Quasi- and Multilevel Monte Carlo methods for Bayesian elliptic inverse problems

Aretha Teckentrup

Mathematics Institute, University of Warwick

Joint work with:

Rob Scheichl (Bath), Andrew Stuart (Warwick)

2 February, 2015

Enabling Quantification of EQUIP Uncertainty for Inverse Problems

THE UNIVERSITY OF WARWICK

Outline

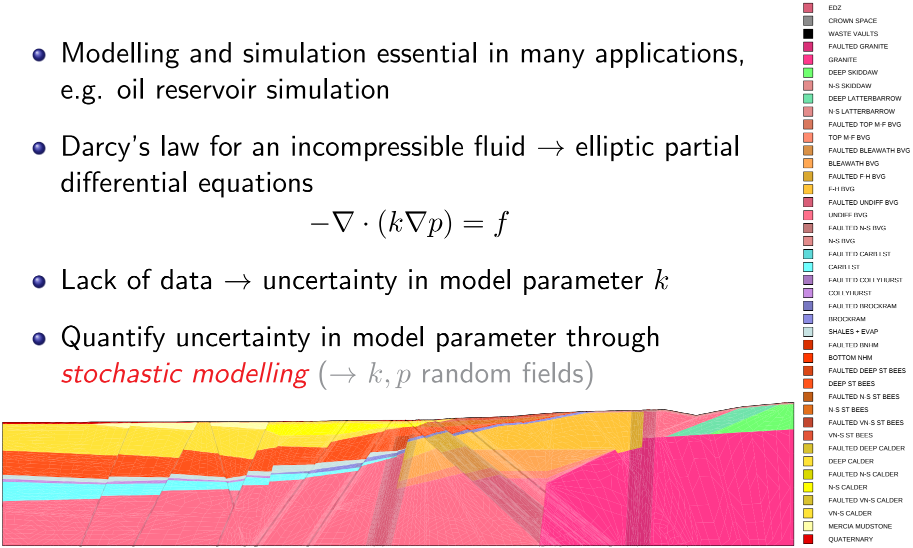
1. Introduction
2. Posterior Expectation as High Dimensional Integration
3. MC, QMC and MLMC
4. Convergence and Complexity
5. Numerical Results
6. Future work

Introduction

- Modelling and simulation essential in many applications, e.g. oil reservoir simulation
- Darcy's law for an incompressible fluid $\rightarrow$ elliptic partial differential equations

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

- Lack of data $\rightarrow$ uncertainty in model parameter k
- Quantify uncertainty in model parameter through *stochastic modelling* ($\rightarrow k, p$ random fields)



Introduction

- Typical simplified model for k is a log–normal random field, $k = \exp[g]$, where g is a scalar, isotropic Gaussian field. E.g.

$$\mathbb{E}[g(x)] = 0, \quad \mathbb{E}[g(x)g(y)] = \sigma^2 \exp[-|x-y|/\lambda].$$

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

- The end goal is usually to estimate the expected value of a quantity of interest (QoI) $\phi(p)$ or $\phi(k, p)$.

Introduction

In addition to presumed log–normal distribution, one usually has available some data $y \in \mathbb{R}^m$ related to the outputs (e.g. pressure data).

Denote by μ_0 the prior log–normal measure on k, and assume

$$y = \mathcal{O}(p) + \eta,$$

where η is a realisation of the Gaussian random variable $\mathcal{N}(0, \sigma_\eta^2 I_m)$.

Bayes' Theorem:

$$\frac{d\mu^y}{d\mu_0}(k) = \frac{1}{Z} \exp[-\frac{|y - \mathcal{O}(p(k))|^2}{2\sigma_\eta^2}] =: \frac{1}{Z} \exp[-\Phi(p(k))]$$

Here,

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

Posterior Expectation as High Dimensional Integration

We are interested in computing $\mathbb{E}_{\mu^y}[\phi(p)]$. Using Bayes' Theorem, we can write this as

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

We have rewritten the posterior expectation as a **ratio of two prior expectations**. We can now approximate

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

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

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

MC, QMC and MLMC [Niederreiter '94], [Graham et al '14]

- The **standard Monte Carlo (MC)** estimator

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

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

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

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

is an equal weighted average of N deterministically chosen samples $\psi(p_h^{(j)})$, with p_h as above.

MC, QMC and MLMC [Giles, '07], [Cliffe et al '11]

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

Linearity of expectation gives us

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

The **multilevel Monte Carlo (MLMC)** estimator

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

is a sum of $L+1$ independent MC estimators. The sequence $\{N_\ell\}$ is decreasing, which means a significant portion of the computational effort is shifted onto the coarse grids.

Convergence and Complexity

- We want to bound the mean square error (MSE)

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

- In the log–normal case, it is *not* sufficient to bound the individual mean square errors $\mathbb{E}[(Q - \widehat{Q})^2]$ and $\mathbb{E}[(Z - \widehat{Z})^2]$.
- We require a bound on $\mathbb{E}[(Z - \widehat{Z})^p]$, for some $p > 2$.
  - **MC**: Follows from results on moments of sample means of i.i.d. random variables (✓)
  - **MLMC**: Follows from results for MC, plus bounds on moments of sum of independent estimators (independent of L) (✓)
  - **QMC**: Requires extension of current QMC theory to non–linear functionals (✓) and higher order moments of the worst-case error (✗)

Convergence and Complexity

Theorem (Scheichl, Stuart, T., in preparation)

Under a log-normal prior, $k = \exp[g]$, we have

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

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

where the convergence rate s is problem dependent. If $k = \exp[g] + c$, for some $c > 0$, then we additionally have

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

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

Convergence and Complexity

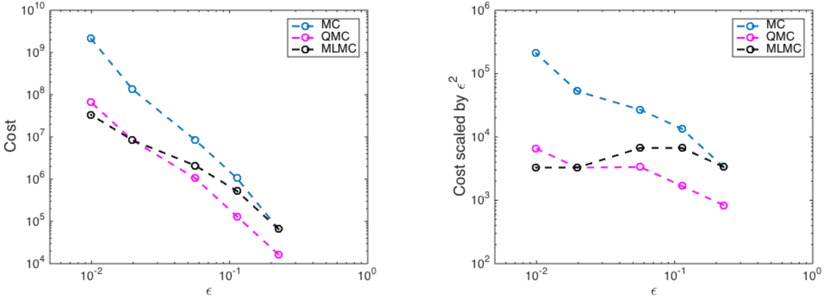
The computational ε-cost is the number of FLOPS required to achieve a MSE of $\mathcal{O}(\varepsilon^2)$.

For the groundwater flow problem in d dimensions, we typically have $s = 2$, and with an optimal linear solver, the computational ε-costs are bounded by:

d	**MLMC**	**QMC**	**MC**
1	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-3})$
2	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-3})$	$\mathcal{O}(\varepsilon^{-4})$
3	$\mathcal{O}(\varepsilon^{-3})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-5})$

Numerical Results

- 2-dimensional flow cell model problem on $(0,1)^2$
- k log-normal random field with exponential covariance function, correlation length $\lambda = 0.3$, variance $\sigma^2 = 1$
- Observed data corresponds to local averages of the pressure p at 9 points, $\sigma_\eta^2 = 0.09$
- QoI is outflow over right boundary



Future work

- As before, we have by Bayes' theorem
$$\frac{d\mu^y}{d\mu_0}(k) = \frac{1}{Z}\exp[-\Phi(p(k))].$$
- Assume we have only a finite number N of forward solver evaluations, and build a surrogate for the log-likelihood, or even for p itself, using a Gaussian process emulator.
- Open questions:
  - Can we quantify the error this introduces in μ^y?
  - Do we recover μ^y as $N \to \infty$?
  - Is there an advantage to using the Gaussian process rather than just the mean?
  - What is the optimal choice of points for the N forward solver evaluations? $\to$ links to optimal design
- $\to$ Shiwei Lan's work on Gaussian process emulators in MCMC algorithms

References

M.B. Giles.
Multilevel Monte Carlo path simulation.
Opererations Research, 256:981–986, 2008.

H. Niederreiter.
Random Number Generation and quasi-Monte Carlo methods.
SIAM, 1994.

R. Scheichl, A.M. Stuart, and A.L. Teckentrup.
Quasi-Monte Carlo and Multilevel Monte Carlo Methods for Computing Posterior Expectations in Elliptic Inverse Problems.
In preparation.

C. Schillings and Ch. Schwab.
Sparse, adaptive Smolyak quadratures for Bayesian inverse problems.
Inverse Problems, 29(6):065011, 2013.