Multilevel Uncertainty Quantification
with Applications in Subsurface Flow

Robert Scheichl
Department of Mathematical Sciences

Innovative Schemes and Highly Performing Methods for the Numerical Simulation of Fluid Flows
MOMAS Workshop, 15-16 Oct 2012, Marseille
Motivation

- **UK Gov Policy:** Allow building **new** nuclear power stations, **but** waste disposal problem has to be solved first!

  Total current UK Waste (intermediate & highly radioactive): ≈ 220,000 m$^3$ would stand about 17m deep on the pitch in Wembley Stadium

  Long term solution: deep geological disposal (CoRWM July 2006, HMG October 2006)

  → Multiple barriers: mechanical, chemical, physical

  Assessing safety of potential sites of utmost importance (long timescales of 1000s of years)

  → modelling essential!!

  Key aspect: How to quantify uncertainties in the models?

  EPSRC Grant with Universities of Nottingham and Oxford

R. Scheichl (Bath, UK) Multilevel Uncertainty Quantification Marseille, Oct 2012 2 / 37
Motivation

- **UK Gov Policy**: Allow building **new** nuclear power stations, but waste disposal problem has to be solved first!

- **Total current UK Waste** (intermediate & highly radioactive):

  $\approx 220,000 \text{ m}^3$

  would stand about 17m deep on the pitch in Wembley Stadium

- **Long term solution**: deep geological disposal
  (CoRWM July 2006, HMG October 2006)

  $\rightarrow$ **Multiple barriers**: mechanical, chemical, physical

- **Assessing safety** of potential sites of utmost importance
  (long timescales of 1000s of years) $\rightarrow$ **modelling essential!!**
Motivation

- **UK Gov Policy**: Allow building new nuclear power stations, but waste disposal problem has to be solved first!
- **Total current UK Waste** (intermediate & highly radioactive):
  \[ \approx 220,000 \text{ m}^3 \]
  would stand about 17m deep on the pitch in Wembley Stadium

- **Long term solution**: deep geological disposal
  (CoRWM July 2006, HMG October 2006)
  \[ \rightarrow \text{ Multiple barriers: mechanical, chemical, physical} \]
- **Assessing safety** of potential sites of utmost importance
  (long timescales of 1000s of years) \[ \rightarrow \text{ modelling essential!!} \]
- **Key aspect**: How to quantify uncertainties in the models?
  EPSRC Grant with Universities of Nottingham and Oxford
Outline

- Uncertainty Quantification in Groundwater Flow
- HPC Challenge: PDEs with Random Coefficients
- NEW: Multilevel Markov Chain Monte Carlo
- Heterogeneities & Multilevel Deterministic Solvers
- Parallel Scalability (in presence of heterogeneity)
- Conclusions & Outlook
Collaborators

Multilevel Markov Chain Monte Carlo Work

- C. Ketelsen (Lawrence Livermore National Lab)
- A.L. Teckentrup (Bath)
- P.S. Vassilevski (LLLNL)

Other Parts

- J. Charrier (Marseille)
- K.A. Cliffe (Nottingham)
- M. Giles (Oxford)
- I.G. Graham (Bath)
- E. Ullmann (Bath)
Uncertainty Quantification in Single Phase Flow
New multilevel tools also applicable to multiphase flow

Darcy’s Law: \( \vec{q} + k \nabla p = f \)
Incompressibility: \( \nabla \cdot \vec{q} = 0 \)

Boundary Conditions

Society of Petroleum Engineers Benchmark SPE10
Uncertainty Quantification in Single Phase Flow

New multilevel tools also applicable to multiphase flow

\[ \tilde{q} + k \nabla p = f \]

**Darcy’s Law:**

**Incompressibility:**

\[ \nabla \cdot \tilde{q} = 0 \]

+ Boundary Conditions

\[ \text{uncertain } k \rightarrow \text{uncertain } p, \tilde{q} \]

Society of Petroleum Engineers Benchmark SPE10
Stochastic Modelling of Uncertainty:
Model uncertain conductivity tensor $k$ as a lognormal random field
Stochastic Modelling of Uncertainty:
Model uncertain conductivity tensor $k$ as a lognormal random field

**Typical simplified model (prior):**

- $k(x, \omega)$ isotropic, scalar
- $\log k(x, \omega) = \text{Gaussian}$
  
  mean-free with exponential covariance:

\[
R(x, y) := \sigma^2 \exp\left(-\frac{\|x - y\|}{\lambda}\right)
\]
Stochastic Modelling of Uncertainty:
Model uncertain conductivity tensor $k$ as a lognormal random field

Typical simplified model (prior):

- $k(x, \omega)$ isotropic, scalar
- $\log k(x, \omega) = \text{Gaussian}$

meanfree with exponential covariance:

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

- Incorporating data (Posterior) $\rightarrow$ MCMC
Stochastic Modelling of Uncertainty:
Model uncertain conductivity tensor $k$ as a lognormal random field

Typical simplified model (prior):

- $k(x, \omega)$ isotropic, scalar
- $\log k(x, \omega) = \text{Gaussian}$
  meanfree with exponential covariance:
  \[
  R(x, y) := \sigma^2 \exp \left( -\frac{\|x - y\|}{\lambda} \right)
  \]
- Incorporating data (Posterior) $\rightarrow$ MCMC

Typical quantities of interest:

- effective conductivity $k_{\text{eff,1}} = \frac{1}{|D|} \int_D q_1$
- breakthrough time; water cut; etc...

Typical realisation
($\lambda = \frac{1}{64}, \sigma^2 = 8$)
Key Computational Challenges

PDEs with Highly Heterogeneous Random Coefficients

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

PDEs with Highly Heterogeneous Random Coefficients

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

- **Sampling** from random field \(k(x, \omega)\):
  - truncated Karhunen-Loève expansion
  - Matrix factorisation, e.g. circulant embedding (FFT)
  - etc . . .
Key Computational Challenges
PDEs with Highly Heterogeneous Random Coefficients

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

- **Sampling** from random field \(k(x, \omega):\)
  - truncated Karhunen-Loève expansion
  - Matrix factorisation, e.g. circulant embedding (FFT)
  - etc . . .

- **High-Dimensional Integration:**
  - stochastic Galerkin (+ sparse versions)
  - stochastic collocation (+ sparse & anisotropic versions)
  - Monte Carlo & Markov Chain MC \(\leftarrow\) Multilevel !
  - etc . . .
Key Computational Challenges

PDEs with Highly Heterogeneous Random Coefficients

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

- **Sampling** from random field $k(x, \omega)$:
  - truncated Karhunen-Loève expansion
  - Matrix factorisation, e.g. circulant embedding (FFT)
  - etc . . .

- **High-Dimensional Integration**:
  - stochastic Galerkin (+ sparse versions)
  - stochastic collocation (+ sparse & anisotropic versions)
  - Monte Carlo & Markov Chain MC $\leftarrow$ **Multilevel** !
  - etc . . .

- **Solution** of large number of **multiscale** deterministic PDEs:
  - Multigrid, Domain Decomposition Methods, **AMG**
Why is this problem so challenging?
Why is this problem so challenging?

Truncated KL expansion of \( \log k(x, \omega) \approx \sum_{j=1}^{J} \sqrt{\mu_j} \phi_j(x) \xi_j(\omega) \)

\((\mu_j, \phi_j(x))\) orthonormal eigenpairs of \( \int_{\Omega} R(x, y) \phi(y) dy \); \( \xi_j(\omega) \) i.i.d. \( N(0, \sigma^2) \)

KL-eigenvalues in 1D

Convergence of \( q|_{x=1} \) w.r.t. \( J \)
Why is this problem so challenging?

Truncated KL expansion of \( \log k(x, \omega) \approx \sum_{j=1}^{J} \sqrt{\mu_j} \phi_j(x) \xi_j(\omega) \)

\((\mu_j, \phi_j(x))\) orthonormal eigenpairs of \( \int_{\Omega} R(x, y) \phi(y) dy \); \( \xi_j(\omega) \) i.i.d. \( N(0, \sigma^2) \)

- Large \#KL-modes for small \( \lambda \) \( \Rightarrow \) high dimension \( J \gg 100 \)
- Low regularity (\( k \) only Hölder with \( \eta < \frac{1}{2} \)) \( \Rightarrow \) fine mesh \( h \ll 1 \)
- Large \( \sigma^2 \) & exponential \( \Rightarrow \) large heterogeneity \( \frac{k_{\max}}{k_{\min}} > 10^6 \)
Spatial discretisation (not important for this talk)

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

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

- Possible extensions to two-phase/multi-phase flow: e.g.

  $$(A_{pp}(\omega) A_{ps} A_{sp} A_{ss})(\delta p(\omega) \delta s(\omega)) = \text{RHS}$$

  with diffusion-type operator

  $$A_{pp} \approx -\nabla_h \cdot (\alpha \eta_s(\delta s) \nabla_h) + ...$$
Spatial discretisation (not important for this talk)

- **Standard FEs** (cts pw. linear) or **cell-centred FVs** on $\mathcal{T}^h$:
  
  $$A(\omega)p(\omega) = b(\omega) \quad M_h \times M_h \text{ linear system}$$

- **Mixed FEs**:

  $$\begin{bmatrix}
  M(\omega) & B^T \\
  B & 0
  \end{bmatrix}
  \begin{bmatrix}
  q(\omega) \\
  p(\omega)
  \end{bmatrix}
  =
  \begin{bmatrix}
  g(\omega) \\
  0
  \end{bmatrix}$$

  - [Graham, RS, Ullmann, in preparation]
Spatial discretisation (not important for this talk)

- **Standard FEs** (cts pw. linear) or **cell-centred FVs** on $\mathcal{T}^h$:
  \[
  A(\omega) p(\omega) = b(\omega) \quad M_h \times M_h \text{ linear system}
  \]

- **Mixed FEs**:
  \[
  \begin{bmatrix}
  M(\omega) & B^T \\
  B & 0 \end{bmatrix}
  \begin{bmatrix}
  q(\omega) \\
  p(\omega) \end{bmatrix}
  =
  \begin{bmatrix}
  g(\omega) \\
  0 \end{bmatrix}
  \]

  ▶ [Graham, RS, Ullmann, in preparation]

- **Possible extensions to two-phase/multi-phase flow**: e.g.
  \[
  \begin{pmatrix}
  A_{pp}(\omega) & A_{ps} \\
  A_{sp} & A_{ss} \end{pmatrix}
  \begin{pmatrix}
  \delta p(\omega) \\
  \delta s(\omega) \end{pmatrix}
  = \text{RHS}
  \]

  with diffusion-type operator $A_{pp} \approx -\nabla^h \cdot (\alpha \eta(s) \nabla^h) + \ldots$
Multilevel Stochastic Solvers
Quantity of interest: Expected value $\mathbb{E}[Q]$ of $Q := G(p)$

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

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

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

Assume optimal PDE solver $\Rightarrow \text{Cost}(Q_h^{(i)}) = \mathcal{O}(M_h) = \mathcal{O}(h^{-d})$
• **Quantity of interest**: Expected value $\mathbb{E}[Q]$ of $Q := G(p)$ some (nonlinear) functional of the PDE solution $p$

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

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

Assume optimal PDE solver $\Rightarrow$ Cost($Q_h^{(i)}$) = $\mathcal{O}(M_h)$ = $\mathcal{O}(h^{-d})$

• **Mean square error (MSE)** for standard MC:

$$\mathbb{E}[(\hat{Q}_h^{MC} - \mathbb{E}[Q])^2] = \frac{\mathbb{V}[Q_h]}{N} + \frac{(\mathbb{E}[Q_h - Q])^2}{N}$$

- **sampling error**
- **FE error (“bias”)**
• **Quantity of interest:** Expected value $\mathbb{E}[Q]$ of $Q := G(p)$ some (nonlinear) functional of the PDE solution $p$

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

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

Assume optimal PDE solver $\Rightarrow$ Cost$(Q_h^{(i)}) = \mathcal{O}(M_h) = \mathcal{O}(h^{-d})$

• **Mean square error (MSE)** for standard MC:

$$\mathbb{E}\left[\left(\hat{Q}_h^{MC} - \mathbb{E}[Q]\right)^2\right] = \frac{\mathbb{V}[Q_h]}{N} + \frac{(\mathbb{E}[Q_h] - Q)^2}{1}$$

  - sampling error
  - FE error (“bias”)

• Need **large $N$ and small $h$** (low regularity) $\Rightarrow$ **Too expensive**:

$$\text{Cost}(\hat{Q}_h^{MC}) = \mathcal{O}(N \cdot h^{-d})$$

(especially in 3D!)
What is cost to get MSE below tolerance $\varepsilon^2$?
What is cost to get MSE below tolerance $\varepsilon^2$?

**Complexity of Standard Monte Carlo**

If $Q_h \rightarrow Q$ with $O(h^\alpha)$ for some $\alpha > 0$, then to obtain MSE $< \varepsilon^2$

$$\text{Cost}(\hat{Q}_h^{MC}) = O(\varepsilon^{-2} - \frac{d}{\alpha})$$
What is cost to get MSE below tolerance $\varepsilon^2$?

**Complexity of Standard Monte Carlo**

If $Q_h \rightarrow Q$ with $\mathcal{O}(h^\alpha)$ for some $\alpha > 0$, then to obtain $\text{MSE} < \varepsilon^2$

$$\text{Cost}(\hat{Q}_h^{MC}) = \mathcal{O}(\varepsilon^{-2 - \frac{d}{\alpha}})$$

**Numerical Example**  ($D = (0,1)^2, Q = k_{\text{eff},1}$, mixed FE & amg1r5)

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

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$h^{-1}$</th>
<th>$N$</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>129</td>
<td>$\sim$ 14000</td>
<td>21 min</td>
</tr>
<tr>
<td>0.002</td>
<td>1025</td>
<td>$\sim$ 350000</td>
<td>30 days</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$h^{-1}$</th>
<th>$N$</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>513</td>
<td>$\sim$ 8500</td>
<td>4 h</td>
</tr>
<tr>
<td>0.002</td>
<td></td>
<td></td>
<td><strong>Prohibitively large!!</strong></td>
</tr>
</tbody>
</table>

Here $d = 2$ & $\alpha \approx \frac{3}{4} \Rightarrow \text{Cost} \approx \mathcal{O}(\varepsilon^{-14/3}) \approx 25 \times$ more work to halve error!
Multilevel Monte Carlo [Heinrich, ’01], [Giles, ’07] [Cliffe, Giles, RS, Teckentrup, ’11]

Note that trivially

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

(where $h_\ell = h_{\ell-1}/2$ and $Q_\ell = Q_{h_\ell}$)
Multilevel Monte Carlo \cite{Heinrich01,Giles07,Cliffe11}

Note that trivially

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

(where \( h_{\ell} = h_{\ell-1}/2 \) and \( Q_{\ell} = Q_{h_{\ell}} \))

Define the following \textbf{multilevel MC} estimator for \( \mathbb{E}[Q] \):

$$ \hat{Q}_L^{\text{ML}} := \hat{Q}_0^{\text{MC}} + \sum_{\ell=1}^{L} \hat{Y}_{\ell}^{\text{MC}} \quad \text{where} \quad Y_{\ell} := Q_{\ell} - Q_{\ell-1} $$
Multilevel Monte Carlo  

[Heinrich, ’01], [Giles, ’07]  

[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 \( h_\ell = h_{\ell-1}/2 \) and \( Q_\ell = Q_{h_\ell} \))  

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

\[ \hat{Q}_L^{ML} := \hat{Q}_0^{MC} + \sum_{\ell=1}^{L} \hat{Y}_{\ell}^{MC} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1} \]  

**Key Observation** (multigrid idea: compute corrections)  

\[ \nabla [Q_\ell - Q_{\ell-1}] \to 0 \quad \text{as} \quad h_\ell \to 0 \]
Complexity Theorem for Multilevel Monte Carlo

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

$$\nabla [Q_\ell - Q_{\ell-1}] = O(h_\ell^\beta).$$

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

$$\text{Cost}(\hat{Q}_L^{\text{MLC}}) = O(\varepsilon^{-2 - \frac{d - \beta}{\alpha}}) \quad \text{(if } \beta < d)$$

For $\beta \geq d$ the cost is $O(\varepsilon^{-2})$ with a possible log-factor.
Complexity Theorem for Multilevel Monte Carlo

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

$$\nabla [Q_\ell - Q_{\ell-1}] = O(h_\ell^\beta).$$

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

$$\text{Cost}(\hat{Q}^{\text{MLMC}}_L) = O(\varepsilon^{-\frac{2-d-\beta}{\alpha}}) \quad \text{(if $\beta < d$)}$$

For $\beta \geq d$ the cost is $O(\varepsilon^{-2})$ with a possible log-factor.

**Cost estimates:** taking $\alpha \approx 0.75$ (as in example above) and $\beta \approx 2\alpha$

<table>
<thead>
<tr>
<th>$d$</th>
<th>MLMC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$O(\varepsilon^{-8/3})$</td>
<td>$O(\varepsilon^{-14/3})$</td>
</tr>
<tr>
<td>3</td>
<td>$O(\varepsilon^{-4})$</td>
<td>$O(\varepsilon^{-6})$</td>
</tr>
</tbody>
</table>
Complexity Theorem for Multilevel Monte Carlo

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

$$\nabla [Q_\ell - Q_{\ell-1}] = \mathcal{O}(h^\beta).$$

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

$$\text{Cost}(\hat{Q}_L^{\text{MLMC}}) = \mathcal{O}(\varepsilon^{-2} - \frac{d-\beta}{\alpha})$$

(if $\beta < d$)

For $\beta \geq d$ the cost is $\mathcal{O}(\varepsilon^{-2})$ with a possible log-factor.

Cost estimates: taking $\alpha \approx 0.75$ (as in example above) and $\beta \approx 2\alpha$

<table>
<thead>
<tr>
<th>$d$</th>
<th>MLMC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\mathcal{O}(\varepsilon^{-8/3})$</td>
<td>$\mathcal{O}(\varepsilon^{-14/3})$</td>
</tr>
<tr>
<td>3</td>
<td>$\mathcal{O}(\varepsilon^{-4})$</td>
<td>$\mathcal{O}(\varepsilon^{-6})$</td>
</tr>
</tbody>
</table>

Asymptotically same as one deterministic solve to accuracy $\varepsilon$!
Numerical Example

$D = (0, 1)^2$, $\sigma^2 = 1$, $\lambda = 0.1$, $h_0 = \frac{1}{16}$, $\#KL = 500$, standard FE, UMFPACK

$Q = k_{\text{eff,1}}$, $\varepsilon = 0.001$

Matlab implementation on 3GHz Intel Core 2 Duo E8400 proc, 3.2GByte RAM

Here with sparse direct solver, i.e. cost/sample $\approx \mathcal{O}(h^{-3})$ (not $\mathcal{O}(h^{-d})$)
Theory: Verifying Assumptions of Complexity Theorem
Theory: 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, 2011]**: \( k \) lognormal, i.e. not uniformly elliptic/bounded and only \( k(\cdot, \omega) \in C^{0,\eta}(D) \), for some \( \eta < 1 \)
Theory: 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, 2011]**: \( k \) lognormal, i.e. not uniformly elliptic/bounded and only \( k(\cdot, \omega) \in C^{0, \eta}(D) \), for some \( \eta < 1 \)

- New regularity result: \((q\text{-th moment of } H^{1+s}\text{-norm})\)

\[
\| p \|_{L^q(\Omega, H^{1+s}(D))} \leq C_{s,q} \| f \|_{L^2(D)}, \quad \forall \ s < \eta, \ q < \infty.
\]

- New FE error result: \((q\text{-th moment of } H^1\text{-norm})\)

\[
\| p - p_h \|_{L^q(\Omega, H^1(D))} \leq C'_{s,q} \| f \|_{L^2(D)} h^s, \quad \forall \ s < \eta, \ q < \infty.
\]

leads to \( \alpha = s, \ \beta = 2s \) for \( Q := \| p \|_{H^1(D)} \) (and for \( L_2 \)-norm twice that)
Theory: 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, 2011]**: \( k \) lognormal, i.e. not uniformly elliptic/bounded and only \( k(\cdot, \omega) \in C^{0,\eta}(D) \), for some \( \eta < 1 \)

- New regularity result: \((q\text{-th moment of } H^{1+s}\text{-norm}) \leftarrow \text{Charrier, Tue}\)
  \[ \|p\|_{L^q(\Omega, H^{1+s}(D))} \leq C_{s,q} \|f\|_{L^2(D)}, \forall s < \eta, \ q < \infty. \]

- New FE error result: \((q\text{-th moment of } H^1\text{-norm})\)
  \[ \|p - p_h\|_{L^q(\Omega, H^1(D))} \leq C'_{s,q} \|f\|_{L^2(D)} \ h^s, \forall s < \eta, \ q < \infty. \]
  leads to \( \alpha = s, \ \beta = 2s \) for \( Q := |p|_{H^1(D)} \) (and for \( L^2 \)-norm twice that)

- **[Teckentrup, RS, Giles, Ullmann, 2012]**: (nonlinear) functionals, corner domains, discontinuous coefficients, level-dependent truncations

- Similar results for mixed FEs **[Graham, RS, Ullmann, 2012]**; FVM, tensor coefficients, \( L^\infty -, \ W^{1,\infty} \)-norms **[Teckentrup, 2012+]**
Incorporating data – Markov Chain Monte Carlo

- Can multilevel idea be extended to MCMC?

Multilevel Metropolis-Hastings algorithm (here only 2-level)

- Split KL modes: \( J_0 \) coarse + 1 \( J_1 \) fine

- As for standard multilevel MC use

- NEW: Separate Markov chain on level \( \ell \), but "coarse" modes of new proposal taken from last state on level \( \ell - 1 \) (different transition probability, but easily computable)

- Use idea recursively (but always start on Level 0 to avoid bias)
Incorporating data – Markov Chain Monte Carlo

- Can **multilevel** idea be extended to **MCMC**? Answer: **YES!**
  (Example: standard Metropolis-Hastings (**MetH**) algorithm)
Can **multilevel** idea be extended to **MCMC**? Answer: **YES!**  
(Example: standard Metropolis-Hastings (MetH) algorithm)

**Multilevel Metropolis-Hastings algorithm** (here only 2-level)

- Split KL modes:
  
  $$\begin{array}{c|cc}
  1 & \text{coarse} & J_0 \\
  \hline
  J_0 + 1 & \text{fine} & J_1 \\
  \end{array}$$

- As for standard multilevel MC use

  $$\mathbb{E}^{\text{post}}[Q_\ell] = \mathbb{E}^{\text{post}}[Q_\ell - Q_{\ell-1}] + \mathbb{E}^{\text{post}}[Q_{\ell-1}]$$

  \[
  \uparrow \quad \uparrow \quad \uparrow \quad \text{NEW} \quad \text{standard MetH} \quad \text{standard MetH}
  \]

- **NEW**: Separate Markov chain on level $\ell$, but “coarse” modes of new proposal taken from last state on level $\ell - 1$ (different transition probability, but easily computable)
Incorporating data – Markov Chain Monte Carlo

- Can **multilevel** idea be extended to **MCMC**? Answer: **YES**!
  
  (Example: standard Metropolis-Hastings (MetH) algorithm)

- **Multilevel Metropolis-Hastings algorithm** (here only 2-level)
  
  - Split KL modes:
    
    \[
    \mathbb{E}^{\text{post}}[Q_\ell] = \mathbb{E}^{\text{post}}[Q_\ell] - Q_{\ell-1} + \mathbb{E}^{\text{post}}[Q_{\ell-1}]
    \]

    **NEW**: Separate Markov chain on level \( \ell \), **but** “coarse” modes of new proposal taken from last state on level \( \ell - 1 \) (different transition probability, but easily computable)

  - Use idea recursively (but always start on Level 0 to avoid bias)
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

We have a genuine Markov chain on every level. The multilevel algorithm is consistent (no bias between levels) and converges for any initial state.

The acceptance probability \( \alpha^\ell \) → 1 as \( h^\ell \to 0 \) and \#KL^\ell → ∞.

Similar complexity theorem (as for standard multilevel MC):

\[
\text{Cost} = O\left(\varepsilon^{-2} - d - \beta \alpha\right)
\]

(as above), but here \( \beta = \alpha \Rightarrow \) less gain!

(Can't show yet) but believe initial “burn-in” also significantly reduced.
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine **Markov chain** on every level
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine **Markov chain** on every level
- The multilevel algorithm is **consistent** (no bias between levels) (requires starting always on the coarsest level, like in full MG)

\[ \alpha_\ell \rightarrow 1 \text{ as } h_\ell \rightarrow 0 \text{ and } \#KL_\ell \rightarrow \infty \]

Similar complexity theorem (as for standard multilevel MC):

\[ \text{Cost} = O(\varepsilon^{-2} - d - \beta \alpha) \text{ (as above), but } \beta = \alpha \Rightarrow \text{less gain!} \]

(Can't show yet) but believe initial "burn-in" also significantly reduced
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine **Markov chain** on every level
- The multilevel algorithm is **consistent** (no bias between levels)
  (requires starting always on the coarsest level, like in full MG)
- The multilevel algorithm **converges** for any initial state
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine Markov chain on every level
- The multilevel algorithm is consistent (no bias between levels) (requires starting always on the coarsest level, like in full MG)
- The multilevel algorithm converges for any initial state
- Acceptance probability $\alpha^\ell \to 1$ as $h^\ell \to 0$ and $\#KL^\ell \to \infty$
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine Markov chain on every level
- The multilevel algorithm is consistent (no bias between levels) (requires starting always on the coarsest level, like in full MG)
- The multilevel algorithm converges for any initial state
- Acceptance probability $\alpha^\ell \rightarrow 1$ as $h^\ell \rightarrow 0$ and $\#KL^\ell \rightarrow \infty$
- Similar complexity theorem (as for standard multilevel MC):
  \[
  \text{Cost} = \mathcal{O}(\varepsilon^{-2-\frac{d-\beta}{\alpha}}) \quad \text{(as above), but here } \beta = \alpha \Rightarrow \text{less gain!}
  \]
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine **Markov chain** on every level

- The multilevel algorithm is **consistent** (no bias between levels)
  (requires starting always on the coarsest level, like in full MG)

- The multilevel algorithm **converges** for any initial state

- **Acceptance probability** $\alpha^\ell \rightarrow 1$ as $h_\ell \rightarrow 0$ and $\#KL_\ell \rightarrow \infty$

- Similar **complexity theorem** (as for standard multilevel MC):

  $\text{Cost} = O\left(\varepsilon^{-2 - \frac{d-\beta}{\alpha}}\right)$ (as above), **but** here $\beta = \alpha \Rightarrow$ less gain!
Multilevel MCMC – What can we prove?

Hot of the Press! [Ketelsen, Scheichl, Teckentrup, Vassilevski, in prep.]

- We have a genuine Markov chain on every level
- The multilevel algorithm is consistent (no bias between levels) (requires starting always on the coarsest level, like in full MG)
- The multilevel algorithm converges for any initial state
- Acceptance probability $\alpha^\ell \to 1$ as $h^\ell \to 0$ and $\#KL^\ell \to \infty$
- Similar complexity theorem (as for standard multilevel MC):
  \[
  \text{Cost} = O\left(\varepsilon^{-2-\frac{d-\beta}{\alpha}}\right) \quad \text{(as above), but here } \beta = \alpha \Rightarrow \text{less gain!}
  \]
- (Can’t show yet) but believe initial “burn-in” also significantly reduced
Numerical Example (multilevel MCMC)

\[ D = (0, 1)^2, \sigma^2 = 1, \lambda = 1, Q = k_{\text{eff}, 1} \]

- **Data (artificial):** Pressure \( p \) at 9 random points in domain
- \#KL_1 = 100 and \#KL_2 = 90
- Averaging over 10000 samples (+ 2500 for “burn-in”)

<table>
<thead>
<tr>
<th>( h^{-1} )</th>
<th>( \sigma^2_{\text{fidelity}} )</th>
<th>( \mathbb{V}^{\text{post}}(Q_h) )</th>
<th>( \mathbb{V}^{\text{post}}(Q_h - Q_{2h}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>0.009</td>
<td>0.51</td>
<td>0.0120</td>
</tr>
<tr>
<td>128</td>
<td>0.007</td>
<td>0.53</td>
<td>0.0100</td>
</tr>
<tr>
<td>256</td>
<td>0.005</td>
<td>0.40</td>
<td>0.0030</td>
</tr>
<tr>
<td>512</td>
<td>0.003</td>
<td>0.57</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

\[ \text{const} \quad \mathcal{O}(h^{-1}) \]
Multilevel Markov Chain Monte Carlo
(Details & Theory)
Incorporating Data – Bayes’ Theorem

- Approximate log-normal $k$, e.g., by truncated KL expansion:

$$k = \exp \left[ \sum_{j=1}^{\infty} \sqrt{\nu_n} \phi_j(x) \xi_j(\omega) \right] \approx \exp \left[ \sum_{j=1}^{J} \sqrt{\nu_j} \phi_j(x) \xi_j(\omega) \right]$$

Parametrised by $\theta_J := [\xi_1, \ldots, \xi_J]$ (“prior”) (other possibilities)

- Usually also some data $F_{\text{obs}}$ related to outputs, e.g. pressure. To reduce uncertainty, incorporate $F_{\text{obs}}$ (“posterior”)

Using Bayes’ Theorem, we have

$$\pi(h, \theta_J) \propto L(h | F_{\text{obs}}) \pi(\theta_J)$$

Terms on RHS computable! Proportionality constant $1/\pi(F_{\text{obs}})$ unknown
Incorporating Data – Bayes’ Theorem

- Approximate log-normal \( k \), e.g., by truncated KL expansion:

\[
k = \exp \left[ \sum_{j=1}^{\infty} \sqrt{\nu_n} \phi_j(x) \xi_j(\omega) \right] \approx \exp \left[ \sum_{j=1}^{J} \sqrt{\nu_j} \phi_j(x) \xi_j(\omega) \right]
\]

Parametrised by \( \theta_J := [\xi_1, \ldots, \xi_J] \) (“prior”) (other possibilities)

- Usually also some data \( F_{\text{obs}} \) related to outputs, e.g. pressure. To reduce uncertainty, incorporate \( F_{\text{obs}} \) (“posterior”)

- Using Bayes’ Theorem, we have

\[
\pi^{h,J}(\theta_J) := \mathcal{P}(\theta_J \mid F_{\text{obs}}) \approx \frac{\mathcal{L}_h(F_{\text{obs}} \mid \theta_J) \mathcal{P}(\theta_J)}{\mathcal{P}(F_{\text{obs}})}
\]

Terms on RHS computable! Proportionality constant \( 1/\mathcal{P}(F_{\text{obs}}) \) unknown
Incorporating Data – Bayes’ Theorem

Need to specify

- **prior distribution** (all information available on $k$)

  Here, for simplicity we chose $k$ **log-normal**:

  $$ P(\theta_J) \approx \frac{1}{(2\pi)^{J/2}} \exp \left[ - \sum_{j=1}^{J} \frac{\xi_j^2}{2} \right]. $$
Incorporating Data – Bayes’ Theorem

Need to specify

- **prior distribution** (all information available on $k$)
  - Here, for simplicity we chose $k$ **log-normal**:
    \[
    P(\theta_J) \approx \frac{1}{(2\pi)^{J/2}} \exp \left[ -\sum_{j=1}^{J} \frac{\xi_j^2}{2} \right].
    \]

- **likelihood model** (quantifying how likely it is that the current value of $\theta_J$ leads to the observed data $F_{\text{obs}}$ on mesh $\mathcal{T}_h$)
  - Here, we choose a **normal model**:
    \[
    L_h(F_{\text{obs}} \mid \theta_J) \approx \exp \left[ -\left\| F_{\text{obs}} - F^h(\theta_J) \right\|^2 \right] \frac{1}{\sigma_{\text{fid},h}^2}.\]

  - $F^h(\theta_J)$ ... model response; $\sigma_{\text{fid},h}$ ... fidelity parameter ($h$-dep.)
  (includes numerical, modelling and measuring errors)
ALGORITHM 1. (Standard Metropolis Hastings MCMC)

- Choose $\theta^0_J$.
- At state $\theta^n_J$ generate proposal $\theta'_J$ from proposal distribution $q(\theta'_J | \theta^n_J)$ (e.g. random walk).
- Accept $\theta'_J$ as a sample with probability

$$\alpha^{h,J} = \min \left( 1, \frac{\pi^{h,J}(\theta'_J) q(\theta^n_J | \theta'_J)}{\pi^{h,J}(\theta^n_J) q(\theta'_J | \theta^n_J)} \right) = \min \left( 1, \frac{\pi^{h,J}(\theta'_J)}{\pi^{h,J}(\theta^n_J)} \right)$$

i.e. $\theta^{n+1}_J = \theta'_J$ with probability $\alpha^{h,J}$; otherwise $\theta^{n+1}_J = \theta^n_J$. 

Pros:
- Produces a Markov chain $\{\theta^n_J\}_{n \in \mathbb{N}}$ with $\theta^n_J \sim \pi^{h,J}$ as $n \to \infty$.

Cons:
- Evaluation of $\alpha^{h,J}$ very expensive for small $h$.
- Acceptance rate $\alpha^{h,J}$ very low for large $J$ ($< 10\%$).
ALGORITHM 1. (Standard Metropolis Hastings MCMC)

- Choose \( \theta^0_J \).

- At state \( \theta^n_J \) generate proposal \( \theta'_J \) from proposal distribution \( q(\theta'_J | \theta^n_J) \) (e.g. random walk).

- Accept \( \theta'_J \) as a sample with probability

\[
\alpha^{h,J} = \min \left( 1, \frac{\pi^{h,J}(\theta'_J) q(\theta^n_J | \theta'_J)}{\pi^{h,J}(\theta^n_J) q(\theta'_J | \theta^n_J)} \right) = \min \left( 1, \frac{\pi^{h,J}(\theta'_J)}{\pi^{h,J}(\theta^n_J)} \right)
\]

i.e. \( \theta^{n+1}_J = \theta'_J \) with probability \( \alpha^{h,J} \); otherwise \( \theta^{n+1}_J = \theta^n_J \).

Pros:
- Produces a Markov chain \( \{\theta^n_J\}_{n \in \mathbb{N}} \), with \( \theta^n_J \sim \pi^{h,J} \) as \( n \to \infty \).

Cons:
- Evaluation of \( \alpha^{h,J} \) very expensive for small \( h \).
- Acceptance rate \( \alpha^{h,J} \) very low for large \( J \) (\( < 10\% \)).
Standard Markov chain Monte Carlo

- Samples $\theta^*_j$ used as usual for inference (even though not i.i.d.):

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

where $Q_{h,J}^{(n)} = \mathcal{G} \left( p_h(\theta^*_j^{(n)}) \right)$ is the $n$th sample of $Q$ on $T_h$. 
Standard Markov chain Monte Carlo

- Samples $\theta^n_j$ used as usual for inference (even though not i.i.d.):
  \[ \mathbb{E}_{\pi^h,J}[Q] \approx \mathbb{E}_{\pi^h,J}[Q_{h,J}] \approx \frac{1}{N} \sum_{n=1}^{N} Q^{(n)}_{h,J} := \hat{Q}^\text{MetH}_{h,J} \]
  where $Q^{(n)}_{h,J} = G(p_h(\theta_j^{(n)}))$ is the $n$th sample of $Q$ on $T_h$.

- **Convergence:** If (as satisfied for the random walk)
  \[ q(\theta'_j | \theta^n_j) > 0, \quad \text{for all } (\theta'_j, \theta^n_j) \text{ with } \pi^h,J(\theta'_j), \pi^h,J(\theta^n_j) > 0 \]
  then
  \[ \lim_{N \to \infty} \hat{Q}^\text{MetH}_{h,J} = \mathbb{E}_{\pi^h,J}[Q_{h,J}] \]
Standard Markov chain Monte Carlo

- Samples $\theta^n_J$ used as usual for inference (even though not i.i.d.):

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

where $Q_{h,J}^{(n)} = G(p_h(\theta_J^{(n)}))$ is the $n$th sample of $Q$ on $T_h$.

- **Convergence:** If (as satisfied for the random walk)

\[
q(\theta'_J | \theta_J^n) > 0, \quad \text{for all } (\theta'_J, \theta_J^n) \text{ with } \pi^h,J(\theta'_J), \pi^h,J(\theta_J^n) > 0
\]

then

\[
\lim_{N \to \infty} \hat{Q}_{h,J}^{\text{MetH}} = \mathbb{E}_{\pi^h,J} [Q_{h,J}]
\]

- Can bound mean square error by

\[
\nabla_{\text{MetH}} \left[ \hat{Q}_{h,J}^{\text{MetH}} \right] + \left( \mathbb{E}_{\text{MetH}} \left[ \hat{Q}_{h,J}^{\text{MetH}} \right] - \mathbb{E}_{\pi^h,J} \left[ \hat{Q}_{h,J}^{\text{MetH}} \right] \right)^2 \quad (\text{MCMC errors})
\]

\[
+ \left( \mathbb{E}_{\pi^h,J} \left[ Q_{h,J} - Q \right] \right)^2 \quad (\text{approx. errors})
\]
Multilevel Markov Chain Monte Carlo

using $h_\ell = h_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, $\ell = 1, \ldots, L$, and setting $Q_\ell = Q_{h_\ell,J_\ell}$

Key ingredients in standard multilevel Monte Carlo:

- **Much cheaper** to solve the PDE on coarser grids
- $\forall [Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty \Rightarrow$ fewer samples on finer grids
- Telescoping sum: $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell] - \mathbb{E}[Q_{\ell-1}]$
Multilevel Markov Chain Monte Carlo

using $h_\ell = h_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, $\ell = 1, \ldots, L$, and setting $Q_\ell = Q_{h_\ell,J_\ell}$

Key ingredients in standard multilevel Monte Carlo:

- **Much cheaper** to solve the PDE on coarser grids
- $\nabla [Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty \Rightarrow$ **fewer samples** on finer grids
- Telescoping sum: $\mathbb{E} [Q_L] = \mathbb{E} [Q_0] + \sum_{\ell=1}^{L} \mathbb{E} [Q_\ell] - \mathbb{E} [Q_{\ell-1}]$

In MCMC setting **target distribution depends on $\ell$**:

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

using $h_\ell = h_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, $\ell = 1, \ldots, L$, and setting $Q_\ell = Q_{h_\ell,J_\ell}$

Key ingredients in standard multilevel Monte Carlo:

- **Much cheaper** to solve the PDE on coarser grids
- $\nabla [Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty \Rightarrow$ **fewer samples** on finer grids
- Telescoping sum: $\mathbb{E} [Q_L] = \mathbb{E} [Q_0] + \sum_{\ell=1}^{L} \mathbb{E} [Q_\ell] - \mathbb{E} [Q_{\ell-1}]$

In MCMC setting **target distribution depends on** $\ell$:

$$
\hat{Q}_{ML}^L := \frac{1}{N_0} \sum_{n=1}^{N_0} Q_0(\theta^n_0) + \sum_{\ell=1}^{L} \frac{1}{N_{\ell}} \sum_{n=1}^{N_{\ell}} (Q_\ell(\theta^n_\ell) - Q_{\ell-1}(\theta^n_{\ell-1}))
$$
ALGORITHM 2. (Multilevel Metropolis Hastings MCMC) (NEW)

At states $\theta^n_0, \ldots, \theta^n_\ell$ with $(\theta^n_\ell)_{n \geq 0} = (\theta^n_\ell, J_\ell)_{n \geq 0}$ the chain on level $\ell$:

- Generate proposal $\theta'_0$ from $q^0(\theta'_0 | \theta^n_0) = q(\theta'_0 | \theta^n_0)$ (random walk)
- Accept $\theta'_0$ using standard MetH with acceptance probability

$$\alpha^0(\theta'_0 | \theta^n_0) = \min \left( 1, \frac{\pi^0(\theta'_0) q^0(\theta^n_0 | \theta'_0)}{\pi^0(\theta^n_0) q^0(\theta'_0 | \theta^n_0)} \right) = \min \left( 1, \frac{\pi^0(\theta'_0)}{\pi^0(\theta^n_0)} \right)$$
At states $\theta_0^n, \ldots, \theta_\ell^n$ with $(\theta_\ell^n)_{n \geq 0} = (\theta_\ell^n, J_\ell)_{n \geq 0}$ the chain on level $\ell$:

- Generate proposal $\theta'_0$ from $q^0(\theta'_0 | \theta_0^n) = q(\theta'_0 | \theta_0^n)$ (random walk)
- Accept $\theta'_0$ using standard MetH with acceptance probability

$$\alpha^0(\theta'_0 | \theta_0^n) = \min \left( 1, \frac{\pi^0(\theta'_0) q^0(\theta_0^n | \theta'_0)}{\pi^0(\theta_0^n) q^0(\theta'_0 | \theta_0^n)} \right) = \min \left( 1, \frac{\pi^0(\theta'_0)}{\pi^0(\theta_0^n)} \right)$$

- \ldots

- Propose $\theta'_\ell = [\theta_{\ell-1}^{n+1}, \theta'_\ell, \bot]$ with $\theta'_\ell, \bot$ generated from $q(\theta'_\ell, \bot | \theta^n_{\ell, \bot})$ (transition prob. $q^\ell$ depends on acceptance prob. $\alpha^{\ell-1}$ on level $\ell - 1$)
- Accept $\theta'_\ell$ with probability

$$\alpha^\ell(\theta'_\ell | \theta^n_\ell) = \min \left( 1, \frac{\pi^\ell(\theta'_\ell) q^\ell(\theta^n_\ell | \theta'_\ell)}{\pi^\ell(\theta^n_\ell) q^\ell(\theta'_\ell | \theta^n_\ell)} \right)$$
ALGORITHM 2. (Multilevel Metropolis Hastings MCMC) (NEW)

At states $\theta^n_0, \ldots, \theta^n_\ell$ with $(\theta^n_\ell)_{n \geq 0} = (\theta^n_\ell, J_\ell)_{n \geq 0}$ the chain on level $\ell$:

- Generate proposal $\theta'_0$ from $q^0(\theta'_0 | \theta^n_0) = q(\theta'_0 | \theta^n_0)$ (random walk)

- Accept $\theta'_0$ using standard MetH with acceptance probability

$$\alpha^0(\theta'_0 | \theta^n_0) = \min \left( 1, \frac{\pi^0(\theta'_0) q^0(\theta^n_0 | \theta'_0)}{\pi^0(\theta^n_0) q^0(\theta'_0 | \theta^n_0)} \right) = \min \left( 1, \frac{\pi^0(\theta'_0)}{\pi^0(\theta^n_0)} \right)$$

- $\ldots$

- Propose $\theta'_\ell = [\theta^{n+1}_{\ell-1}, \theta'_\ell, \bot]$ with $\theta'_\ell, \bot$ generated from $q(\theta'_\ell, \bot | \theta^n_\ell, \bot)$ (transition prob. $q^\ell$ depends on acceptance prob. $\alpha^{\ell-1}$ on level $\ell - 1$)

- Accept $\theta'_\ell$ with probability

$$\alpha^\ell(\theta'_\ell | \theta^n_\ell) = \min \left( 1, \frac{\pi^\ell(\theta'_\ell) \pi^{\ell-1}(\theta^n_{\ell, \text{coarse}})}{\pi^\ell(\theta^n_\ell) \pi^{\ell-1}(\theta^{n+1}_{\ell-1})} \right)$$

in the case of a random walk (proved by induction).
Multilevel Markov Chain Monte Carlo

- **Convergence**: If (as satisfied for the random walk)

\[ q^\ell(\theta_\ell | \theta^n_\ell) > 0 \quad \text{for all } (\theta_\ell, \theta^n_\ell) \text{ with } \pi^\ell(\theta_\ell), \pi^\ell(\theta^n_\ell) > 0 \]

then

\[ \lim_{\{N_\ell\} \to \infty} \hat{Q}^{ML} = \mathbb{E}_{\pi^L}[Q_L]. \]
Multilevel Markov Chain Monte Carlo

- **Convergence:** If (as satisfied for the random walk)
  \[ q^\ell(\theta_\ell | \theta_\ell^n) > 0 \]  for all \((\theta_\ell, \theta_\ell^n)\) with \(\pi^\ell(\theta_\ell), \pi^\ell(\theta_\ell^n) > 0\)
  then
  \[ \lim_{\{N_\ell\} \to \infty} \hat{Q}^{ML} = \mathbb{E}_{\pi^\ell}[Q_\ell]. \]

- **But** coarse modes of fine chain may differ on level \(\ell\) and \(\ell - 1\):

  \[
  \begin{array}{|c|c|c|}
  \hline
  \text{accept/accept} & \theta_{\ell-1}^{n+1} & \theta_\ell^{n+1} \\
  \hline
  \text{accept/reject} & \theta_{\ell-1} & [\theta_{\ell-1}^n, \theta_1^\ell, \bot] \\
  \hline
  \text{reject/accept} & \theta_{\ell-1}^n & [\theta_{\ell-1}^n, \theta_1^\ell, \bot] \\
  \hline
  \text{reject/reject} & \theta_{\ell-1}^n & [\theta_{\ell-1}^n, \theta_1^\ell, \bot] = \theta_1^n \\
  \hline
  \end{array}
  \]

  No guarantee that in that case variance is small (see below!)
Complexity Theorem for Multilevel MCMC

Let \( Y_\ell := Q_\ell - Q_{\ell-1} \) and assume there are constants \( \alpha, \beta, \gamma > 0 \) s.t. \( \alpha \geq \frac{1}{2} \min(\beta, \gamma) \) and

**M1.** \[ |\mathbb{E}_{\pi_\ell} [Q_\ell - Q]| \lesssim h_\ell^\alpha \quad \text{(FE-error)} \]

**M2.** \[ \nabla_{\text{MetH}} [\hat{Y}_\ell] + (\mathbb{E}_{\text{MetH}} [\hat{Y}_\ell] - \mathbb{E}_{\pi_\ell, \pi_{\ell-1}} [\hat{Y}_\ell])^2 \lesssim \frac{\nabla_{\pi_\ell, \pi_{\ell-1}} [Y_\ell]}{N_\ell} \]

**M3.** \[ \nabla_{\pi_\ell, \pi_{\ell-1}} [Y_\ell] \lesssim h_\ell^\beta \quad \text{(Variance decay)} \]

Then, there exist a value \( L \) and a sequence \( \{N_\ell\}_L^{L=0} \) such that the mean square error is less than \( \varepsilon^2 \) and

\[
C(\hat{Q}_L^{\text{ML}}) \lesssim \begin{cases} 
\varepsilon^{-2}, & \text{if } \beta > d, \\
\varepsilon^{-2}(\log \varepsilon)^2, & \text{if } \beta = d, \\
\varepsilon^{-2-(d-\beta)/\alpha}, & \text{if } \beta < d.
\end{cases}
\]
Analysis (to satisfy assumptions M1-M3)

Lemma (Prior/Posterior)

(a) For any random variable $Y = Y(\theta_{\ell})$ with $\mathbb{E}_{\mathcal{P}_\ell} [|Y|^p] < \infty$

$$|\mathbb{E}_{\pi_{\ell}} [Y^p]| \lesssim \mathbb{E}_{\mathcal{P}_\ell} [|Y|^p].$$

(b) For any $Y = Y(\theta_{\ell}, \theta_{\ell-1})$ with $\mathbb{E}_{\mathcal{P}_\ell, \mathcal{P}_{\ell-1}} [|Y|^p] < \infty$

$$|\mathbb{E}_{\pi_{\ell}, \pi_{\ell-1}} [Y^p]| \lesssim \mathbb{E}_{\mathcal{P}_\ell, \mathcal{P}_{\ell-1}} [|Y|^p].$$
Analysis (to satisfy assumptions M1-M3)

Lemma (Prior/Posterior)

(a) For any random variable $Y = Y(\theta_\ell)$ with $E_{P_\ell} [|Y|^p] < \infty$

$$|E_{\pi_\ell} [Y^p]| \lesssim E_{P_\ell} [|Y|^p].$$

(b) For any $Y = Y(\theta_\ell, \theta_{\ell-1})$ with $E_{P_\ell, P_{\ell-1}} [|Y|^p] < \infty$

$$|E_{\pi_\ell, \pi_{\ell-1}} [Y^p]| \lesssim E_{P_\ell, P_{\ell-1}} [|Y|^p].$$

• Thus to bound the bias (M1) we can use the earlier results.

• For the variance reduction (M3) we need on further result.
Analysis (to satisfy assumptions M1-M3)

Key Lemma
Suppose observation operator $F^h$ is linear and $F^h(u) \lesssim \|u\|_{L^2(D)}$ for $u \in H^1_0(D)$. Then

$$\lim_{\ell \to \infty} \alpha^\ell (\theta_\ell | \theta^\ell_n) = 1,$$
for $\mathcal{P}_\ell$-almost all $\theta_\ell, \theta^\ell_n$.

Furthermore,

$$\mathbb{E}_{\mathcal{P}_\ell, \mathcal{P}_\ell} \left[ (1 - \alpha^\ell)^q \right]^{1/q} \lesssim h^{1-\delta}_\ell + J_{\ell-1}^{-1/2+\delta},$$
for any $q < \infty$ and $\delta > 0$.

Proof: Uses explicit expressions for the likelihood and the prior, together with error bounds shown previously.
Analysis (to satisfy assumptions M1-M3)

Lemma (Variance Reduction)

Let $\theta^n_\ell$ and $\theta^n_{\ell-1}$ be from Algorithm 2. Then

$$\nabla_{\pi^\ell,\pi^{\ell-1}} \left[ Q_\ell(\theta^n_\ell) - Q_{\ell-1}(\theta^n_{\ell-1}) \right] \preceq h_\ell^{1-\delta} + J_\ell^{-1/2+\delta},$$

for any $\delta > 0$.

Proof: Use $\nabla[Y] \leq \mathbb{E}[Y^2]$, and prior/posterior lemma above.
Analysis (to satisfy assumptions M1-M3)

Lemma (Variance Reduction)

Let $\theta_n^\ell$ and $\theta_{n-1}^\ell$ be from Algorithm 2. Then

$$\nabla_{\pi^\ell, \pi_{\ell-1}} \left[ Q_\ell(\theta_n^\ell) - Q_{\ell-1}(\theta_{n-1}^\ell) \right] \lesssim h^{1-\delta}_\ell + J_{\ell}^{-1/2+\delta},$$

for any $\delta > 0$.

Proof: Use $\nabla[Y] \leq \mathbb{E}[Y^2]$, and prior/posterior lemma above.

Two cases:

- $\theta_n^\ell$ and $\theta_{n-1}^\ell$ have the same coarse modes $\Rightarrow$ result follows
- $\theta_n^\ell$ and $\theta_{n-1}^\ell$ differ on coarse modes; this only happens with probability $1 - \alpha^\ell$ (very small on fine levels):

$$\mathbb{E} \left[ I_{\{\text{differ}\}} \right] \leq \mathbb{E}_{P_{\ell}, P_{\ell}} \left[ 1 - \alpha^\ell(\theta_n^\ell | \theta') \right].$$
Verification of multilevel assumptions

- **M1.** \( |E_{π^ℓ}[Q^ℓ - Q]| \lesssim h_ℓ^α \) satisfied with \( α = 1 - δ \)

- **M2.** \( \nabla_{\text{alg}}[\hat{Y}_ℓ] + (E_{\text{alg}}[\hat{Y}_ℓ] - E_{π^ℓ,π_ℓ−1}[\hat{Y}_ℓ])^2 \lesssim N_ℓ^{-1} \nabla_{π^ℓ,π_ℓ−1}[Y_ℓ] \)
  satisfied for certain proposal distributions (e.g. preconditioned random walks), see recent work by Hairer, Stuart et al

- **M3.** \( \nabla_{π^ℓ,π_ℓ−1}[Y_ℓ] \lesssim h_ℓ^β \) satisfied with \( β = 1 - δ \)
Verification of multilevel assumptions

- **M1.** \( |\mathbb{E}_{\pi^\ell}[Q^\ell - Q]| \lesssim h^\alpha_\ell \) satisfied with \( \alpha = 1 - \delta \)

- **M2.** \( \nabla_{\text{alg}}[\hat{Y}^\ell] + (\mathbb{E}_{\text{alg}}[\hat{Y}^\ell] - \mathbb{E}_{\pi^\ell,\pi^{\ell-1}}[\hat{Y}^\ell])^2 \lesssim N^{-1}_\ell \nabla_{\pi^\ell,\pi^{\ell-1}}[Y^\ell] \)
  satisfied for certain proposal distributions (e.g. preconditioned random walks), see recent work by Hairer, Stuart et al

- **M3.** \( \nabla_{\pi^\ell,\pi^{\ell-1}}[Y^\ell] \lesssim h^\beta_\ell \) satisfied with \( \beta = 1 - \delta \)

This gives the following (theoretical) costs to get a RMSE of \( \mathcal{O}(\varepsilon) \).

<table>
<thead>
<tr>
<th>( d )</th>
<th>MLMC MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \mathcal{O}(\varepsilon^{-2}) )</td>
</tr>
<tr>
<td>2</td>
<td>( \mathcal{O}(\varepsilon^{-3}) )</td>
</tr>
<tr>
<td>3</td>
<td>( \mathcal{O}(\varepsilon^{-4}) )</td>
</tr>
</tbody>
</table>

Less savings than in standard MC case due to divergence of chains!
What about final assumption in complexity theorem that \( \text{cost/sample} = O(h^{-d}) \)? Is it realistic?
Multilevel Deterministic Solvers
Theory – How realistic is cost/sample = \( O(h^{-d}) \)?

Multilevel deterministic solvers for elliptic problems with rough coefficients

- Sampling \( k \): FFT-based circulant embedding \( \Rightarrow \sim O(h^{-d}) \)
  (but only for stationary correlation functions & parallel efficiency?)
Theory – How realistic is $\text{cost/sample} = \mathcal{O}(h^{-d})$?

Multilevel deterministic solvers for elliptic problems with rough coefficients

- **Sampling $k$:** FFT-based **circulant embedding** $\Rightarrow \sim \mathcal{O}(h^{-d})$
  (but only for stationary correlation functions & parallel efficiency?)

- **PDE solve:** **Multigrid** methods $\Rightarrow \sim \mathcal{O}(h^{-d})$
  **but** constant depends strongly on variability of coefficient!
Theory — How realistic is \( \text{cost/sample} = \mathcal{O}(h^{-d}) \)?

Multilevel deterministic solvers for elliptic problems with rough coefficients

- Sampling \( k \): FFT-based **circulant embedding** \( \Rightarrow \sim \mathcal{O}(h^{-d}) \)
  (but only for stationary correlation functions & parallel efficiency?)

- PDE solve: **Multigrid** methods \( \Rightarrow \sim \mathcal{O}(h^{-d}) \)
  but constant depends strongly on variability of coefficient!

- Analysis of this robustness is talk in itself
  - Geom. MG: [Xu, Zhu, 08], [RS, Vassilevski, Zikatanov, 12], ...
  - Coeff depend: [Graham, Lechner, RS, 07], [Galvis, Efendiev, 10],...
Theory – How realistic is \( \text{cost/sample} = \mathcal{O}(h^{-d}) \)?

Multilevel deterministic solvers for elliptic problems with rough coefficients

- Sampling \( k \): FFT-based **circulant embedding** \( \Rightarrow \sim \mathcal{O}(h^{-d}) \)
  (but only for stationary correlation functions & parallel efficiency?)

- PDE solve: **Multigrid** methods \( \Rightarrow \sim \mathcal{O}(h^{-d}) \)
  but constant depends strongly on variability of coefficient!

- Analysis of this robustness is talk in itself
  - Geom. MG: [Xu, Zhu, 08], [RS, Vassilevski, Zikatanov, 12], ...
  - Coeff depend: [Graham, Lechner, RS, 07], [Galvis, Efendiev, 10], ...

- In practice: **Algebraic Multigrid (AMG)** \( \Rightarrow \mathcal{O}(h^{-d}) \) robustly
Scalability of AMG (lognormal coefficients)

Algorithmic Scalability (i.e. #iterations in practice)

- Classical Ruge-Stüben AMG (e.g. BoomerAMG, LLNL) fully robust
Scalability of AMG (lognormal coefficients)

**Algorithmic Scalability** (i.e. \#iterations in practice)

- Classical Ruge-Stüben AMG (e.g. BoomerAMG, LLNL) fully robust
- Unsmoothed aggregation-type AMG (e.g. DUNE, Heidelberg) also fully robust and scales **quasi-optimally** \#its= \( \mathcal{O}(|\log(h)|) \)!
Scalability of AMG (lognormal coefficients)

**Algorithmic Scalability** (i.e. \#iterations in practice)

- Classical Ruge-Stüben AMG (e.g. BoomerAMG, LLNL) fully robust
- Unsmoothed aggregation-type AMG (e.g. DUNE, Heidelberg) also fully robust and scales **quasi-optimally** \#its= \( O(\mid \log(h) \mid) \)!
- In MCMC case: possible to “recycle” AMG-setup or use **adaptive AMG** [Ketelsen, Vassilevski, 2012]
Scalability of AMG (lognormal coefficients)

Parallel (& Algorithmic) Scalability

- Unsmoothed aggregation AMG (DUNE, [Blatt, Ippisch, 2011]):
  
  \[ (d = 3, \sigma^2 = 1, h = 1/80 \text{ on 1 proc, largest problem } M_h \sim 1.5 \times 10^{11}) \]

Weak scaling (JUGENE, Jülich, 290K procs)
Scalability of AMG (lognormal coefficients)

Parallel (& Algorithmic) Scalability

- Unsmoothed aggregation AMG ([DUNE, [Blatt, Ippisch, 2011]]): 
  
  \( (d = 3, \sigma^2 = 1, h = 1/80 \text{ on 1 proc, largest problem } M_h \sim 1.5 \times 10^{11}) \)

Weak scaling ([JUGENE, Jülich, 290K procs]) – BoomerAMG, LLNL similar

- Extension to multiphase via CPR precond.
Conclusions

- UQ & uncertainty propagation → PDEs with random coeffs
- Multilevel MC: Same cost as deterministic solver
- **New**: Multilevel Markov chain Monte Carlo (w. theory!)
- Multilevel deterministic solvers: scalability & coeff. robustness
Conclusions

- UQ & uncertainty propagation → PDEs with random coeffs
- Multilevel MC: Same cost as deterministic solver
- **New**: Multilevel Markov chain Monte Carlo (w. theory!)
- Multilevel deterministic solvers: scalability & coeff. robustness

Further Work

- Large 3D MLMC computations (parallelise whole algorithm)
- Quasi Monte Carlo: deterministic sampling rules (w. theory!)
- Efficient sampling of $k$ (in parallel): GPUs, PDE-based,...
- **Multilevel MCMC** for multiphase flow
- Massively parallel multilevel solvers for multiphase flow
Thank You!
Thank You!

Preprints available on my website:

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

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