1}{64}, \sigma^2 = 8)$

- e.g. truncated Karhunen-Loève expansion

$$\log k(x,\omega) \approx \sum_{j=1}^{s} \sqrt{\mu_j}\phi_j(x) Z_j(\omega), \quad Z_j(\omega) \text{ iid } N(0,\sigma^2)$$

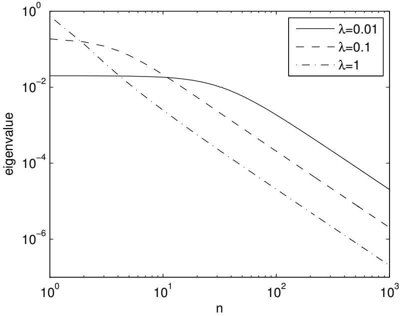
Typical quantities of interest:

- $p(x^*)$, $\vec{q}(x^*)$, travel time, water cut,. . .
- outflow through Γ_{out}: $Q_{\text{out}} = \int_{\Gamma_{\text{out}}} \vec{q} \cdot d\vec{n}$

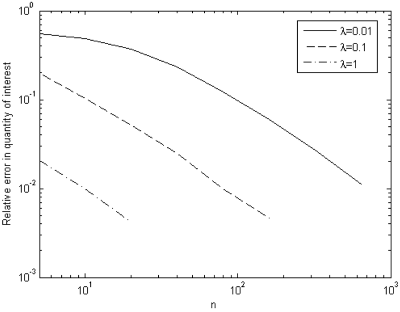


Why is this problem so challenging?

Why is this problem so challenging?

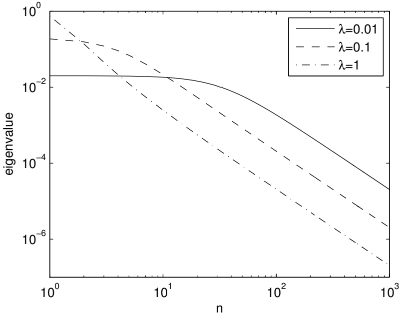


KL-eigenvalues in 1D

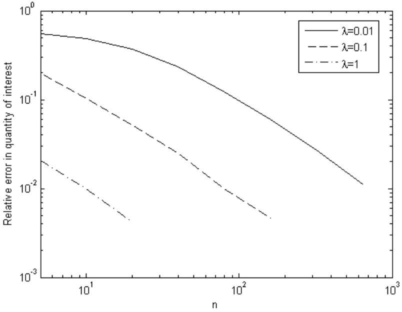


Convergence of $q|_{x=1}$ w.r.t. s

Why is this problem so challenging?



KL-eigenvalues in 1D



Convergence of $q|_{x=1}$ w.r.t. s

- Small correlation length λ $\implies$ **high** dimension $s \gg 10$
  and **fine** mesh $h \ll 1$
- Large σ^2 & exponential $\implies$ **large** heterogeneity $\frac{k_{\max}}{k_{\min}} > 10^6$

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)

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

$$\underbrace{\mathbb{E}\left[\left(\hat{Q}^{\mathrm{MC}} - \mathbb{E}[Q]\right)^2\right]}_{=:\ \mathsf{MSE}} \quad = \quad \underbrace{\frac{\mathbb{V}[Q_{h,s}]}{N}}_{\textbf{sampling error}} \ + \ \underbrace{\left(\mathbb{E}[Q_{h,s} - Q]\right)^2}_{\textbf{model error (“bias”)}}$$

- Typical (2D): $\alpha = 1 \ \Rightarrow \ \mathsf{MSE} = \mathcal{O}(N^{-1}) + \mathcal{O}(M_h^{-1}) = \ \mathcal{O}(\varepsilon^2)$

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

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

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

- Thus $M_h \sim N \sim \varepsilon^{-2}$ and $\mathrm{Cost} = \mathcal{O}(NM_h) = \mathcal{O}(\varepsilon^{-4})$ (w. MG solver)

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

- Quickly becomes **prohibitively expensive !**

Complexity Theorem for (plain vanilla) Monte Carlo

Assume that $\mathbb{E}[Q_{h,s}] \to \mathbb{E}[Q]$ with $\mathcal{O}(h^\alpha)$ and cost per sample is $\mathcal{O}(h^{-\gamma})$. Then

$$\mathrm{Cost}(\hat{Q}^{\mathrm{MC}}) \;=\; \mathcal{O}\big(\varepsilon^{-2-\frac{\gamma}{\alpha}}\big) \quad \text{to obtain} \quad \mathrm{MSE} = \mathcal{O}(\varepsilon^2).$$

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]

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: $\approx \mathcal{O}(M_h) = \mathcal{O}(h^{-2}) \implies \gamma \approx 2$.

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: $\approx \mathcal{O}(M_h) = \mathcal{O}(h^{-2}) \implies \gamma \approx 2$.
- **Total cost** to get RMSE $\mathcal{O}(\varepsilon)$: $\approx \mathcal{O}(\varepsilon^{-14/3})$
  to get error reduction by a factor 2 → cost grows by a factor 25!

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

ε	h^{-1}	N	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	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 Stochastic Solvers

Multilevel Monte Carlo [Heinrich, '01], [Giles, '07] [Barth, Schwab, Zollinger, '11], [Cliffe, Giles, RS, Teckentrup, '11]

Note that trivially

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

where $h_{\ell-1} = m h_\ell$ (hierarchy of grids) and $Q_\ell := Q_{h_\ell, s_\ell}$

Multilevel Monte Carlo [Heinrich, '01], [Giles, '07]

[Barth, Schwab, Zollinger, '11], [Cliffe, Giles, RS, Teckentrup, '11]

Note that trivially

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

where $h_{\ell-1} = m h_\ell$ (hierarchy of grids) and $Q_\ell := Q_{h_\ell, s_\ell}$

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

$$\widehat{Q}_L^{\mathcal{ML}} := \hat{Q}_0^{\mathrm{MC}} \; + \; \sum_{\ell=1}^{L} \hat{Y}_\ell^{\mathrm{MC}} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1}$$

Multilevel Monte Carlo [Heinrich, '01], [Giles, '07] [Barth, Schwab, Zollinger, '11], [Cliffe, Giles, RS, Teckentrup, '11]

Note that trivially

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

where $h_{\ell-1} = m h_\ell$ (hierarchy of grids) and $Q_\ell := Q_{h_\ell, s_\ell}$

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

$$\widehat{Q}_L^{\mathcal{ML}} := \hat{Q}_0^{\mathrm{MC}} \; + \; \sum_{\ell=1}^{L} \hat{Y}_\ell^{\mathrm{MC}} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1}$$

Key Observation: (Variance Reduction! Corrections cheaper!)

If $Q_\ell \to Q$ then $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$ **!**

Complexity Theorem for Multilevel Monte Carlo

Assume FE error $\mathcal{O}(h^\alpha)$ and Cost/sample $\mathcal{O}(h^{-\gamma})$ (as above) **as well as**

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(h_\ell^\beta) \qquad \text{(variance reduction)}.$$

There exist L, $\{N_\ell\}_{\ell=0}^L$ (computable on the fly) to obtain MSE $< \varepsilon^2$ with

$$\mathrm{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \; = \; \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) \; + \text{ possible log-factor}$$

(**Note.** This is completely **abstract**! Applies also in other applications!)

Complexity Theorem for Multilevel Monte Carlo

Assume FE error $\mathcal{O}(h^\alpha)$ and Cost/sample $\mathcal{O}(h^{-\gamma})$ (as above) **as well as**

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(h_\ell^\beta) \qquad \text{(variance reduction)}.$$

There exist L, $\{N_\ell\}_{\ell=0}^L$ (computable on the fly) to obtain MSE $< \varepsilon^2$ with

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) + \text{possible log-factor}$$

(**Note.** This is completely **abstract**! Applies also in other applications!)

If $\beta \sim 2\alpha$ and $\gamma \approx d$ (as in example above with AMG) then

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\varepsilon^{-\max\left(2,\frac{d}{\alpha}\right)}\right) \;=\; \mathcal{O}\left(\max(N_0, M_L)\right)$$

For $\alpha \approx 3/4$ (in example above): $\mathcal{O}(\varepsilon^{-8/3})$ instead of $\mathcal{O}(\varepsilon^{-14/3})$

Complexity Theorem for Multilevel Monte Carlo

Assume FE error $\mathcal{O}(h^\alpha)$ and Cost/sample $\mathcal{O}(h^{-\gamma})$ (as above) **as well as**

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(h_\ell^\beta) \qquad \text{(variance reduction).}$$

There exist L, $\{N_\ell\}_{\ell=0}^L$ (computable on the fly) to obtain MSE $< \varepsilon^2$ with

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) \quad + \text{ possible log-factor}$$

(**Note.** This is completely **abstract**! Applies also in other applications!)

If $\beta \sim 2\alpha$ and $\gamma \approx d$ (as in example above with AMG) then

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\varepsilon^{-\max\left(2,\frac{d}{\alpha}\right)}\right) \;=\; \mathcal{O}\left(\max(N_0, M_L)\right)$$

For $\alpha \approx 3/4$ (in example above): $\mathcal{O}(\varepsilon^{-8/3})$ instead of $\mathcal{O}(\varepsilon^{-14/3})$

Optimality: Same asymptotic cost as **one deterministic solve** (tol$=\varepsilon$) **!**

Numerical Example (Multilevel MC)

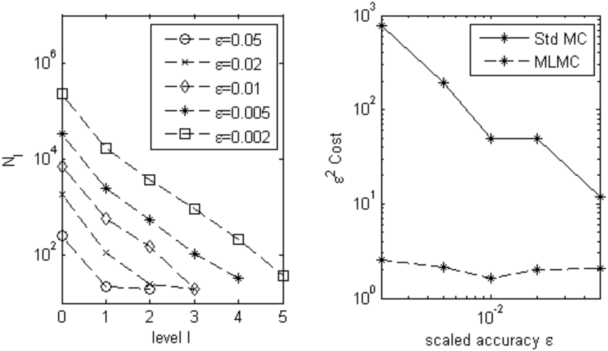
$D = (0,1)^2$; covariance $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|_2}{\lambda}\right)$; $Q = \|p\|_{L_2(D)}$

Std. FE discretisation, circulant embedding $(s_\ell = \mathcal{O}(M_\ell))$

Numerical Example (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 $(s_\ell = \mathcal{O}(M_\ell))$



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

Analysis: Verifying Assumptions of Complexity Theorem

Analysis: Verifying Assumptions of Complexity Theorem

- **[Barth, Schwab, Zollinger, 2011]**: case of **uniformly elliptic** and **bounded** $k(\cdot,\omega) \in W^{1,\infty}(D)$ (**not** satisfied here!)
- **[Charrier, RS, Teckentrup, 2013]**: **lognormal** $k \Rightarrow$ not uniformly elliptic/bdd. **and** only $k(\cdot,\omega) \in C^{0,\eta}(D)$, with $\eta < 1/2$ (exponen. covar.)

Analysis: Verifying Assumptions of Complexity Theorem

- **[Barth, Schwab, Zollinger, 2011]**: case of **uniformly elliptic** and **bounded** $k(\cdot,\omega) \in W^{1,\infty}(D)$ (**not** satisfied here!)
- **[Charrier, RS, Teckentrup, 2013]**: **lognormal** $k \Rightarrow$ not uniformly elliptic/bdd. **and** only $k(\cdot,\omega) \in C^{0,\eta}(D)$, with $\eta < 1/2$ (exponen. covar.)
- New **regularity** result: (q-th moment of H^{1+t}-norm)

$$\|p\|_{L^q(\Omega,H^{1+t}(D))} \leq C_{t,q}\|f\|_{L^2(D)}, \ \forall\, t < 1/2, \ q < \infty.$$

- New **FE error** result: (q-th moment of H^1-norm)

$$\|p - p_h\|_{L^q(\Omega,H^1(D))} \leq C'_{t,q}\|f\|_{L^2(D)}\, h^t, \ \forall\, t < 1/2, \ q < \infty.$$

Analysis: Verifying Assumptions of Complexity Theorem

- **[Barth, Schwab, Zollinger, 2011]**: case of **uniformly elliptic** and **bounded** $k(\cdot,\omega) \in W^{1,\infty}(D)$ (**not** satisfied here!)
- **[Charrier, RS, Teckentrup, 2013]**: **lognormal** $k \Rightarrow$ not uniformly elliptic/bdd. **and** only $k(\cdot,\omega) \in C^{0,\eta}(D)$, with $\eta < 1/2$ (exponen. covar.)
- New **regularity** result: (q-th moment of H^{1+t}-norm)

$$\|p\|_{L^q(\Omega,H^{1+t}(D))} \leq C_{t,q}\|f\|_{L^2(D)}, \ \forall\, t < 1/2, \ q < \infty.$$

- New **FE error** result: (q-th moment of H^1-norm)

$$\|p - p_h\|_{L^q(\Omega,H^1(D))} \leq C'_{t,q}\|f\|_{L^2(D)}\, h^t, \ \forall\, t < 1/2, \ q < \infty.$$

- **[Teckentrup, RS, Giles, Ullmann, 2013]**: (nonlinear) functionals, corner domains, discontinuous coefficients, level-dependent truncations
- **[Teckentrup, 2013]**: L^∞-, $W^{1,\infty}$-norms, random interfaces,...
- **[Graham, RS, Ullmann, 2013]**: extension to mixed FEs

Analysis: Verifying Assumptions of Complexity Theorem

- **[Barth, Schwab, Zollinger, 2011]**: case of **uniformly elliptic** and **bounded** $k(\cdot,\omega) \in W^{1,\infty}(D)$ (**not** satisfied here!)
- **[Charrier, RS, Teckentrup, 2013]**: **lognormal** k $\Rightarrow$ not uniformly elliptic/bdd. **and** only $k(\cdot,\omega) \in C^{0,\eta}(D)$, with $\eta < 1/2$ (exponen. covar.)
- New **regularity** result: (q-th moment of H^{1+t}-norm)

$$\|p\|_{L^q(\Omega,H^{1+t}(D))} \le C_{t,q}\|f\|_{L^2(D)}, \ \forall\, t < 1/2,\ q < \infty.$$

- New **FE error** result: (q-th moment of H^1-norm)

$$\|p - p_h\|_{L^q(\Omega,H^1(D))} \le C'_{t,q}\|f\|_{L^2(D)}\ h^t, \ \forall\, t < 1/2,\ q < \infty.$$

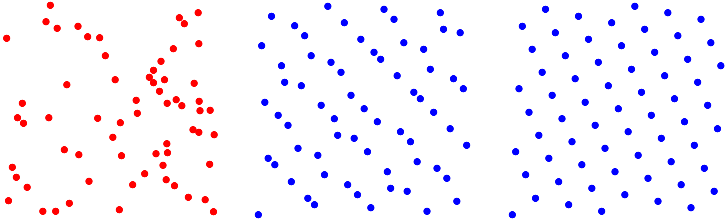
- **[Teckentrup, RS, Giles, Ullmann, 2013]**: (nonlinear) functionals, corner domains, discontinuous coefficients, level-dependent truncations
- **[Teckentrup, 2013]**: L^∞-, $W^{1,\infty}$-norms, random interfaces,...
- **[Graham, RS, Ullmann, 2013]**: extension to mixed FEs

For Fréchet diff'ble functional $Q = \mathcal{G}(p)$, assumptions hold for any $\alpha < 1,\ \beta < 2$.

Reducing # Samples (Quasi–Monte Carlo)

[Graham, Kuo, Nuyens, RS, Sloan, '10] & [Gr., Ku., Nichols, RS, Schwab, Sl., '13]

random $\omega^{(i)}$ $\longrightarrow$ **deterministically chosen** $\widetilde{\omega}^{(i)}$

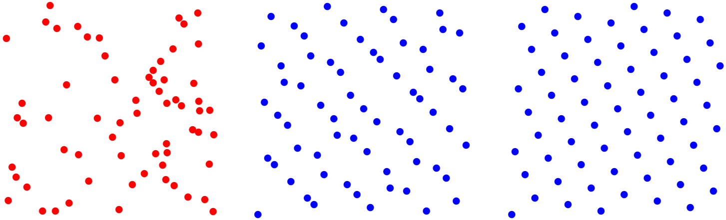


64 random points   64 Sobol′ points   64 lattice points

Reducing # Samples (Quasi–Monte Carlo)

[Graham, Kuo, Nuyens, RS, Sloan, '10] & [Gr., Ku., Nichols, RS, Schwab, Sl., '13]

random $\omega^{(i)}$ $\longrightarrow$ **deterministically chosen** $\widetilde{\omega}^{(i)}$



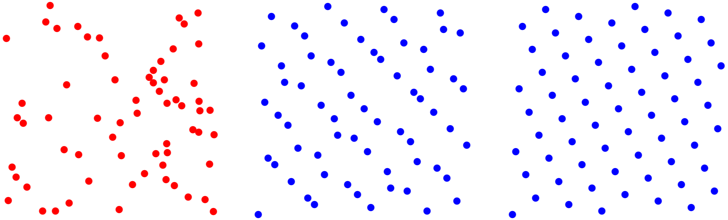
64 random points   64 Sobol$'$ points   64 lattice points

- Provided KL-eigenvalues decay sufficiently fast (e.g. Matérn): QMC estimator converges with $O(N^{-1})$ instead of $O(N^{-1/2})$
- **New rigorous theory** (for $s \to \infty$) in weighted Sobolev spaces

Reducing # Samples (Quasi–Monte Carlo)

[Graham, Kuo, Nuyens, RS, Sloan, '10] & [Gr., Ku., Nichols, RS, Schwab, Sl., '13]

random $\omega^{(i)}$ $\longrightarrow$ **deterministically chosen** $\widetilde{\omega}^{(i)}$



64 random points 64 Sobol′ points 64 lattice points

- Provided KL-eigenvalues decay sufficiently fast (e.g. Matérn): QMC estimator converges with $O(N^{-1})$ instead of $O(N^{-1/2})$
- **New rigorous theory** (for $s \to \infty$) in weighted Sobolev spaces
- **In practice** #samples (and thus cost) always **significantly smaller**

Multilevel Quasi-Monte Carlo (Gains complimentary!)

[Giles, Waterhouse '09] (SDE), [Kuo, Schwab, Sloan '12] (uniform affine), [Harbrecht et al, '13] (**lognormal**, but **not** *s*-independent & no efficiency gains)

Multilevel Quasi-Monte Carlo (Gains complimentary!)

[Giles, Waterhouse '09] (SDE), [Kuo, Schwab, Sloan '12] (uniform affine), [Harbrecht et al, '13] (**lognormal**, but **not** s-independent & no efficiency gains)

NEW: Complexity Theorem for **Multilevel QMC** (lognormal; $\mathcal{G}$ linear)
[Kuo, RS, Schwab, Sloan, Ullmann, in prep.]

Assume FE error $\mathcal{O}(h^\alpha)$ and Cost/sample $\mathcal{O}(h^{-\gamma})$ (as above) **as well as**

$$\mathbb{V}_\Delta\left[\mathcal{Q}^s_{N_\ell}(\mathcal{G}(p_\ell - p_{\ell-1}))\right] = \mathcal{O}(N_\ell^{-\eta} h_\ell^\beta), \quad \text{with } 1 \le \eta < 2.$$

There exist L, $\{N_\ell\}_{\ell=0}^L$ (computable on the fly) to obtain MSE $< \varepsilon^2$ with

$$\text{Cost}(\widehat{Q}_L^{\mathcal{MLQ}}) \;=\; \mathcal{O}\left(\varepsilon^{-\frac{2}{\eta} - \max\left(0, \frac{\eta\gamma - \beta}{\eta\alpha}\right)}\right) \quad + \text{ possible log's}$$

Multilevel Quasi-Monte Carlo (Gains complimentary!)

[Giles, Waterhouse '09] (SDE), [Kuo, Schwab, Sloan '12] (uniform affine), [Harbrecht et al, '13] (**lognormal**, but **not** s-independent & no efficiency gains)

NEW: Complexity Theorem for **Multilevel QMC** (lognormal; $\mathcal{G}$ linear)
[Kuo, RS, Schwab, Sloan, Ullmann, in prep.]

Assume FE error $\mathcal{O}(h^\alpha)$ and Cost/sample $\mathcal{O}(h^{-\gamma})$ (as above) **as well as**

$$\mathbb{V}_\Delta\left[\mathcal{Q}^s_{N_\ell}\big(\mathcal{G}(p_\ell - p_{\ell-1})\big)\right] = \mathcal{O}(N_\ell^{-\eta} h_\ell^\beta), \quad \text{with } 1 \le \eta < 2.$$

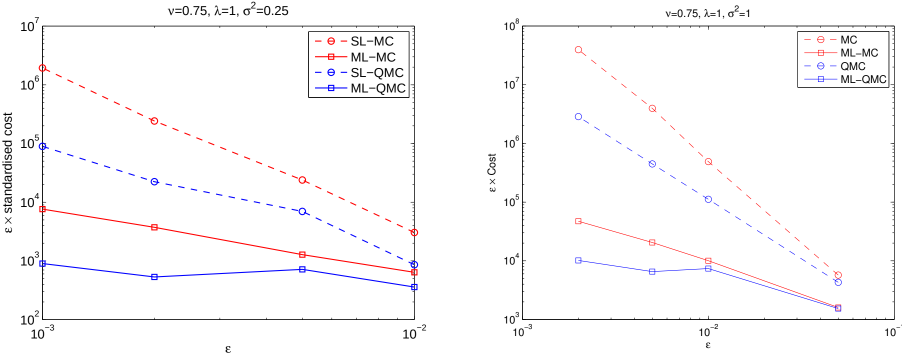
There exist L, $\{N_\ell\}_{\ell=0}^L$ (computable on the fly) to obtain MSE $< \varepsilon^2$ with

$$\text{Cost}(\widehat{Q}_L^{\mathcal{MLQ}}) \;=\; \mathcal{O}\left(\varepsilon^{-\frac{2}{\eta} - \max\left(0, \frac{\eta\gamma-\beta}{\eta\alpha}\right)}\right) \quad + \text{ possible log's}$$

- If $\eta \approx 2$, $\beta \approx 2\alpha$ and $\gamma \approx d$ then Cost $= \mathcal{O}\left(\varepsilon^{-\max\left(1, \frac{d}{\alpha}\right)}\right)$.
- **Better than MLMC** complexity for $\alpha > d/2$. **Optimal** for $\alpha \le d$!

Numerical Examples

$D = (0,1)^2$; mixed BCs; std. p.w. lin. FE discretisation; $Q = \int_0^1 k \nabla p \, dx_2$
Matérn covariance; truncated KLE w. $s = 400$; randomised lattice rule w. $\gamma_j = 1/j^2$



$\nu = 0.75,\ \lambda = 1,\ \sigma^2 = 0.25$ (left) and $\sigma^2 = 1$ (right)

Multilevel Markov Chain Monte Carlo

The Inverse Problem

Incorporating Data – Bayesian Inversion

- Model was parametrised by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (the **"prior"**).
  In the subsurface flow application a lognormal coefficient

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

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

Incorporating Data – Bayesian Inversion

- Model was parametrised by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (the **"prior"**).
  In the subsurface flow application a lognormal coefficient

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

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

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

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

Incorporating Data – Bayesian Inversion

- Model was parametrised by $\mathbf{Z}_s := [Z_1, \ldots, Z_s]$ (the **"prior"**).
  In the subsurface flow application a lognormal coefficient

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

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

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

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

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

  $$\mathcal{L}_h(F_{\mathrm{obs}} \,|\, \mathbf{Z}_s) \eqsim \exp(-\|F_{\mathrm{obs}} - F_h(\mathbf{Z}_s)\|^2/\sigma_{\mathrm{fid}}^2)$$

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

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

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

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

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

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

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

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

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

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

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

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

Comments on Metropolis-Hastings MCMC

Pros:

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

Comments on Metropolis-Hastings MCMC

Pros:

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

Cons:

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

Comments on Metropolis-Hastings MCMC

Pros:

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

Cons:

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

Prohibitively expensive – significantly more than plain-vanilla MC !

Multilevel Markov Chain Monte Carlo

choose $h_{\ell-1} = mh_\ell$ and $s_\ell \geq s_{\ell-1}$, and set $Q_\ell := Q_{h_\ell,s_\ell}$ and $\mathbf{Z}_\ell := \mathbf{Z}_{s_\ell}$

Multilevel Markov Chain Monte Carlo

choose $h_{\ell-1} = mh_\ell$ and $s_\ell \geq s_{\ell-1}$, 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

Multilevel Markov Chain Monte Carlo

choose $h_{\ell-1} = mh_\ell$ and $s_\ell \geq s_{\ell-1}$, 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:

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

Multilevel Markov Chain Monte Carlo

choose $h_{\ell-1} = m h_\ell$ and $s_\ell \geq s_{\ell-1}$, 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 $h_{\ell-1} = mh_\ell$ and $s_\ell \geq s_{\ell-1}$, 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 $\mathbf{Z}'_\ell = [\mathbf{z}^{n+1}_{\ell-1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$ 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

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

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

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

1. On level $\ell - 1$: Generate new state $\mathbf{z}_{\ell-1}^{n+1}$ using Algorithm 1.
2. On level ℓ: Propose $\mathbf{Z}'_\ell = [\mathbf{z}_{\ell-1}^{n+1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$ 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

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

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

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

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

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

Unfortunately we discovered an **error** in our proof, so that this **algorithm creates a small bias** in the fine-level posterior !! (not noticable in numerics)

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

At n^{th} state $\mathbf{Z}_\ell^n$ (of a Markov chain on level ℓ):

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

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

At n^{th} state $\mathbf{Z}_\ell^n$ (of a Markov chain on level ℓ):

1. On level $\ell - 1$: Generate an **independent** sample $\mathbf{z}_{\ell-1}^{n+1} \sim \pi^{\ell-1}$ (from coarse posterior)
2. On level ℓ: Propose $\mathbf{Z}'_\ell = [\mathbf{z}_{\ell-1}^{n+1}, \mathbf{Z}'_{\ell,\mathrm{F}}]$ with $\mathbf{Z}'_{\ell,\mathrm{F}}$ as before (e.g. generated via a Crank-Nicholson preconditioned random walk)

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

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

At n^{th} state $\mathbf{Z}_\ell^n$ (of a Markov chain on level ℓ):

1. On level $\ell - 1$: Generate an **independent** sample $\mathbf{z}_{\ell-1}^{n+1} \sim \pi^{\ell-1}$ (from coarse posterior)
2. On level ℓ: Propose $\mathbf{Z}'_\ell = [\mathbf{z}_{\ell-1}^{n+1}, \mathbf{Z}'_{\ell,\mathsf{F}}]$ with $\mathbf{Z}'_{\ell,\mathsf{F}}$ as before (e.g. generated via a Crank-Nicholson preconditioned random walk)
   Transition prob. q^{ML} depends on posterior on level $\ell - 1$!
3. Accept $\mathbf{Z}'_\ell$ with probability
$$\alpha_{\mathsf{F}}^{\ell}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}_\ell^n) = \min\left(1, \frac{\pi^\ell(\mathbf{Z}'_\ell)\,\mathrm{q}^{\mathsf{ML}}(\mathbf{Z}_\ell^n \,|\, \mathbf{Z}'_\ell)}{\pi^\ell(\mathbf{Z}_\ell^n)\,\mathrm{q}^{\mathsf{ML}}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}_\ell^n)}\right)$$

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

At n^{th} state $\mathbf{Z}_\ell^n$ (of a Markov chain on level ℓ):

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

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

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

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

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

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

Comments on NEW Multilevel MCMC

Revised version of [Ketelsen, RS, Teckentrup, arXiv:1303.7343], in preperation

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

Comments on NEW Multilevel MCMC

Revised version of [Ketelsen, RS, Teckentrup, arXiv:1303.7343], in preperation

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

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

Comments on NEW Multilevel MCMC

Revised version of [Ketelsen, RS, Teckentrup, arXiv:1303.7343], in preperation

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

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

In second case 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).

Comments on NEW Multilevel MCMC

Revised version of [Ketelsen, RS, Teckentrup, arXiv:1303.7343], in preperation

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

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

  In second case 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).

- **Practical algorithm:** Use sub-sampling on level $\ell-1$ to get "independent" samples (**see below for more details**).

Complexity Theorem for Multilevel MCMC

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

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

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. MSE $< \varepsilon^2$ and

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

(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

w. exponential covariance, standard FEs & Fréchet-diff'ble functionals on $H^{\frac{1}{2}-\delta}(D)$

Verifying (M1-M4) for the subsurface flow problem

w. exponential covariance, standard FEs & Fréchet-diff'ble functionals on $H^{\frac{1}{2}-\delta}(D)$

- First split bias into truncation and discretization error:

$$|\mathbb{E}_{\pi^\ell}[Q_\ell] - \mathbb{E}_{\pi^\infty}[Q]| \leq |\mathbb{E}_{\pi^\ell}[Q_\ell - Q(\mathbf{Z}_\ell)]| \qquad \textbf{(M1a)}$$

$$+ |\mathbb{E}_{\pi^\ell}[Q(\mathbf{Z}_\ell)] - \mathbb{E}_{\pi^\infty}[Q]| \qquad \textbf{(M1b)}$$

Verifying (M1-M4) for the subsurface flow problem

w. exponential covariance, standard FEs & Fréchet-diff'ble functionals on $H^{\frac{1}{2}-\delta}(D)$

- First split bias into truncation and discretization error:

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

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

- For **M1a** use $\mathbb{E}_{\pi^\ell}[|X|^q] \lesssim \mathbb{E}_{\mathcal{P}_\ell}[|X|^q]$ (prior 'bounds' posterior) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \lesssim h_\ell^{q-\delta}$ [Teckentrup, RS, et al '13] (see above)

Verifying (M1-M4) for the subsurface flow problem

w. exponential covariance, standard FEs & Fréchet-diff'ble functionals on $H^{\frac{1}{2}-\delta}(D)$

- First split bias into truncation and discretization error:

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

- For **M1a** use $\mathbb{E}_{\pi^\ell}[|X|^q] \lesssim \mathbb{E}_{\mathcal{P}_\ell}[|X|^q]$ (prior 'bounds' posterior) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \lesssim h_\ell^{q-\delta}$ [Teckentrup, RS, et al '13] (see above)

- For **M1b** bound truncation error in posterior [Teckentrup, Thesis, '13]

Verifying (M1-M4) for the subsurface flow problem

w. exponential covariance, standard FEs & Fréchet-diff'ble functionals on $H^{\frac{1}{2}-\delta}(D)$

- 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) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \lesssim h_\ell^{q-\delta}$ [Teckentrup, RS, et al '13] (see above)

- For **M1b** bound truncation error in posterior [Teckentrup, Thesis, '13]

- **M2** **not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved so far for lognormal case!)

Verifying (M1-M4) for the subsurface flow problem

w. exponential covariance, standard FEs & Fréchet-diff'ble functionals on $H^{\frac{1}{2}-\delta}(D)$

- First split bias into truncation and discretization error:

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

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

- For **M1a** use $\mathbb{E}_{\pi^\ell}\left[|X|^q\right] \lesssim \mathbb{E}_{\mathcal{P}_\ell}\left[|X|^q\right]$ (prior 'bounds' posterior) & $\mathbb{E}_{\mathcal{P}_\ell}[|Q_\ell - Q(\mathbf{Z}_\ell)|^q] \lesssim h_\ell^{q-\delta}$ [Teckentrup, RS, et al '13] (see above)
- For **M1b** bound truncation error in posterior [Teckentrup, Thesis, '13]
- **M2 not** specific to multilevel MCMC; first steps to prove it are in [Hairer, Stuart, Vollmer, '11] (but still unproved so far for lognormal case!)
- **M4** holds (with suitable multigrid solver)

Key assumption for multilevel MCMC is (M3)

Key Lemma

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

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

Key assumption for multilevel MCMC is (M3)

Key Lemma

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

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

Proof. First note

$$1 - \alpha_{\mathsf{F}}^{\ell}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) \lesssim \max_{\mathbf{Z}_\ell = \mathbf{Z}'_\ell, \mathbf{Z}^n_\ell} \left| 1 - \frac{\pi^\ell(\mathbf{Z}_\ell)}{\pi^{\ell-1}(\mathbf{Z}_{\ell,C})} \right|$$

Key assumption for multilevel MCMC is (M3)

Key Lemma

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

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

Proof. First note

$$1 - \alpha_{\mathsf{F}}^\ell(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) \lesssim \max_{\mathbf{Z}_\ell = \mathbf{Z}'_\ell, \mathbf{Z}^n_\ell} \left| 1 - \frac{\pi^\ell(\mathbf{Z}_\ell)}{\pi^{\ell-1}(\mathbf{Z}_{\ell,C})} \right|$$

Then recall that $\pi^\ell(\mathbf{Z}_\ell) \eqsim \exp(-\|F_{\text{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2/\sigma^2_{\text{fid}})$ and use

$$\|F_{\text{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2 - \|F_{\text{obs}} - F_{\ell-1}(\mathbf{Z}_{\ell,C})\|^2 \lesssim \|F_\ell(\mathbf{Z}_\ell) - F_{\ell-1}(\mathbf{Z}_{\ell,C})\|$$

Key assumption for multilevel MCMC is (M3)

Key Lemma

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

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

Proof. First note

$$1 - \alpha_{\mathsf{F}}^{\ell}(\mathbf{Z}'_\ell \,|\, \mathbf{Z}^n_\ell) \lesssim \max_{\mathbf{Z}_\ell = \mathbf{Z}'_\ell, \mathbf{Z}^n_\ell} \left| 1 - \frac{\pi^\ell(\mathbf{Z}_\ell)}{\pi^{\ell-1}(\mathbf{Z}_{\ell,C})} \right|$$

Then recall that $\pi^\ell(\mathbf{Z}_\ell) \eqsim \exp(-\|F_{\text{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2/\sigma^2_{\text{fid}})$ and use

$$\|F_{\text{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2 - \|F_{\text{obs}} - F_{\ell-1}(\mathbf{Z}_{\ell,C})\|^2 \lesssim \|F_\ell(\mathbf{Z}_\ell) - F_{\ell-1}(\mathbf{Z}_{\ell,C})\|$$

Since $|1 - \exp(x)| \leq |x| \exp|x|$ it finally follows from [Teckentrup, RS et al '13]

$$\mathbb{E}_{\mathcal{P}_\ell} \left| 1 - \frac{\pi^\ell(\mathbf{Z}_\ell)}{\pi^{\ell-1}(\mathbf{Z}_{\ell,C})} \right| \lesssim h_{\ell-1}^{1-\delta} + s_{\ell-1}^{-1/2+\delta}.$$

Key assumption for multilevel MCMC is (M3)

Theorem

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

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

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

Key assumption for multilevel MCMC is (M3)

Theorem

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

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

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

Proof. Use the facts that $\mathbb{V}[Y] \leq \mathbb{E}[Y^2]$ and that all moments w.r.t. the posterior are bounded by the moments w.r.t. the prior.

Key assumption for multilevel MCMC is (M3)

Theorem

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

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

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

Proof. Use the facts that $\mathbb{V}[Y] \leq \mathbb{E}[Y^2]$ and that all moments w.r.t. the posterior are bounded by the moments w.r.t. the prior.

Then distinguish **two cases**: The coarse modes of $\mathbf{Z}_\ell^n$ and $\mathbf{z}_{\ell-1}^n$ are

- **the same** $\Rightarrow$ result follows again from [Teckentrup, RS et al '13]
- **different** $\Rightarrow$ this only happens with probability $1 - \boldsymbol{\alpha}_\mathrm{F}^\ell$ and

$$\mathbb{E}\left[\mathbf{1}_{\{\text{differ}\}}\right] \leq \mathbb{E}_{\mathcal{P}_\ell,\mathcal{P}_\ell}\left[1 - \boldsymbol{\alpha}^\ell(\mathbf{Z}_\ell^n|\mathbf{Z}_\ell')\right]$$

Key assumption for multilevel MCMC is (M3)

Theorem

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

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

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

Proof. Use the facts that $\mathbb{V}[Y] \leq \mathbb{E}[Y^2]$ and that all moments w.r.t. the posterior are bounded by the moments w.r.t. the prior.

Then distinguish **two cases**: The coarse modes of $\mathbf{Z}_\ell^n$ and $\mathbf{z}_{\ell-1}^n$ are

- **the same** $\Rightarrow$ result follows again from [Teckentrup, RS et al '13]
- **different** $\Rightarrow$ this only happens with probability $1 - \boldsymbol{\alpha}_\mathsf{F}^\ell$ and

$$\mathbb{E}\left[\mathbf{1}_{\{\text{differ}\}}\right] \leq \mathbb{E}_{\mathcal{P}_\ell,\mathcal{P}_\ell}\left[1 - \boldsymbol{\alpha}^\ell(\mathbf{Z}_\ell^n|\mathbf{Z}_\ell')\right]$$

The result then follows from the **Key Lemma**, by applying Hölder's inequality to

$$\mathbb{E}[\mathbf{1}_{\{\text{differ}\}}\,(Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^n))^2].$$

Numerical Example (**OLD** method with bias)

$D = (0,1)^2$, exponential covariance with $\sigma^2 = 1$ & $\lambda = 0.5$, $Q = Q_{\text{out}}$, $h_0 = \frac{1}{16}$

Numerical Example (**OLD** method with bias)

$D = (0,1)^2$, exponential covariance with $\sigma^2 = 1$ & $\lambda = 0.5$, $Q = Q_{\text{out}}$, $h_0 = \frac{1}{16}$

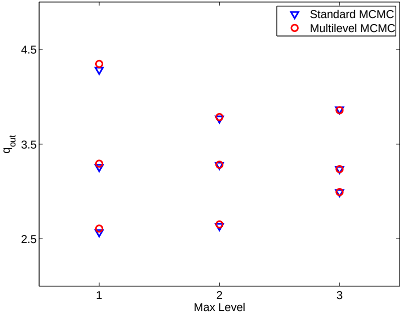
- **"Data"** F_{obs}: Pressure $p(x^*)$ at 9 random points $x^* \in D$.
- # modes: $s_0 = 96$, $s_1 = 121$, $s_2 = 153$ and $s_3 = 169$

Numerical Example (**OLD** method with bias)

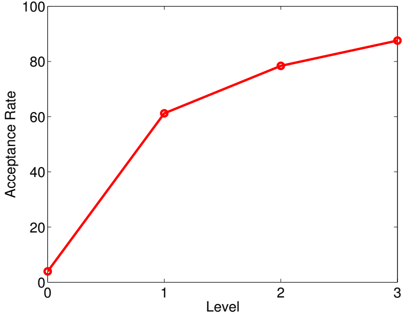
$D = (0,1)^2$, exponential covariance with $\sigma^2 = 1$ & $\lambda = 0.5$, $Q = Q_{\text{out}}$, $h_0 = \frac{1}{16}$

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

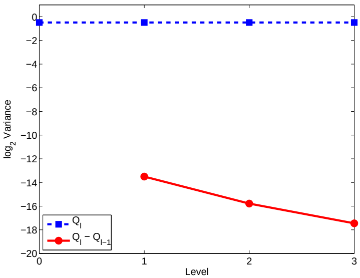
Comparison single- vs. multi-level



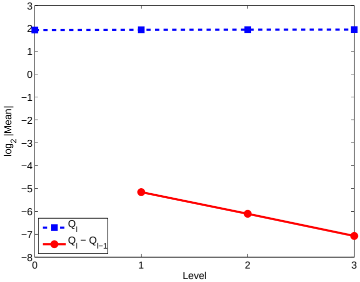
Acceptance rate α_F^ℓ in multilevel estim.



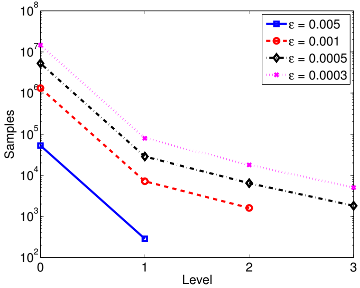
Variance



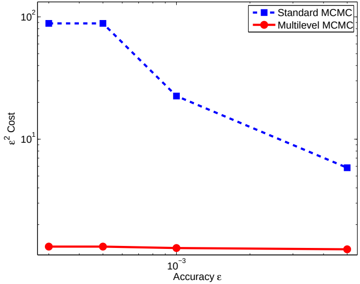
Mean



samples



scaled cost



Proposed NEW Practical Method

- **Recall:**

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

- **Note. No independence** of estimators needed in multilevel method
  $\Rightarrow$ can use same samples on all levels (just extra log in total cost)

Proposed NEW Practical Method

- **Recall:**

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

- **Note.** **No independence** of estimators needed in multilevel method
  $\Rightarrow$ can use same samples on all levels (just extra log in total cost)

- **Practical Algorithm:**

  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{N_0}$ on level 0.

Proposed NEW Practical Method

- **Recall:**

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

- **Note. No independence** of estimators needed in multilevel method
  $\Rightarrow$ can use same samples on all levels (just extra log in total cost)

- **Practical Algorithm:**
  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{N_0}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$

     Above $N_0/N_1 \approx 200$; thus could use period 200 with no extra cost!

Proposed NEW Practical Method

- **Recall:**

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

- **Note. No independence** of estimators needed in multilevel method
  $\Rightarrow$ can use same samples on all levels (just extra log in total cost)

- **Practical Algorithm:**

  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{N_0}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$
     Above $N_0/N_1 \approx 200$; thus could use period 200 with no extra cost!
  3. Use Alg. 2 to get Markov chain $\{\mathbf{Z}_1^n\}_{n=1}^{N_1}$. Continue w. **Step 2**.
     Can use shorter period, since $\alpha_{\mathrm{F}}^1 \approx 1$ and so autocorrelation smaller!

Proposed NEW Practical Method

- **Recall:**

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

- **Note. No independence** of estimators needed in multilevel method
  $\Rightarrow$ can use same samples on all levels (just extra log in total cost)

- **Practical Algorithm:**

  1. Use Algorithm 1 to obtain Markov chain $\{\mathbf{Z}_0^n\}_{n=1}^{N_0}$ on level 0.
  2. Sub-sample this chain (with sufficiently large period) to get (essentially) independent set $\{\mathbf{z}_0^n\}_{n=1}^{N_1}$
     Above $N_0/N_1 \approx 200$; thus could use period 200 with no extra cost!
  3. Use Alg. 2 to get Markov chain $\{\mathbf{Z}_1^n\}_{n=1}^{N_1}$. Continue w. **Step 2**.
     Can use shorter period, since $\alpha_{\mathrm{F}}^1 \approx 1$ and so autocorrelation smaller!
  4. May need extra samples on levels $0, \ldots, L-1$. **Not** on level L **!**
     e.g. $N_{\ell-1}/N_\ell \approx 4$, for $\ell > 1$ above, which may be too short a period.

Additional Comments

- In all our tests 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 our tests 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 our tests 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])
- Related theoretical work by [Hoang, Schwab, Stuart, 2013] (different multilevel splitting and so far no numerics to compare)

Conclusions

- UQ in subsurface flow $\longrightarrow$ PDEs with random coefficients
  (with very high-dimensional parameter space)
- Incorporating data – Bayesian inverse problem
- Multilevel idea extends to Markov chain Monte Carlo
- Theory for lognormal subsurface model problem

Conclusions

- UQ in subsurface flow $\longrightarrow$ PDEs with random coefficients
  (with very high-dimensional parameter space)
- Incorporating data – Bayesian inverse problem
- Multilevel idea extends to Markov chain Monte Carlo
- Theory for lognormal subsurface model problem

Future Work

- Numerical tests w. NEW method; circulant embedding instead of KL
- 3D, parallelisation, application to radwaste case studies
  [w. Gmeiner, Rüde, Wohlmuth]
- Other proposal distributions (e.g. likelihood informed)
  [w. Cui, Law, Marzouk]
- Other applic. (PDE & non-PDE): statisticians, chemists,...
  [w. Lindgren, Simpson]

Thank You!

Preprints available on my website:

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

(revised version of relevant MLMCMC preprint will be available very soon)

I would like to thank the UK Research Council **EPSRC**, as well as **Lawrence Livermore National Lab** (CA) for the financial support of this work.