Multilevel Monte Carlo Methods for Large Scale Problems

Robert Scheichl

Department of Mathematical Sciences
University of Bath

Collaborators:
A. Teckentrup & E. Ullmann (Bath); J. Charrier (Marseille);
K.A. Cliffe & M. Park (Nottingham); M. Giles (Oxford);
C. Ketelsen & P. Vassilevski (LLNL)

Newton Institute Workshop on Weather & Climate Prediction on Next Generation Supercomputers

Met Office, Exeter, 24th October 2012

Stochastic Modelling and Supercomputers

- Many reasons for stochastic modelling in earth sciences:
  - **lack of data** (e.g. data assimilation for weather prediction)
  - **uncertainty in data** (e.g. uncertainty quantification in subsurface flow)
  - **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **Input:** best knowledge about system, statistics of input parameters, measured data with error statistics, etc...
- **Output:** statistics of quantities of interest or of entire state space
  - Data assimilation in NWP: data misfit, rainfall at some location
  - Radioactive waste disposal: flow at repository, travel time
  - Oil reservoir simulation: production rate
  - Atmospheric dispersion: amount of ash over Heathrow

Stochastic Modelling and Supercomputers

- Many reasons for stochastic modelling in earth sciences:
  - **lack of data** (e.g. data assimilation for weather prediction)
  - **uncertainty in data** (e.g. uncertainty quantification in subsurface flow)
  - **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **Input:** best knowledge about system, statistics of input parameters, measured data with error statistics, etc...
- **Output:** statistics of quantities of interest or of entire state space

Main workhorse: Monte Carlo type Methods

☺ **No** curse of dimensionality
☹ Convergence rate $O(N^{-1/2})$

Stochastic Modelling and Supercomputers

- Many reasons for stochastic modelling in earth sciences:
  - **lack of data** (e.g. data assimilation for weather prediction)
  - **uncertainty in data** (e.g. uncertainty quantification in subsurface flow)
  - **unresolvable scales** (e.g. atmospheric dispersion modelling)
- **Input:** best knowledge about system, statistics of input parameters, measured data with error statistics, etc...
- **Output:** statistics of quantities of interest or of entire state space

Main workhorse: Monte Carlo type Methods

☺ **No** curse of dimensionality

☹ Convergence rate $O(N^{-1/2})$

Novel Variants!

Monte Carlo for large scale problems (plain vanilla)

$$\mathbf{Z}_J(\omega) \in \mathbb{R}^J \xrightarrow{\text{Model}(M)} \mathbf{X}_M(\omega) \in \mathbb{R}^M \xrightarrow{\text{Output}} Q_{M,J}(\omega) \in \mathbb{R}$$

random input | state vector | quantity of interest

- e.g. $\mathbf{Z}_J$ multivariate Gaussian; $\mathbf{X}_M$ numerical solution of PDE; $Q_{M,J}$ a (non)linear functional of $\mathbf{X}_M$

Monte Carlo for large scale problems (plain vanilla)

$$\mathbf{Z}_J(\omega) \in \mathbb{R}^J \quad \stackrel{\text{Model}(M)}{\longrightarrow} \quad \mathbf{X}_M(\omega) \in \mathbb{R}^M \quad \stackrel{\text{Output}}{\longrightarrow} \quad Q_{M,J}(\omega) \in \mathbb{R}$$

random input — state vector — quantity of interest

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

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

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

Monte Carlo for large scale problems (plain vanilla)

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

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

$$= \underbrace{\frac{\mathbb{V}[Q_{M,J}]}{N}}_{\textbf{sampling error}} + \underbrace{\left(\mathbb{E}[Q_{M,J} - Q]\right)^2}_{\textbf{model error ("bias")}}$$

- Typical: $\alpha = 1/2 \ \Rightarrow\ \mathrm{RMSE} = \mathcal{O}(N^{-1/2}) + \mathcal{O}(M^{-1/2})$

Monte Carlo for large scale problems (plain vanilla)

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

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

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

- Typical: $\alpha = 1/2 \;\Rightarrow\; \mathrm{RMSE} = \mathcal{O}(N^{-1/2}) + \mathcal{O}(M^{-1/2})$

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

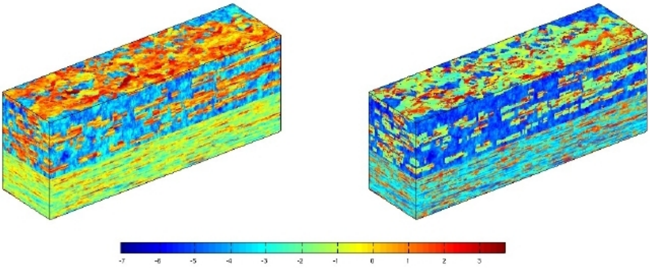
  (e.g. for $\mathrm{TOL} = 10^{-3}$ we get $M \sim N \sim 10^6$ and $\mathrm{Cost} = \mathcal{O}(10^{12})$!!)

- Quickly becomes **prohibitively expensive !**

Example: Uncertainty in Subsurface Flow

(eg. risk analysis of radiocative waste disposal or oil reservoir simulation)

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



Society of Petroleum Engineers Benchmark SPE10

Example: Uncertainty in Subsurface Flow

(eg. risk analysis of radiocative waste disposal or oil reservoir simulation)

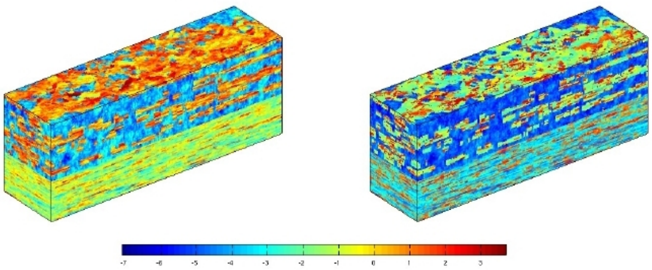
uncertain k $\rightarrow$

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

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

$+$ **Boundary Conditions**

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



Society of Petroleum Engineers Benchmark SPE10

Stochastic Modelling of Uncertainty:

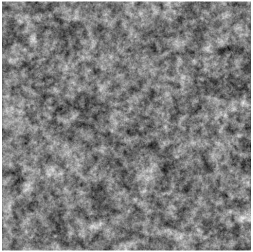
Model uncertain conductivity tensor k as a **lognormal** random field

Typical simplified model (prior):

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

  meanfree with exponential covariance:

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



typical realisation

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

Typical quantities of interest:

- effective conductivity $k_{\mathrm{eff},1} = \frac{1}{|D|}\int_D q_1$
- $q_1(x^*)$; travel time; water cut; etc...

Stochastic Modelling of Uncertainty:

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

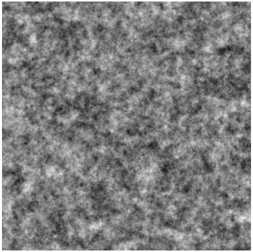
Typical simplified model (prior):

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

  meanfree with exponential covariance:

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

- Incorporating data (**Posterior**) → **MCMC**



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

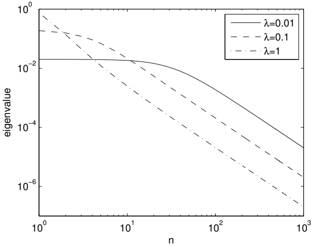
Typical quantities of interest:

- effective conductivity $k_{\text{eff},1} = \frac{1}{|D|} \int_D q_1$
- $q_1(x^*)$; travel time; water cut; etc...

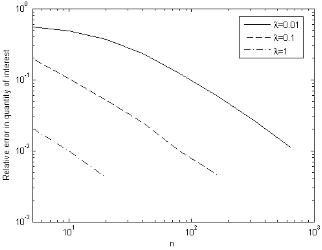
Sampling from k

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

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



KL-eigenvalues in 1D

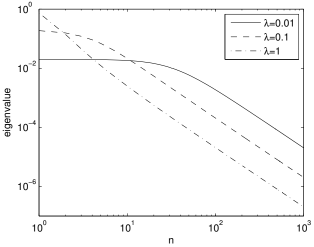


Truncation error for $q|_{x=1}$ wrt. J

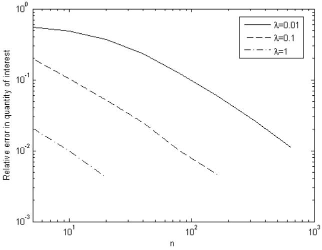
Sampling from k

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

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



KL-eigenvalues in 1D

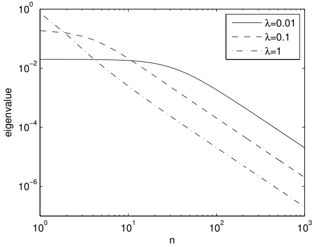


Truncation error for $q|_{x=1}$ wrt. J

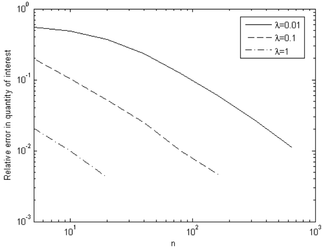
Sampling from k

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

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



KL-eigenvalues in 1D



Truncation error for $q|_{x=1}$ wrt. J

- Large #KL-modes for small λ $\Longrightarrow$ **high** dimension $J \gg 100$
- Low regularity (Hölder with $\eta < \frac{1}{2}$) $\Longrightarrow$ **fine** numerical mesh $\Delta x \ll 1$

Model: Spatial discretisation with $M = \mathcal{O}(\Delta x^{-d})$ DOFs

- **Standard pw. linear FEs** or **cell-centred FVs**:

$$-\nabla \cdot (k(x,\omega)\nabla p(x,\omega)) = f(x) \qquad \longrightarrow \qquad A(\omega)\,\mathbf{X}_M(\omega) = \mathbf{b}(\omega)$$

random elliptic PDE — random $M \times M$ linear system

(similarly with Mixed FEs)

- **Quantity of interest:** Expected value $\mathbb{E}[Q]$ of $Q := \mathcal{G}(p)$
functional of the PDE solution p; in practice compute $Q_{M,J} := \mathcal{G}_M(\mathbf{X}_M)$

Model: Spatial discretisation with $M = \mathcal{O}(\Delta x^{-d})$ DOFs

- **Standard pw. linear FEs** or **cell-centred FVs**:

$$-\nabla \cdot (k(x,\omega)\nabla p(x,\omega)) = f(x) \qquad \longrightarrow \qquad A(\omega)\,\mathbf{X}_M(\omega) = \mathbf{b}(\omega)$$

  random elliptic PDE — random $M \times M$ linear system

  (similarly with Mixed FEs)

- **Quantity of interest:** Expected value $\mathbb{E}[Q]$ of $Q := \mathcal{G}(p)$
  functional of the PDE solution p; in practice compute $Q_{M,J} := \mathcal{G}_M(\mathbf{X}_M)$

- Recall **(plain vanilla) Monte Carlo (MC)** estimate for $\mathbb{E}[Q]$:

$$\hat{Q}^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_{M,J}^{(i)}, \qquad Q_{M,J}^{(i)} \text{ i.i.d. samples with Model}(M)\,.$$

- Assume optimal **multigrid solver** $\Rightarrow$ $\mathrm{Cost}(Q_{M,J}^{(i)}) = \mathcal{O}(M)$

What is cost to get RMSE $<$ TOL?

Complexity Theorem for (plain vanilla) Monte Carlo

Assume that $Q_{M,J} \to Q$ with $\mathcal{O}(M^{-\alpha})$ for some $\alpha > 0$, then to obtain RMSE $<$ TOL

$$\text{Cost}(\hat{Q}^{\mathrm{MC}}) \;=\; \mathcal{O}\left(\text{TOL}^{-2-\frac{1}{\alpha}}\right)$$

What is cost to get RMSE < TOL?

Complexity Theorem for (plain vanilla) Monte Carlo

Assume that $Q_{M,J} \to Q$ with $\mathcal{O}(M^{-\alpha})$ for some $\alpha > 0$, then to obtain RMSE < TOL

$$\text{Cost}(\hat{Q}^{\text{MC}}) = \mathcal{O}\left(\text{TOL}^{-2-\frac{1}{\alpha}}\right)$$

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

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

TOL	M	N	Cost
0.01	1.7×10^4	1.4×10^4	21 min
0.002	1.1×10^6	3.5×10^5	30 days

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

TOL	M	N	Cost
0.01	2.6×10^5	8.5×10^3	4 h
0.002	**Prohibitively large!!**		

Here $d = 2$ & $\alpha \approx 3/8 \Rightarrow$ Cost $\approx \mathcal{O}(\text{TOL}^{-14/3}) \approx 25 \times$ more work to halve error!

Multilevel Monte Carlo

[Heinrich, '01], [Giles, '07]
[Cliffe, Giles, RS, Teckentrup, '11]

Note that trivially

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

(where $\Delta x_\ell = \Delta x_{\ell-1}/2$, $M_\ell = 2^d M_{\ell-1}$ and $Q_\ell := Q_{M_\ell, J_\ell}$)

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 $\Delta x_\ell = \Delta x_{\ell-1}/2$, $M_\ell = 2^d M_{\ell-1}$ and $Q_\ell := Q_{M_\ell, J_\ell}$)

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]
[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 $\Delta x_\ell = \Delta x_{\ell-1}/2$, $M_\ell = 2^d M_{\ell-1}$ and $Q_\ell := Q_{M_\ell, J_\ell}$)

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 (multigrid idea: compute corrections)

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

Complexity Theorem for Multilevel Monte Carlo

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

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(M_\ell^{-\beta}).$$

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

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \; = \; \mathcal{O}\left(\text{TOL}^{-2-\max\left(0, \frac{1-\beta}{\alpha}\right)}\right) \; + \text{possible log-factor}$$

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

Complexity Theorem for Multilevel Monte Carlo

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

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(M_\ell^{-\beta}).$$

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

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\text{TOL}^{-2-\max\left(0,\frac{1-\beta}{\alpha}\right)}\right) \quad + \text{ possible log-factor}$$

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

If $\beta \sim 2\alpha$ (as in example above) then

$$\text{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\text{TOL}^{-\max\left(2,\frac{1}{\alpha}\right)}\right) \;=\; \mathcal{O}\left(\max(N_0, M_L)\right)$$

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

Complexity Theorem for Multilevel Monte Carlo

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

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(M_\ell^{-\beta}).$$

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

$$\mathrm{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\mathrm{TOL}^{-2-\max\left(0,\frac{1-\beta}{\alpha}\right)}\right) \quad + \text{ possible log-factor}$$

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

If $\beta \sim 2\alpha$ (as in example above) then

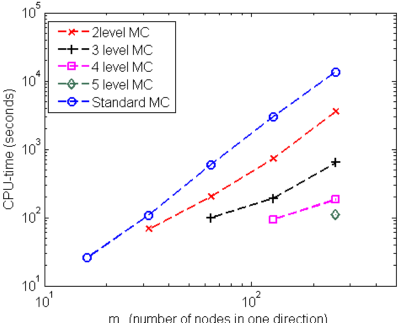
$$\mathrm{Cost}(\widehat{Q}_L^{\mathcal{ML}}) \;=\; \mathcal{O}\left(\mathrm{TOL}^{-\max\left(2,\frac{1}{\alpha}\right)}\right) \;=\; \mathcal{O}\left(\max(N_0, M_L)\right)$$

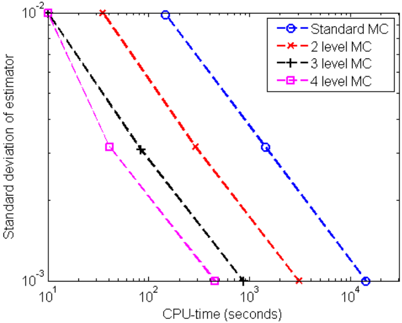
For $\alpha \approx 3/8$ (in example above): $\mathcal{O}(\mathrm{TOL}^{-8/3})$ instead of $\mathcal{O}(\mathrm{TOL}^{-14/3})$

Asymptotically **same** as **one deterministic solve** to the same accuracy **!**

Numerical Example

$D = (0,1)^2$, $\sigma^2 = 1$, $\lambda = 0.1$, $\Delta x_0 = \frac{1}{16}$, $J_L = 500$, standard FE, UMFPACK





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

$Q = k^2_{\text{eff},1}$, $\Delta x_L^{-1} = 256$
(second moment!)

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

Here with **sparse direct solver**, i.e. cost/sample $\approx \mathcal{O}(M^{3/2})$ (not $\mathcal{O}(M)$)

Theory: Verifying Assumptions for Subsurface Application

- **[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 all $\eta < \frac{1}{2}$
- **[Teckentrup, RS, Giles, Ullmann, 2012]**: extension to (nonlinear) functionals, corner domains, discont. coeffs, level-dependent truncations
- New **regularity** result: all finite moments of H^{1+s}-norm of p bdd. $(\forall s < \frac{1}{2})$
- New **FE error** result: all finite moments of H^1-error are $O(\Delta x^s)$ $(\forall s < \frac{1}{2})$

Leads to $\alpha = 1/d$, $\beta = 2/d$ for linear & for Fréchet diff'ble functionals $Q := \mathcal{G}(p)$

Multilevel Markov Chain Monte Carlo

Can the multilevel idea be extended to MCMC?

Multilevel Markov Chain Monte Carlo

Can the multilevel idea be extended to MCMC? **YES!**

Incorporating Data – Bayes' Theorem

- Our model was parametrised by $\mathbf{Z}_J := [Z_1, \ldots, Z_J]$ (the **"prior"**).
  In the subsurface flow application:

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

- Usually also **some data** $F_{\rm obs}$ related to **outputs** (e.g. pressure), available. To reduce uncertainty, incorporate $F_{\rm obs}$ (the **"posterior"**)

Incorporating Data – Bayes' Theorem

- Our model was parametrised by $\mathbf{Z}_J := [Z_1, \dots, Z_J]$ (the **"prior"**).
  In the subsurface flow application:
  $$\log k \approx \sum_{j=1}^{J} \sqrt{\nu_j}\phi_j(x) Z_j(\omega) \quad \text{and} \quad \mathcal{P}(\mathbf{Z}_J) \eqsim (2\pi)^{-J/2} \prod_{j=1}^{J} \exp\left(-\frac{Z_j^2}{2}\right)$$
- Usually also **some data** F_{obs} related to **outputs** (e.g. pressure), available. To reduce uncertainty, incorporate F_{obs} (the **"posterior"**)

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

$$\underbrace{\pi^{M,J}(\mathbf{Z}_J)}_{\textbf{posterior}} := \mathcal{P}(\mathbf{Z}_J \,|\, F_{\mathrm{obs}}) \eqsim \underbrace{\mathcal{L}_M(F_{\mathrm{obs}} \,|\, \mathbf{Z}_J)}_{\textbf{likelihood}} \underbrace{\mathcal{P}(\mathbf{Z}_J)}_{\textbf{prior}}$$

Incorporating Data – Bayes' Theorem

- Our model was parametrised by $\mathbf{Z}_J := [Z_1, \ldots, Z_J]$ (the **"prior"**).
  In the subsurface flow application:

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

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

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

$$\underbrace{\pi^{M,J}(\mathbf{Z}_J)}_{\text{posterior}} := \mathcal{P}(\mathbf{Z}_J \,|\, F_{\mathrm{obs}}) \eqsim \underbrace{\mathcal{L}_M(F_{\mathrm{obs}} \,|\, \mathbf{Z}_J)}_{\text{likelihood}} \underbrace{\mathcal{P}(\mathbf{Z}_J)}_{\text{prior}}$$

- **Likelihood model** (e.g. Gaussian):

$$\mathcal{L}_M(F_{\mathrm{obs}} \,|\, \mathbf{Z}_J) \eqsim \exp\left(\frac{-\|F_{\mathrm{obs}} - F_M(\mathbf{Z}_J)\|^2}{\sigma^2_{\mathrm{fid},M}}\right)$$

$F_M(\mathbf{Z}_J)$ ... model response; $\sigma_{\mathrm{fid},M}$ ... fidelity parameter (M-dep.)

ALGORITHM 1. (Standard Metropolis Hastings MCMC)

- Choose $\mathbf{Z}_J^0$.
- At state $\mathbf{Z}_J^n$ generate proposal $\mathbf{Z}'_J$ from distribution $q(\mathbf{Z}'_J \mid \mathbf{Z}_J^n)$ (e.g. random walk).
- Accept $\mathbf{Z}'_J$ as a sample with probability

$$\alpha^{M,J} = \min\left(1, \frac{\pi^{M,J}(\mathbf{Z}'_J)\, q(\mathbf{Z}_J^n \mid \mathbf{Z}'_J)}{\pi^{M,J}(\mathbf{Z}_J^n)\; q(\mathbf{Z}'_J \mid \mathbf{Z}_J^n)}\right) \;=\; \overbrace{\min\left(1, \frac{\pi^{M,J}(\mathbf{Z}'_J)}{\pi^{M,J}(\mathbf{Z}_J^n)}\right)}^{\text{for random walk}}$$

i.e. $\mathbf{Z}_J^{n+1} = \mathbf{Z}'_J$ with probability $\alpha^{M,J}$; otherwise stay at $\mathbf{Z}_J^{n+1} = \mathbf{Z}_J^n$.

ALGORITHM 1. (Standard Metropolis Hastings MCMC)

- Choose $\mathbf{Z}_J^0$.
- At state $\mathbf{Z}_J^n$ generate proposal $\mathbf{Z}'_J$ from distribution $q(\mathbf{Z}'_J \,|\, \mathbf{Z}_J^n)$ (e.g. random walk).
- Accept $\mathbf{Z}'_J$ as a sample with probability

$$\alpha^{M,J} = \min\left(1, \frac{\pi^{M,J}(\mathbf{Z}'_J)\, q(\mathbf{Z}_J^n \,|\, \mathbf{Z}'_J)}{\pi^{M,J}(\mathbf{Z}_J^n)\; q(\mathbf{Z}'_J \,|\, \mathbf{Z}_J^n)}\right) = \overbrace{\min\left(1, \frac{\pi^{M,J}(\mathbf{Z}'_J)}{\pi^{M,J}(\mathbf{Z}_J^n)}\right)}^{\text{for random walk}}$$

i.e. $\mathbf{Z}_J^{n+1} = \mathbf{Z}'_J$ with probability $\alpha^{M,J}$; otherwise stay at $\mathbf{Z}_J^{n+1} = \mathbf{Z}_J^n$.

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

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

where $Q_{M,J}^{(n)} = \mathcal{G}\big(\mathbf{X}_M(\mathbf{Z}_J^{(n)})\big)$ is the nth sample of Q using Model(M).

ALGORITHM 1. (Standard Metropolis Hastings MCMC)

- Choose $\mathbf{Z}_J^0$.
- At state $\mathbf{Z}_J^n$ generate proposal $\mathbf{Z}'_J$ from distribution $q(\mathbf{Z}'_J \,|\, \mathbf{Z}_J^n)$ (e.g. random walk).
- Accept $\mathbf{Z}'_J$ as a sample with probability

$$\alpha^{M,J} = \min\left(1, \frac{\pi^{M,J}(\mathbf{Z}'_J)\, q(\mathbf{Z}_J^n \,|\, \mathbf{Z}'_J)}{\pi^{M,J}(\mathbf{Z}_J^n)\, q(\mathbf{Z}'_J \,|\, \mathbf{Z}_J^n)}\right) = \overbrace{\min\left(1, \frac{\pi^{M,J}(\mathbf{Z}'_J)}{\pi^{M,J}(\mathbf{Z}_J^n)}\right)}^{\text{for random walk}}$$

i.e. $\mathbf{Z}_J^{n+1} = \mathbf{Z}'_J$ with probability $\alpha^{M,J}$; otherwise stay at $\mathbf{Z}_J^{n+1} = \mathbf{Z}_J^n$.

Pros:

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

Cons:

- Evaluation of $\alpha^{M,J}$ very expensive for large M.
- Acceptance rate $\alpha^{M,J}$ very low for large J $(< 10\%)$.

Multilevel Markov Chain Monte Carlo

as above $\Delta x_\ell = \Delta x_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, and we set $Q_\ell := Q_{M_\ell, J_\ell}$

What were the **key ingredients**?

- Models with less DOFs on coarser levels **much cheaper** to solve.
- $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$ $\Rightarrow$ **fewer samples** on finer levels
- **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

as above $\Delta x_\ell = \Delta x_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, and we set $Q_\ell := Q_{M_\ell, J_\ell}$

What were the **key ingredients**?

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

Important: In MCMC setting **target distribution depends on** ℓ:

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

Multilevel Markov Chain Monte Carlo

as above $\Delta x_\ell = \Delta x_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, and we set $Q_\ell := Q_{M_\ell, J_\ell}$

What were the **key ingredients**?

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

Important: In MCMC setting **target distribution depends on** ℓ:

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

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

Multilevel Markov Chain Monte Carlo

as above $\Delta x_\ell = \Delta x_{\ell-1}/2$ and $J_\ell = 2J_{\ell-1}$, and we set $Q_\ell := Q_{M_\ell, J_\ell}$

What were the **key ingredients**?

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

Important: In MCMC setting **target distribution depends on** ℓ:

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

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

Idea: Split $\mathbf{Z}_\ell^n = [\mathbf{z}_{\ell,\mathrm{lo}}^n, \mathbf{z}_{\ell,\mathrm{hi}}^n] = \boxed{Z_{\ell,1}^n, \ldots \text{ coarse } \ldots, Z_{\ell,J_0}^n, \; Z_{\ell,J_0+1}^n, \text{ fine}, Z_{\ell,J_0}^n}$

(proposals use coarse modes from level $\ell - 1 \to$ different proposal distribution q^ℓ)

ALGORITHM 2. (Multilevel Metropolis Hastings MCMC) (NEW)

At states $\mathbf{Z}_0^n, \ldots, \mathbf{Z}_\ell^n$

- Generate proposal $\mathbf{Z}_0'$ from $\mathrm{q}^0(\mathbf{Z}_0' \,|\, \mathbf{Z}_0^n) = q(\mathbf{Z}_0' \,|\, \mathbf{Z}_0^n)$ (random walk)
- Accept $\mathbf{Z}_0'$ using standard MetH with acceptance probability

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

ALGORITHM 2. (Multilevel Metropolis Hastings MCMC) (NEW)

At states $\mathbf{Z}_0^n, \ldots, \mathbf{Z}_\ell^n$

- Generate proposal $\mathbf{Z}_0'$ from $\mathrm{q}^0(\mathbf{Z}_0' \,|\, \mathbf{Z}_0^n) = q(\mathbf{Z}_0' \,|\, \mathbf{Z}_0^n)$ (random walk)
- Accept $\mathbf{Z}_0'$ using standard MetH with acceptance probability

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

- ...
- Propose $\mathbf{Z}_\ell' = [\mathbf{Z}_{\ell-1}^{n+1}, \mathbf{z}_{\ell,\mathrm{hi}}']$ with $\mathbf{z}_{\ell,\mathrm{hi}}'$ generated from $q(\mathbf{z}_{\ell,\mathrm{hi}}' \,|\, \mathbf{z}_{\ell,\mathrm{hi}}^n)$

  (now transition prob. q^ℓ depends on acceptance prob. $\alpha^{\ell-1}$ on level $\ell - 1$)
- Accept $\mathbf{Z}_\ell'$ with probability

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

ALGORITHM 2. (Multilevel Metropolis Hastings MCMC) (NEW)

At states $\mathbf{Z}_0^n, \dots, \mathbf{Z}_\ell^n$

- Generate proposal $\mathbf{Z}_0'$ from $\mathrm{q}^0(\mathbf{Z}_0' \,|\, \mathbf{Z}_0^n) = q(\mathbf{Z}_0' \,|\, \mathbf{Z}_0^n)$ (random walk)
- Accept $\mathbf{Z}_0'$ using standard MetH with acceptance probability

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

- ...
- Propose $\mathbf{Z}_\ell' = [\mathbf{Z}_{\ell-1}^{n+1}, \mathbf{z}_{\ell,\mathrm{hi}}']$ with $\mathbf{z}_{\ell,\mathrm{hi}}'$ generated from $q(\mathbf{z}_{\ell,\mathrm{hi}}' \,|\, \mathbf{z}_{\ell,\mathrm{hi}}^n)$

  (now transition prob. q^ℓ depends on acceptance prob. $\alpha^{\ell-1}$ on level $\ell - 1$)

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

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

  in the case of a random walk **(Proof by induction!)**.

Multilevel MCMC – What can we prove?

Hot off the press! **[Ketelsen, RS, 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, a bit like in full MG.
- The multilevel algorithm **converges** for any initial state.

Multilevel MCMC – What can we prove?

Hot off the press! **[Ketelsen, RS, 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, a bit like in full MG.
- The multilevel algorithm **converges** for any initial state.
- **But** "coarse modes" may differ between states on level ℓ and $\ell - 1$:

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

Multilevel MCMC – What can we prove?

Hot off the press! **[Ketelsen, RS, 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, a bit like in full MG.
- The multilevel algorithm **converges** for any initial state.
- **But** "coarse modes" may differ between states on level ℓ and $\ell - 1$:

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

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

Complexity Theorem for Multilevel MCMC

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

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

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

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

M4. $\mathcal{C}_\ell \lesssim M_\ell^{\gamma}$. (cost per sample)

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

$$\mathcal{C}(\widehat{Q}_L^{\mathrm{ML}}) \lesssim \varepsilon^{-2-\max(0,(\gamma-\beta)/\alpha)} \quad \text{(same as before)}$$

(This is again completely **abstract** and applies not only to UQ for subsurface flow!)

For standard MCMC (under the same assumptions) have again Cost $\lesssim \varepsilon^{-2-\gamma/\alpha}$.

Analysis for subsurface flow model problem

- Moments w.r.t. posterior bounded by moments w.r.t. prior:

$$\left| \mathbb{E}_{\pi^\ell,\pi^{\ell-1}}[Y^q] \right| \lesssim \mathbb{E}_{\mathcal{P}_\ell,\mathcal{P}_{\ell-1}}[|Y|^q] \quad \Rightarrow \quad \textbf{M1}$$

- **M2** is hard! Recent result by **[Hairer, Stuart, Vollmer, 2011]**.
- For the variance reduction assumption **M3** we need a further result:

Analysis for subsurface flow model problem

- Moments w.r.t. posterior bounded by moments w.r.t. prior:

$$\left| \mathbb{E}_{\pi^\ell,\pi^{\ell-1}} [Y^q] \right| \lesssim \mathbb{E}_{\mathcal{P}_\ell,\mathcal{P}_{\ell-1}} [|Y|^q] \quad \Rightarrow \quad \mathbf{M1}$$

- **M2** is hard! Recent result by **[Hairer, Stuart, Vollmer, 2011]**.
- For the variance reduction assumption **M3** we need a further result:

Key Lemma

Assume F^M is linear and $F^M(u) \lesssim \|u\|_{L^2(D)}$ for all $u \in H_0^1(D)$. Then

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

and

$$\mathbb{E}_{\mathcal{P}_\ell}\left[(1-\alpha^\ell)^q\right]^{1/q} \lesssim M_\ell^{-1/d+\delta} + J_\ell^{-1/2+\delta} + (\sigma_{\text{fid},\ell}^{-2} - \sigma_{\text{fid},\ell-1}^{-2}),$$

for any $q < \infty$ and $\delta > 0$.

Lemma (Variance Reduction – **M2**)

Let $\mathbf{Z}_\ell^n$ and $\mathbf{Z}_{\ell-1}^n$ be from Algorithm 2. Then, for any $\delta > 0$,

$$\mathbb{V}_{\pi^\ell,\pi^{\ell-1}}\left[Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{Z}_{\ell-1}^n)\right] \lesssim M_\ell^{-1/d+\delta} + J_\ell^{-1/2+\delta} + (\sigma_{\mathrm{fid},\ell}^{-2} - \sigma_{\mathrm{fid},\ell-1}^{-2}).$$

Lemma (Variance Reduction – **M2**)

Let $\mathbf{Z}_\ell^n$ and $\mathbf{Z}_{\ell-1}^n$ be from Algorithm 2. Then, for any $\delta > 0$,

$$\mathbb{V}_{\pi^\ell,\pi^{\ell-1}}\left[Q_\ell(\mathbf{Z}_\ell^n) - Q_{\ell-1}(\mathbf{Z}_{\ell-1}^n)\right] \lesssim M_\ell^{-1/d+\delta} + J_\ell^{-1/2+\delta} + (\sigma_{\mathrm{fid},\ell}^{-2} - \sigma_{\mathrm{fid},\ell-1}^{-2}).$$

Proof. Use $\mathbb{V}[Y] \leq \mathbb{E}[Y^2]$, and result on prior/posterior above and distinguish two cases:

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

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

and result follows from the “key lemma”.

Numerical Example (multilevel MCMC)

$D = (0,1)^2$, $\sigma^2 = 1$, $\lambda = 1$, $Q = k_{\mathrm{eff},1}$

- **Data (artificial)**: Pressure p at 9 random points in domain
- $J_1 = 100$ and $J_0 = 90$
- Averaging over 10000 samples (+ 2500 for "burn-in")

Δx_1^{-1}	σ^2_{fid}	$\mathbb{V}^{\mathrm{post}}(Q_1)$	$\mathbb{V}^{\mathrm{post}}(Q_1 - Q_0)$
64	0.009	0.51	0.0120
128	0.007	0.53	0.0100
256	0.005	0.40	0.0030
512	0.003	0.57	0.0015
		const	$\mathcal{O}(\Delta x_1)$

In absolute terms need 50× less samples for $\Delta x = \frac{1}{64}$ and 400× less for $\Delta x = \frac{1}{512}$

Conclusions

- Stochastic modelling ubiquitous in earth sciences
- **Next generation MC Methods:** Same cost as deterministic solver
- **New**: **Multilevel Markov chain Monte Carlo** (w. **theory!**)
- **Quasi Monte Carlo**: deterministic sampling rules (w. **theory!**)
  $\mathcal{O}(N^{-1})$ possible! Can be combined with ML to gain one extra power!

Conclusions

- Stochastic modelling ubiquitous in earth sciences
- **Next generation MC Methods:** Same cost as deterministic solver
- **New**: **Multilevel Markov chain Monte Carlo** (w. **theory!**)
- **Quasi Monte Carlo**: deterministic sampling rules (w. **theory!**)
  $\mathcal{O}(N^{-1})$ possible! Can be combined with ML to gain one extra power!

Further Work

- Large scale 3D MLMC computations (parallelisation) & multiphase
- More efficient sampling of k (in parallel): GPUs, PDE-based,...
- **New** w. Met Office: MLMC for **atmospheric dispersion** (NAME)
- Other possible applications: Multilevel MCMC for DA in NWP

Thank You!

Thank You!

Preprints available on my website:

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

I would like to thank **EPSRC** and **Lawrence Livermore National Lab** (CA)
for the financial support for this work.