Multilevel Monte Carlo Methods and Applications to Elliptic PDEs with Random Coefficients

K.A. Cliffe*, M.B. Giles†, R. Scheichl‡ and A.L. Teckentrup‡

*School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK, Andrew.Cliffe@nottingham.ac.uk

†The Oxford-Man Institute, University of Oxford, Eagle House, Walton Well Road, Oxford OX2 6ED, UK, Mike.Giles@maths.ox.ac.uk

‡Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK, R.Scheichl@bath.ac.uk, A.L.Teckentrup@bath.ac.uk

May 09, 2011

Abstract

We consider the numerical solution of elliptic partial differential equations with random coefficients. Such problems arise, for example, in uncertainty quantification for groundwater flow. We describe a novel variance reduction technique for the standard Monte Carlo method, called the multilevel Monte Carlo method, and demonstrate numerically its superiority. The asymptotic cost of solving the stochastic problem with the multilevel method is always significantly lower than that of the standard method and grows only proportionally to the cost of solving the deterministic problem in certain circumstances. Numerical calculations demonstrating the effectiveness of the method for one- and two-dimensional model problems arising in groundwater flow are presented.

1 Introduction

There are many situations in which modelling and computer simulation are indispensable tools and where the mathematical models employed have been demonstrated to give adequate representations of reality. However, the parameters appearing in the models often have to be estimated from measurements and are, therefore, subject to uncertainty. This uncertainty propagates through the simulations and quantifying its impact on the results is frequently of great importance.

A good example is provided by the problem of assessing the safety of a potential deep geological repository for radioactive waste. Any radionuclides leaking from such a repository could be transported back to the human environment by groundwater flowing through the rocks beneath the earth's surface. The very long timescales involved mean that modelling and simulation are essential in evaluating repository performance. The study of groundwater flow is well established and there is general scientific consensus that in many situations Darcy's Law can be expected to lead to an accurate description of the flow [7]. The main parameter appearing in Darcy's Law is the hydraulic conductivity, which characterises how easily water can flow through the rock under a given pressure gradient. In practice it is only possible to measure the hydraulic conductivity at a limited number of spatial locations, but it is required at all points of the computational domain for the simulation.

This fact is the primary source of uncertainty in groundwater flow calculations. Understanding and quantifying the impact of this uncertainty on predictions of radionuclide transport is essential for reliable repository safety assessments.

A widely used approach for dealing with uncertainty in groundwater flow is to represent the hydraulic conductivity as a random field [9, 8]. The law of the field has to be estimated from the available data, a significant undertaking in its own right, but the major computational challenge is solving the partial differential equations (PDEs) that govern the pressure field. These are elliptic PDEs with random coefficients. Realistic random field models often need a rather large number of stochastic degrees of freedom (> 100s) for their accurate representation (cf. Section 3.2). Consequently stochastic Galerkin and stochastic collocation approaches, based for example on polynomial chaos expansion [19, 23], are impractical since their cost grows exponentially with the number of stochastic degrees of freedom, and truncating to any feasible number leads to large systematic errors (bias). To the best of our knowledge, there are currently no results with stochastic collocation methods available in the literature that can accurately treat the random field models considered in this paper characterised by short correlation length, high variance and low regularity.

Hence, standard Monte Carlo (MC) simulation is still the method of choice in applications. These MC calculations are, however, very expensive because the individual realisations of the random field have low spatial regularity and significant spatial variation making the problem of solving for the pressure very costly. Furthermore, the notoriously slow rate of convergence of the standard MC method means that many such realisations are required to obtain accurate results. The computational cost of solving elliptic PDEs with random coefficients is therefore a major challenge in uncertainty quantification for groundwater flow studies.

In this paper we address the problem of the large cost of solving elliptic PDEs with random coefficients. Our approach is based on the multilevel Monte Carlo method (MLMC) for infinite-dimensional integration introduced by Giles in connection with stochastic differential equations arising in mathematical finance [13, 12]. Similar ideas have been introduced by Heinrich for finite-dimensional parametric integration and to solve integral equations [16], and by Brandt and his co-workers to accelerate statistical mechanics calculations [2, 3]. In parallel to our work, Barth et al. have recently also provided a theoretical analysis of the multilevel Monte Carlo method in the context of elliptic PDEs with random coefficients [1]. However, they assume smoother coefficient fields than we consider in this paper (see Section 5). For an analysis of the case considered here see the recent paper [5].

In many applications, the quantity of interest is the expected value of a functional of the PDE solution. The MLMC method exploits the linearity of expectation, by expressing the quantity of interest on the finest spatial grid in terms of the same quantity on a relatively coarse grid and "correction" terms. The dramatic reduction in cost associated with the MLMC method over standard MC is due to the fact that most of the uncertainty can be captured on the coarse grids and so the number of realisations needed on the finest grid is greatly reduced. In this paper we explain how these savings in computational cost arise and demonstrate the effectiveness of the MLMC method by a set of numerical results for an elliptic PDE with random coefficients.

The outline of the rest of this paper is as follows. In section 2 we describe the MLMC algorithm in a general context and present a theorem that estimates the cost of the algorithm under certain, problem-dependent, assumptions, which we carefully explain. In section 3 we set out the equations for a model problem arising from groundwater flow, describe our stochastic model, and present the numerical method used for spatial discretisation. We present our numerical results for one and two

dimensional problems in section 4. In section 5 we give our conclusions and make some suggestions for future work.

A final comment is that the main novelty in this paper lies in the use of the highly efficient multilevel Monte Carlo method for a particularly important scientific application. However, this is only one example of how it may be used in connection with stochastic PDEs; a future paper will discuss its use for a stochastic parabolic PDE which arises in a computational finance setting.

2 Monte Carlo Simulations

We will start in this section with a review of the standard Monte Carlo (MC) method and then go on to describe the Multilevel Monte Carlo (MLMC) method. Both methods are not restricted to differential equations with random coefficients, and so we describe them in more abstract terms.

To simplify the notation we will write $a \lesssim b$ for two positive quantities a and b, if a/b is uniformly bounded independent of any parameters, in particular independent of the number of samples N and the number of spatial degrees of freedom M below. Furthermore, we write $a \eqsim b$, if $a \lesssim b$ and $b \lesssim a$.

Let $\mathbf{X}_M$ be a random vector over an infinite dimensional probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $\mathbb{R}^M$. Furthermore let $Q_M = \mathcal{G}(\mathbf{X}_M)$ be some linear or nonlinear functional of $\mathbf{X}_M$. This may be a single component or a norm of $\mathbf{X}_M$, or it may be a more complicated nonlinear functional (e.g. a higher order moment). We assume that as $M \to \infty$ the expected value $\mathbb{E}[Q_M] \to \mathbb{E}[Q]$, for some (inaccessible) random variable $Q : \Omega \to \mathbb{R}$, and that (in mean) the order of convergence is α, i.e.

$$\big|\mathbb{E}[Q_M - Q]\big| \ \lesssim \ M^{-\alpha}.$$

We are interested in estimating $\mathbb{E}(Q)$. Thus, given $M \in \mathbb{N}$ sufficiently large, we compute approximations (or *estimators*) $\widehat{Q}_M$ of $\mathbb{E}(Q_M)$ and quantify the accuracy of our approximations via the root mean square error (RMSE)

$$e(\widehat{Q}_M) := \Big(\mathbb{E}[(\widehat{Q}_M - \mathbb{E}(Q))^2]\Big)^{1/2}.$$

In our PDE application, choosing M sufficiently large corresponds to choosing a fine enough spatial approximation. The random variable Q will in this case typically be a functional of the solution, and Q_M will be the same functional of the discretised solution.

The computational ε-cost $\mathcal{C}_\varepsilon(\widehat{Q}_M)$ is then quantified by the number of floating point operations that are needed to achieve a RMSE of $e(\widehat{Q}_M) < \varepsilon$.

2.1 Standard Monte Carlo Simulation

The standard Monte Carlo estimator for $\mathbb{E}(Q_M)$ is

$$\widehat{Q}^{\mathrm{MC}}_{M,N} := \frac{1}{N}\sum_{i=1}^{N} Q_M^{(i)}, \tag{2.1}$$

where $Q_M^{(i)}$ is the ith sample of Q_M and N independent samples are computed in total. We assume that the cost to compute one sample $Q_M^{(i)}$ of Q_M is

$$\mathcal{C}(Q_M^{(i)}) \ \lesssim \ M^\gamma, \qquad \text{for some} \ \ \gamma > 0.$$

There are two sources of error in the estimator (2.1): the approximation of Q by Q_M, which is related to the spatial discretisation in the case of our PDE application; and the sampling error due to replacing the expected value by a finite sample average. The contribution of both of these errors becomes clear when we expand the mean square error (MSE):

$$\begin{aligned} e(\widehat{Q}^{\mathrm{MC}}_{M,N})^2 &= \mathbb{E}\left[\left(\widehat{Q}^{\mathrm{MC}}_{M,N} - \mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{M,N}] \;+\; \mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{M,N}] - \mathbb{E}[Q]\right)^2\right] \\ &= \mathbb{E}\left[(\widehat{Q}^{\mathrm{MC}}_{M,N} - \mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{M,N}])^2\right] + \left(\mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{M,N}] - \mathbb{E}[Q]\right)^2 \\ &= \mathbb{V}[\widehat{Q}^{\mathrm{MC}}_{M,N}] + \left(\mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{M,N}] - \mathbb{E}[Q]\right)^2. \end{aligned} \tag{2.2}$$

Since $\mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{M,N}] = \mathbb{E}[Q_M]$ and $\mathbb{V}[\widehat{Q}^{\mathrm{MC}}_{M,N}] = N^{-1}\,\mathbb{V}[Q_M]$, we get

$$e(\widehat{Q}^{\mathrm{MC}}_{M,N})^2 = N^{-1}\mathbb{V}[Q_M] + \left(\mathbb{E}[Q_M - Q]\right)^2, \tag{2.3}$$

and so the first term in the MSE is the variance of the MC estimator, which represents the sampling error and decays inversely with the number of samples. The second term is the square of the error in mean between Q_M and Q.

Hence, a sufficient condition to achieve a RMSE of ε with this estimator is that both of the terms are less than $\varepsilon^2/2$. Under the assumption that $\mathbb{V}[Q_M]$ is approximately constant, independent of M, this can be achieved by choosing $N \gtrsim \varepsilon^{-2}$ and $M \gtrsim \varepsilon^{-1/\alpha}$, where the convergence rate α is as defined previously and problem dependent. In other words, we need to take a large enough number of samples N, as well as a large enough value for M, so that $\widehat{Q}^{\mathrm{MC}}_{M,N}$ is a sufficiently accurate approximation of our quantity of interest $\mathbb{E}[Q]$.

Since the cost to compute one sample of Q_M was assumed to satisfy $\mathcal{C}(Q^{(i)}_M) \lesssim M^\gamma$, we have $\mathcal{C}(\widehat{Q}^{\mathrm{MC}}_{M,N}) \lesssim NM^\gamma$ and so the total computational cost of achieving a RMSE of $O(\varepsilon)$ is

$$\mathcal{C}_\varepsilon(\widehat{Q}^{\mathrm{MC}}_{M,N}) \lesssim \varepsilon^{-2-\gamma/\alpha}.$$

2.2 Multilevel Monte Carlo Simulation

The main idea of multilevel Monte Carlo (MLMC) simulation is very simple. We sample not just from one approximation Q_M of Q, but from several. Let us recall the main ideas and the main theorem from [13].

Let $\{M_\ell : \ell = 0, \ldots, L\}$ be an increasing sequence in $\mathbb{N}$ called *levels*, i.e. $M_0 < M_1 < \ldots < M_L =: M$, and assume for simplicity that there exists an $s \in \mathbb{N}\backslash\{1\}$ such that

$$M_\ell = s\, M_{\ell-1}\,, \qquad \text{for all } \ell = 1, \ldots, L. \tag{2.4}$$

As for multigrid methods applied to discretised (deterministic) PDEs, the key is to avoid estimating $\mathbb{E}[Q_{M_\ell}]$ directly on level ℓ, but instead to estimate the correction with respect to the next lower level, i.e. $\mathbb{E}[Y_\ell]$ where $Y_\ell := Q_{M_\ell} - Q_{M_{\ell-1}}$. Linearity of the expectation operator then implies that

$$\mathbb{E}[Q_M] = \mathbb{E}[Q_{M_0}] + \sum_{\ell=1}^{L} \mathbb{E}[Q_{M_\ell} - Q_{M_{\ell-1}}] = \sum_{\ell=0}^{L} \mathbb{E}[Y_\ell], \tag{2.5}$$

where for simplicity we have set $Y_0 := Q_{M_0}$.

Hence, the expectation on the finest level is equal to the expectation on the coarsest level, plus a sum of corrections adding the difference in expectation between simulations on consecutive levels. The multilevel idea is now to independently estimate each of these expectations such that the overall variance is minimised for a fixed computational cost.

Let $\widehat{Y}_\ell$ be an unbiased estimator for $\mathbb{E}[Y_\ell]$, e.g. the standard MC estimator

$$\widehat{Y}^{\mathrm{MC}}_{\ell,N_\ell} \;:=\; \frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(Q^{(i)}_{M_\ell} - Q^{(i)}_{M_{\ell-1}}\right) \tag{2.6}$$

with N_ℓ samples. Then the *multilevel* estimator is simply defined as

$$\widehat{Q}^{\mathrm{ML}}_M \;:=\; \sum_{\ell=0}^{L}\widehat{Y}_\ell. \tag{2.7}$$

If the individual terms are estimated using standard MC, i.e. (2.6) with N_ℓ samples on level ℓ, this is the *multilevel Monte Carlo* (MLMC) estimator and we denote it by $\widehat{Q}^{\mathrm{MLMC}}_{M,\{N_\ell\}}$. It is important to note that the quantity $Q^{(i)}_{M_\ell} - Q^{(i)}_{M_{\ell-1}}$ in (2.6) comes from using the same random sample $\omega^{(i)} \in \Omega$ on both levels M_ℓ and $M_{\ell-1}$.

For the rest of this paper, for simplicity, we will always use standard MC to estimate the terms on the different levels. Note however, that this could be replaced by any other unbiased estimator, e.g. randomised quasi-Monte Carlo (cf. [14, 15]).

Since all the expectations $\mathbb{E}[Y_\ell]$ are estimated independently, the variance of the MLMC estimator is $\mathbb{V}[\widehat{Q}^{ML}_M] = \sum_{\ell=0}^{L} N_\ell^{-1}\,\mathbb{V}[Y_\ell]$, and expanding as in (2.2-2.3) in the previous section leads again to the following form for the MSE:

$$e(\widehat{Q}^{\mathrm{ML}}_M)^2 \;:=\; \mathbb{E}\Big[\big(\widehat{Q}^{\mathrm{ML}}_M - \mathbb{E}[Q]\big)^2\Big] \;=\; \sum_{\ell=0}^{L} N_\ell^{-1}\,\mathbb{V}[Y_\ell] \;+\; \Big(\mathbb{E}[Q_M - Q]\Big)^2. \tag{2.8}$$

As in the standard MC case before, we see that the MSE consists of two terms, the variance of the estimator and the approximation error. Note that the second term is exactly the same as before in (2.2), and so it is sufficient to choose $M = M_L \gtrsim \varepsilon^{-1/\alpha}$ again. To then achieve an overall RMSE of ε, the first term in (2.8) has to be less than $\varepsilon^2/2$ as well. We claim that this is cheaper to achieve in MLMC for two reasons:

- If Q_M converges to Q in mean square, then $\mathbb{V}[Y_\ell] = \mathbb{V}[Q_{M_\ell} - Q_{M_{\ell-1}}] \to 0$ as $\ell \to \infty$, and so fewer samples are required on finer levels to estimate $\mathbb{E}[Y_\ell]$;
- The coarsest level $\ell = 0$ can be kept fixed for all ε, and so the cost per sample on level $\ell = 0$ does not grow as $\varepsilon \to 0$.

In practical applications, M_0 must be chosen sufficiently large to provide a minimal level of resolution of the problem. In our PDE application, this cut-off point is related to the spatial regularity of the PDE solution, which in turn depends on the regularity of the covariance function of the conductivity field and on the correlation length. We will return to this point in Section 4.1.

The computational cost of the multilevel Monte Carlo estimator is

$$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \;=\; \sum_{\ell=0}^{L} N_\ell \, \mathcal{C}_\ell.$$

where $\mathcal{C}_\ell := \mathcal{C}(Y_\ell^{(i)})$ represents the cost of a single sample of Y_ℓ. Treating the N_ℓ as continuous variables, the variance of the MLMC estimator is minimised for a fixed computational cost by choosing

$$N_\ell \;\eqsim\; \sqrt{\mathbb{V}[Y_\ell]/\mathcal{C}_\ell}\,, \tag{2.9}$$

with the constant of proportionality chosen so that the overall variance is $\varepsilon^2/2$. The total cost on level ℓ is proportional to $\sqrt{\mathbb{V}[Y_\ell]\,\mathcal{C}_\ell}$ and hence

$$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \;\lesssim\; \sum_{\ell=0}^{L} \sqrt{\mathbb{V}[Y_\ell]\,\mathcal{C}_\ell}.$$

If the variance $\mathbb{V}[Y_\ell]$ decays faster with ℓ than $\mathcal{C}_\ell$ increases, the dominant term will be on level 0. Since $N_0 \eqsim \varepsilon^{-2}$, the cost savings compared to standard MC will in this case be approximately $\mathcal{C}_0/\mathcal{C}_L \eqsim (M_0/M_L)^\gamma \eqsim \varepsilon^{\gamma/\alpha}$, reflecting the ratio of the costs of samples on level 0 compared to samples on level L.

If the variance $\mathbb{V}[Y_\ell]$ decays slower than the cost $\mathcal{C}_\ell$ increases, the dominant term will be on the finest level L, and the cost savings compared to standard MC will be approximately $\mathbb{V}[Y_L]/\mathbb{V}[Y_0]$ which will be small. Hence, in both cases we have a significant gain.

This outline analysis is made more precise in the following theorem:

Theorem 2.1 *Let $\widehat{Y}_\ell := \widehat{Y}_{\ell,N_\ell}^{\mathrm{MC}}$ and suppose that there are positive constants $\alpha, \beta, \gamma > 0$ such that $\alpha \geq \frac{1}{2}\min(\beta,\gamma)$ and*

i) $\big|\mathbb{E}[Q_{M_\ell} - Q]\big| \;\lesssim\; M_\ell^{-\alpha}$

ii) $\mathbb{V}[Y_\ell] \;\lesssim\; M_\ell^{-\beta}$

iii) $\mathcal{C}_\ell \;\lesssim\; M_\ell^{\gamma}$,

Then, for any $\varepsilon < e^{-1}$, there exist a value L (and corresponding $M \equiv M_L$)and a sequence $\{N_\ell\}_{\ell=0}^L$ such that

$$e(\widehat{Q}_M^{\mathrm{ML}})^2 \;:=\; \mathbb{E}\left[\left(\widehat{Q}_M^{\mathrm{ML}} - \mathbb{E}[Q]\right)^2\right] < \varepsilon^2,$$

and

$$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \;\lesssim\; \begin{cases} \varepsilon^{-2}, & \text{if } \beta > \gamma, \\ \varepsilon^{-2}(\log\varepsilon)^2, & \text{if } \beta = \gamma, \\ \varepsilon^{-2-(\gamma-\beta)/\alpha}, & \text{if } \beta < \gamma. \end{cases}$$

Proof The proof, which is given in Appendix A, is a slight generalisation of the proof in [13]. □

The (optimal) values of L and $\{N_\ell\}_{\ell=0}^L$ can be computed "on the fly" from the sample averages and the (unbiased) sample variances of Y_ℓ. To do this we need to assume further that there exists an $M' \in \mathbb{N}$ such that the decay in $|\mathbb{E}[Q_M - Q]|$ is actually monotonic for $M \geq M'$ and satisfies

$$\big|\mathbb{E}[Q_M - Q]\big| \;\eqsim\; M^{-\alpha}.$$

This ensures (via the triangle inequality) that $|\mathbb{E}[Y_L]| \eqsim M^{-\alpha}$ (since $s > 1$ in (2.4)), and thus $|\widehat{Y}_L| \eqsim M^{-\alpha}$ for N_L sufficiently large, providing us with a computable error estimator to determine whether M is sufficiently large or whether the number of levels L needs to be increased. It can in fact even be used to further improve the MLMC estimate by eliminating the leading order bias term via Richardson extrapolation (see [13, §4.2] for details).

Putting these ideas together, the MLMC algorithm can be implemented in practice as follows:

1. Start with L=0.
2. Estimate $\mathbb{V}[Y_L]$ by the sample variance of an initial number of samples.
3. Calculate the optimal N_ℓ, $\ell = 0, 1, \ldots, L$ using (2.9).
4. Evaluate extra samples at each level as needed for the new N_ℓ.
5. If $L \geq 1$, test for convergence using $\widehat{Y}_L \eqsim M^{-\alpha}$.
6. If not converged, set $L = L + 1$ and go back to 2.

Note that in the above algorithm, step 3 aims to make the variance of the MLMC estimator less than $\frac{1}{2}\varepsilon^2$, while step 5 tries to ensure that the remaining bias is less than $\frac{1}{\sqrt{2}}\varepsilon$.

3 Application to PDEs – A model problem

In this section we will apply Multilevel Monte Carlo to elliptic PDEs with random coefficients arising in subsurface flow.

Probabilistic uncertainty quantification in subsurface flow is of interest in a number of situations, as for example in risk analysis for radioactive waste disposal or in oil reservoir simulation. The classical equations governing (steady state) single phase subsurface flow consist of Darcy's law coupled with an incompressibility condition (see e.g. [7, 6]):

$$\mathbf{q} + k\nabla p = \mathbf{g} \quad \text{and} \quad \nabla \cdot \mathbf{q} = 0, \quad \text{in} \;\; D \subset \mathbb{R}^d, \; d = 1, 2, 3, \tag{3.1}$$

subject to suitable boundary conditions. In physical terms, p denotes the pressure (or more precisely the pressure head) of the fluid, k is the hydraulic conductivity tensor, $\mathbf{q}$ is the filtration velocity (or Darcy flux) and $\mathbf{g}$ are the source terms.

3.1 Model problem

A typical approach to quantify uncertainty in p and $\mathbf{q}$ is to model the hydraulic conductivity as a random field $k = k(\mathbf{x}, \omega)$ on $D \times \Omega$ with a certain mean and covariance structure that has to be

inferred from the data. This means that (3.1) becomes a system of PDEs with random coefficients, which can be written in second order form as

$$-\nabla \cdot (k(\mathbf{x},\omega)\nabla p(\mathbf{x},\omega)) = f(\mathbf{x}), \qquad \text{in} \quad D, \tag{3.2}$$

with $f := -\nabla \cdot \mathbf{g}$. The solution p itself will also be a random field on $D \times \Omega$. For simplicity we assume that the boundary conditions and the sources $\mathbf{g}$ are known (and thus deterministic), and restrict ourselves to the case $D = (0,1)^d$.

In this general form solving (3.2) is extremely challenging computationally and so in practice it is common to use relatively simple models for $k(\mathbf{x},\omega)$ that are as faithful as possible to the measured data. One model that has been studied extensively is a lognormal distribution for $k(\mathbf{x},\omega)$, i.e. replacing the conductivity tensor by a scalar valued field whose log is Gaussian. It guarantees that $k > 0$ almost surely (a.s.) in Ω and it allows the conductivity to vary over many orders of magnitude, which is typical in subsurface flow.

When modelling a whole aquifer, a whole oil reservoir, or a sufficiently large region around a potential radioactive waste repository, the correlation length scale for k is typically significantly smaller than the size of the computational region. However, the correlation is typically large enough to fall outside the domain of stochastic homogenisation techniques. In addition, typical sedimentation processes lead to fairly irregular structures and pore networks, and faithful models should therefore also only assume limited spatial regularity of k. A covariance function that has been proposed in the application literature (cf. [17]) is the following exponential two-point covariance function

$$C(\mathbf{x},\mathbf{y}) \;:=\; \sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|_p}{\lambda}\right), \qquad \mathbf{x},\mathbf{y} \in D, \tag{3.3}$$

where $\|\cdot\|_p$ denotes the ℓ_p-norm in $\mathbb{R}^d$ and typically $p = 1$ or 2; throughout this paper we use $p = 1$ for simplicity. The parameters σ^2 and λ denote the *variance* and the *correlation length*, respectively, and in subsurface flow applications typically only $\sigma^2 \geq 1$ and $\lambda \leq \operatorname{diam} D = 1$ will be of interest. This choice of covariance function implies that k is *homogeneous* and it follows from Kolmogorov's theorem [20] that $k(\cdot,\omega) \in C^{0,\eta}(D)$ a.s. with $\eta < 1/2$.

In order to apply the proposed Multilevel Monte Carlo Method to this application, we need to be able to do two things:

- sample from the input random field $k(\mathbf{x},\omega)$, and
- for a given sample, i.e. for fixed ω, perform a spatial discretisation of the PDE in (3.2) on two consecutive grids and solve it.

3.2 Sampling from the input random field

Several techniques exist to produce samples of k, including circulant embedding as studied in [15] or the Karhunen-Loève (KL) expansion [11]. We only describe and apply the KL-expansion here. Let $Z(\mathbf{x},\omega) := \log k(\mathbf{x},\omega)$. We can then expand Z in terms of a countable set of uncorrelated, zero mean random variables $\{\xi_n\}_{n\in\mathbb{N}}$ such that

$$Z(\mathbf{x},\omega) = \mathbb{E}\left[Z(\mathbf{x},\cdot)\right] + \sum_{n=1}^{\infty} \sqrt{\theta_n}\,\xi_n(\omega)\,b_n(\mathbf{x}), \tag{3.4}$$

where $\{\theta_n\}_{n\in\mathbb{N}}$ are the eigenvalues and $\{b_n\}_{n\in\mathbb{N}}$ the normalised eigenfunctions of the covariance operator with kernel function $C(\mathbf{x},\mathbf{y})$ defined in (3.3). For more details on the derivation and for properties of the KL–expansion, see e.g. [11]. However, an important point to note is that for Gaussian random fields Z the random variables $\{\xi_n\}_{n\in\mathbb{N}}$ are a set of independent standard Gaussian variables.

In the case of the 1-norm $\|\cdot\|_1$, i.e. $p = 1$, in (3.3), analytic expressions for the eigenpairs of the covariance operator are available. Following e.g. [11], we get for $d = 1$, $D = (0, 1)$ and $\sigma^2 = 1$ in (3.3):

$$\begin{aligned} \theta_n^{1D} &= \frac{2\lambda}{\lambda^2 w_n^2 + 1}, \qquad n \in \mathbb{N}, \\ b_n(x)^{1D} &= A_n\big(\sin(w_n x) + \lambda w_n \cos(w_n x)\big), \qquad n \in \mathbb{N}, \end{aligned} \tag{3.5}$$

where $\{w_n\}_{n\in\mathbb{N}}$ are the (real) solutions of the transcendental equation

$$\tan(w) = \frac{2\lambda w}{\lambda^2 w^2 - 1},$$

and the constant A_n is chosen so that $\|b_n\|_{L_2(0,1)} = 1$. For $d = 2$, and $D = (0, 1)^2$ the eigenpairs can then be expressed as

$$\theta_n^{2D} = \theta_{i_n}^{1D}\theta_{j_n}^{1D} \quad \text{and} \quad b_n^{2D}(\mathbf{x}) = b_{i_n}^{1D}(x_1)b_{j_n}^{1D}(x_2), \quad \text{for some } i_n, j_n \in \mathbb{N}.$$

For σ^2 different to 1, the eigenfunctions are the same as above and the eigenvalues (both in 1D and 2D) are simply multiplied by σ^2. In the (practically more realistic) case $p = 2$ and for many other covariance functions, efficient ways to compute KL-expansions using multipole or matrix compression techniques can be found in [21, 18, 10].

In practice we have to truncate the expansion (3.4) after a finite number m_{KL} of terms. Let $Z_{m_{\mathrm{KL}}}$ denote the KL-expansion of Z truncated after m_{KL} terms. Since $\{\xi_n\}_{n\in\mathbb{N}}$ is a sequence of i.i.d. standard Gaussian random variables and $\|b_n\|_{L_2(D)} = 1$, the accuracy of the truncated KL-expansion depends directly on the decay of the eigenvalues θ_n. With the kernel function $C(\mathbf{x},\mathbf{y})$ in (3.3) the covariance operator is self-adjoint, non-negative and compact, which implies that it has a countable sequence of real, non-negative eigenvalues that tend to 0. Furthermore it is of trace class, i.e. the sum of all eigenvalues is finite. In order to decide how many modes to include we make the following observations. (They follow easily from (3.5) in the case $p = 1$ for our model problem; more details on the general case can be found in [21].)

- The eigenvalues θ_n decay quadratically with respect to n, e.g. $\theta_n \lesssim n^{-2}$.
- If $\lambda < \mathrm{diam}(D)$, then there is a pre-asymptotic phase where the KL-eigenvalues do not decay significantly. This is clearly visible in the left plot in Figure 1.
- Moreover
$$\sum_{n=1}^{\infty} \theta_n = \sigma^2 \,\mathrm{meas}(D). \tag{3.6}$$
where $\mathrm{meas}(D) := \int_D \mathrm{d}x$ (see e.g. [10]).

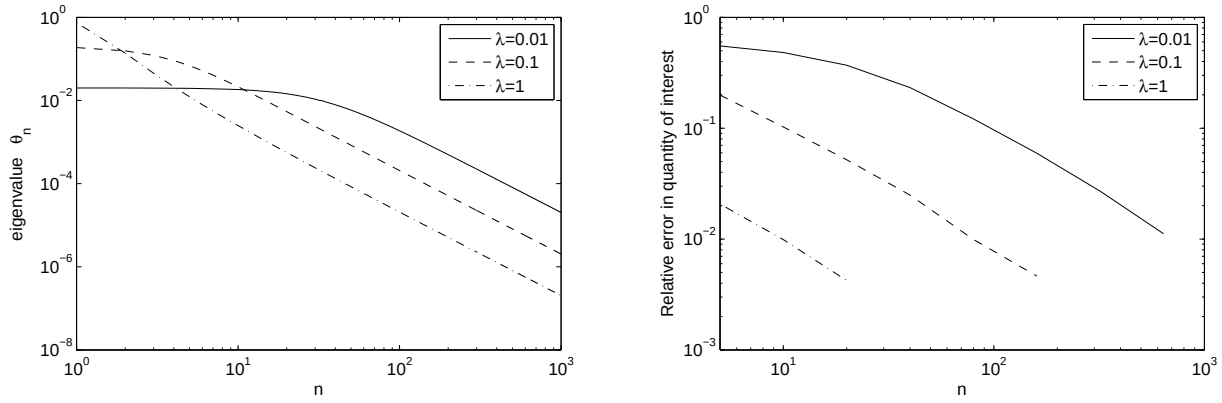

Figure 1: Left: Plot of the KL eigenvalues in decreasing order as a function of their index, for different choices of the correlation length λ and for $d = 1$. Right: Corresponding relative error (or bias) $\big|\mathbb{E}[k_{\text{eff}}(Z)] - \mathbb{E}[k_{\text{eff}}(Z_{m_{\text{KL}}})]\big|/\mathbb{E}[k_{\text{eff}}(Z)]$ in a typical quantity of interest (here the effective hydraulic conductivity k_{eff} defined in Section 4) as a function of the number of KL modes included.

The identity (3.6) can be used in practice to ensure that a sufficient fraction of the variance is captured by the first m_{KL} terms. This is what we did in our numerical experiments later. In order to get an idea about the error resulting from truncating the KL–expansion, note first that

$$\mathbb{E}\left[\|Z - Z_{m_{\text{KL}}}\|^2_{L_2(D)}\right] = \mathbb{E}\left[\Big\| \sum_{n=m_{\text{KL}}+1}^{\infty} \sqrt{\theta_n}\, \xi_n(\omega)\, b_n \Big\|^2_{L_2(D)}\right] = \sum_{n=m_{\text{KL}}+1}^{\infty} \theta_n$$

Since $\theta_n = \mathcal{O}(n^{-2})$, this implies that the RMSE of $Z_{m_{\text{KL}}}$ in the $L_2(D)$–norm is $\mathcal{O}(m_{\text{KL}}^{-1/2})$. It is known that the error in the expected value of functionals of the solution resulting from a truncation of the KL-expansion usually decays more rapidly (see e.g. [4]). Indeed, our numerical computations show that the expected value of a typical quantity of interest, i.e. the effective hydraulic conductivity k_{eff} defined in (4.1) in Section 4, decays like $\mathcal{O}(m_{\text{KL}}^{-1})$ in 1D (cf. Figure 1, right, in which the reference value $\mathbb{E}[k_{\text{eff}}(Z)]$ is evaluated using 5000 KL modes). However, in absolute terms, even in 1D and especially for short correlation lengths λ, a very large number of KL-modes needs to be included to achieve even just a relative error (or bias) of 10^{-2}.

3.3 Spatial discretisation

The particular choice of spatial discretisation scheme is not essential to the multilevel MC approach. However, many quantities of interest in subsurface flow depend on an accurate and mass-conservative representation of the flux $\mathbf{q}$, and so in this context finite volume (FV) or mixed finite elements (FEs) are usually preferred over standard Lagrange FEs. For a short and simple description and a fast implementation of mixed FEs applied to our model problem see eg. [6] or [15]. In this paper we will describe and use a standard cell-centred FV method instead.

Let us briefly describe our discretisation for $d = 2$. The one-dimensional case is analogous. We start by subdividing $[0,1]^2$ uniformly into a mesh of $m \times m$ square cells and denote by $D_{i,j}$ the cell $(\frac{i-1}{m}, \frac{i}{m}) \times (\frac{j-1}{m}, \frac{j}{m})$ with $i, j = 1, \dots m$, and by $\mathbf{x}_{i,j}$ its centre. To discretise (3.2) we integrate (3.2)

over each cell to obtain a set of m^2 algebraic equations

$$\int_{D_{i,j}} -\nabla \cdot (k\nabla p) = \int_{D_{i,j}} f\,, \qquad \text{for all } 1 \le i,j \le m. \tag{3.7}$$

Then, using the Divergence Theorem, we transform the left hand side integral into a boundary integral $\int_{\partial D_{i,j}} -k\nabla p \cdot \mathbf{n}$ and approximate all the resulting integrals in (3.7) by quadrature.

Let $k_{i,j}$ and $f_{i,j}$ be the values of k and f at $\mathbf{x}_{i,j}$, respectively, and let $p_{i,j}$ denote our approximation to p at $\mathbf{x}_{i,j}$. Then the right hand side in (3.7) can be approximated by the midpoint rule as $f_{i,j}/m^2$. To approximate the left hand side we treat each edge of $\partial D_{i,j}$ separately. The contribution from the edge between $D_{i,j}$ and $D_{i+1,j}$ can be approximated again by the midpoint rule, and as it is customary in subsurface flow applications, we use the harmonic average $\overline{k}_{i+\frac{1}{2},j}$ of $k_{i,j}$ and $k_{i+1,j}$ to approximate k on the edge. To approximate $\nabla p \cdot \mathbf{n}$ on the edge we use the central finite difference $(p_{i+1,j} - p_{i,j})/|\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}|$. The contributions from the other edges are approximated similarly, leading to the following final form of the (i,j)th equation:

$$-\overline{k}_{i,j-\frac{1}{2}} p_{i,j-1} - \overline{k}_{i-\frac{1}{2},j} p_{i-1,j} + \Sigma_{i,j} p_{i,j} - \overline{k}_{i+\frac{1}{2},j} p_{i+1,j} - \overline{k}_{i,j+\frac{1}{2}} p_{i,j+1} = f_{i,j}/m^2 \tag{3.8}$$

where $\Sigma_{i,j} = \overline{k}_{i,j-\frac{1}{2}} + \overline{k}_{i-\frac{1}{2},j} + \overline{k}_{i+\frac{1}{2},j} + \overline{k}_{i,j+\frac{1}{2}}$.

A Neumann boundary condition, i.e. a prescribed flux $-k\nabla p \cdot \mathbf{n} = g_N$, on any part of the outer boundary of $(0,1)^2$ is straightforward to incorporate. We simply replace the respective flux term on the left hand side of (3.8) by $g_N(\mathbf{x}_{m+\frac{1}{2},j})/m$ (again obtained via the midpoint rule). To enforce a Dirichlet boundary condition, i.e. a prescribed pressure $p = g_D$, we simply replace the harmonic average on the respective edge by $k_{i,j}$ and the central difference by a one-sided difference.

The resulting linear system takes the standard five-point stencil form. It is sparse and highly ill-conditioned, but it can be solved efficiently and robustly either with algebraic multigrid methods [22] or a sparse direct solver. The solution is an $M = m^2$ dimensional vector $\mathbf{X}_M$ containing approximations $p_{i,j}$ of the pressure p at the points $\mathbf{x}_{i,j}$, $i,j = 1,\dots m$. Typical quantities of interest $Q_M = \mathcal{G}(\mathbf{X}_M)$ to derive from this solution vector will be discussed in the next section.

For the MLMC method we need a sequence of such spatial approximations to construct our levels. We choose a coarsest mesh size m_0 and set $m_\ell = 2^\ell m_0$, for all $\ell \in \mathbb{N}$. Then the mesh size on level ℓ is $h_\ell = m_\ell^{-1}$ and the length of the random vector $\mathbf{X}_{M_\ell}$ is $M_\ell = m_\ell^2$.

4 Numerical Results

In this section we examine the performance of the MLMC method in computing the expected values of some quantities of interest for our model problem in 1D and 2D. In particular, we consider (3.2) on $D = (0,1)^2$ with $f \equiv 0$ and subject to the boundary conditions

$$p|_{x_1=0} = 1, \qquad p|_{x_1=1} = 0, \qquad \frac{\partial p}{\partial \mathbf{n}}\Big|_{x_2=0} = 0, \qquad \frac{\partial p}{\partial \mathbf{n}}\Big|_{x_2=1} = 0,$$

and the corresponding ODE in 1D with $p(0) = 1$ and $p(1) = 0$. To discretise our model problems in space we use the finite volume method described in the previous section.

The statistics of several functionals of the solution are commonly of interest, e.g. the variance of the pressure or of the flow rate at a certain point in the domain, or the average travel time of a

particle convected in the fluid. Here we will mainly focus on the expected value of the cumulative outflow from the region D on the boundary $x_1 = 1$. This is related to the effective (horizontal) conductivity of the region D (see e.g. [15]) which is defined as

$$k_{\text{eff}} := -\int_0^1 k \frac{\partial p}{\partial x_1}\Big|_{x_1=1} \mathrm{d}x_2. \tag{4.1}$$

In addition, we will also look at the horizontal flux $-k\frac{\partial p}{\partial x_1}$ at the centre of the domain.

To quantify the cost of the algorithms in the figures below, we assume that the number of operations to compute one sample on level ℓ is $\mathcal{C}_\ell = C^* M_\ell^\gamma$ for some fixed constant C^* that may depend on λ and σ^2 but is independent of ℓ. In the case of 1D we have $\gamma = 1$. In 2D, for an optimal iterative linear solver such as algebraic multigrid we also have $\gamma \approx 1$. A sparse direct solver on the other hand, such as the one provided by `Matlab` through the backslash operation, will usually be slightly suboptimal in 2D, but will have at worst $\gamma = 1.5$. In the results presented below, unless otherwise stated, we always present the standardised costs, scaled by $1/C^*$, and assume $\gamma = 1$.

The relative performance of the individual methods is very similar if actual CPU times are used instead (see Figure 7). However, since our code is not optimised, we did not want to use these to assess the performance directly.

4.1 Results in 1D

Let us start by solving the 1D version of (3.2) on $D = (0, 1)$ with boundary conditions $p(0) = 1$ and $p(1) = 0$, and choose as the quantity of interest $Q = -k\frac{\partial p}{\partial x}|_{x=1}$. We will first numerically confirm the assumptions in Theorem 2.1, and estimate values of the parameters α and β . We also confirm the predicted bound on the cost of the MLMC estimator.

Figure 2 shows results for the case $\lambda = 0.3$, $\sigma^2 = 1$, $m_{\text{KL}} = 800$ and $m_0 = 16$. The top left plot shows the behaviour of the variance of Q_ℓ and of $Y_\ell = Q_\ell - Q_{\ell-1}$ for each level ℓ. The slope of the line for $\mathbb{V}[Y_\ell]$ is approximately equal to -2, indicating that $\mathbb{V}[Y_\ell] \lesssim h_\ell^2 \eqsim M_\ell^{-2}$, or $\beta \approx 2$. We also see that $\mathbb{V}[Q_\ell]$ is approximately constant on all levels shown, numerically verifying the assumption made in Section 2.1 for large enough values of M. The top right plot shows the expected value of Q_ℓ and of $Y_\ell = Q_\ell - Q_{\ell-1}$. The slope of the line for $\mathbb{E}[Q_\ell - Q_{\ell-1}]$ is roughly equal to -1.75, indicating that $\mathbb{E}[Q - Q_\ell] \lesssim h_\ell^{1.75} \eqsim M_\ell^{-1.75}$, or $\alpha \approx 1.75$.

The bottom two plots are related to the implementation of the MLMC algorithm and to its cost. The left plot shows the number of samples N_ℓ used on each level, and the right plot shows a comparison of the cost of standard MC with the cost of MLMC. Note that the MLMC algorithm does not only result in large savings in the computational cost, but that the cost of the MLMC estimator also grows more slowly than the cost of the standard MC estimator as $\varepsilon \to 0$.

Before moving on to 2D, we briefly return to the point made in Section 2.2 about the choice of the mesh size $h_0 = m_0^{-1}$ on the coarsest level. As we see in the top left plot of Figure 2, for large values of h_ℓ, the variances of Q_ℓ and Y_ℓ are close. Increasing h_ℓ even further, the two graphs will eventually cross, and $\mathbb{V}[Y_\ell]$ will be larger than $\mathbb{V}[Q_\ell]$. In this situation, the contributions to the cost of the MLMC method from level ℓ will actually be bigger than those using standard MC, rendering any further coarsening useless. It turns out that the two graphs cross when $h_\ell \approx \lambda$. It is in fact also at this same point $h_\ell \approx \lambda$ where $\mathbb{V}[Q_\ell]$ ceases to be constant. Thus, the optimal choice for the coarsest level is such that h_0 is slightly smaller than λ.

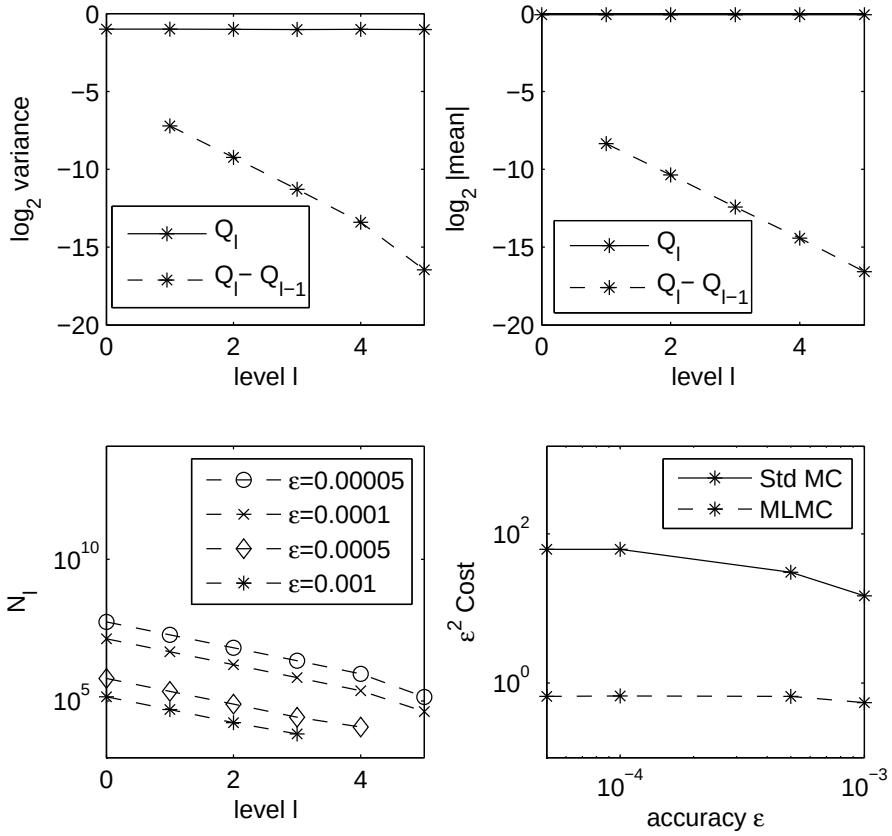

Figure 2: Performance plots for $\lambda = 0.3,\ \sigma^2 = 1,\ m_{\mathrm{KL}} = 800$ and $m_0 = 16$ in 1D. The quantity of interest is the outflow $-k\frac{\partial p}{\partial x}$ at $x = 1$.

4.2 Results in 2D

As in 1D we choose $\lambda = 0.3$ and $\sigma^2 = 1$, but we set $m_0 = 8$ and include $m_{\mathrm{KL}} = 1400$ KL-modes. The quantity of interest is the effective hydraulic conductivity k_{eff} defined in (4.1). We start again by numerically estimating the rates α and β for Theorem 2.1 and by comparing the costs of the MLMC method to standard MC. Figure 3 is similar to Figure 2 for 1D. The top two plots give graphs of the variances and the expected values of Q_ℓ and Y_ℓ. They suggest that $\mathbb{V}[Y_\ell] \lesssim h_\ell^2 \eqsim M_\ell^{-1}$ and thus $\beta \approx 1$, and $\mathbb{E}[Q - Q_\ell] \lesssim h_\ell^{1.75} \eqsim M_\ell^{-0.875}$, or $\alpha \approx 0.875$. Note that in terms of M_ℓ the rates are exactly $1/2$ those in 1D. In terms of h_ℓ they are the same. In the bottom right plot, the savings of the MLMC algorithm over standard MC are again considerable.

We now take a step away from Theorem 2.1, and analyse the gains of introducing different numbers of levels in the MLMC algorithm in more detail. First in Figure 4, we fix the standard deviation of our multilevel estimator $\widehat{Q}_M^{ML}$ (i.e. the sampling part of the error in (2.8)), and study how the computational cost of the MLMC method grows with grid size $M = m_L^2$ for various numbers L of levels. It is very clearly visible that the multilevel methods outperform standard MC dramatically. Note that the cost to estimate $\mathbb{E}[Q_L]$ to the required accuracy on a (finest) grid of

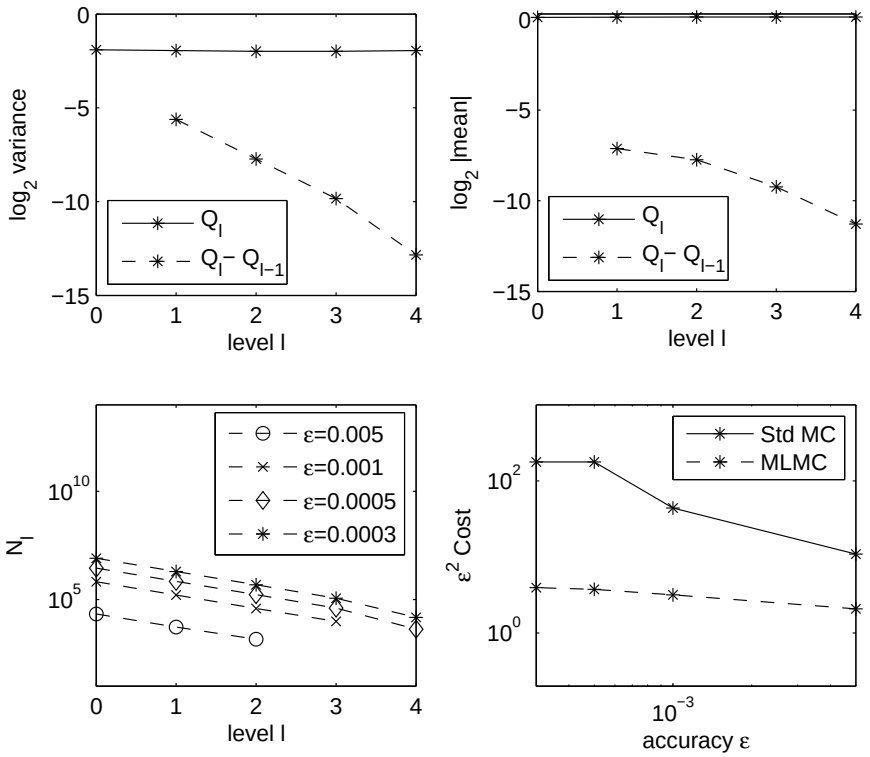

Figure 3: Performance plots for $\lambda = 0.3,\ \sigma^2 = 1,\ m_{\mathrm{KL}} = 1400$ and $m_0 = 8$ in 2D. The quantity of interest is the effective hydraulic conductivity k_{eff} defined in (4.1).

size $m_L = 32$ with standard MC is about the same as that of the 4–level method on a grid of size $m_L = 128$. In the left plot in Figure 4 we use $\gamma = 1$ (typical for an optimal iterative method such as AMG), whereas in the right plot we use $\gamma = 1.5$ (worst case for a sparse direct solver). We see that the gain is actually larger in the second case. For example, on the finest mesh with $m_L = 128$, the ratio of the costs of standard MC and the MLMC method with 4 levels is 67 for $\gamma = 1.5$, whereas this ratio is only 20 for $\gamma = 1$.

In Figure 5, we keep the spatial discretisation on the finest level m_L fixed and study how the computational cost grows as the tolerance on the required standard deviation of the estimator is decreased (using $\gamma = 1$). We can see in the left plot that for $Q = k_{\mathrm{eff}}$ the standard MC estimator only achieves a standard deviation of 3.7×10^{-4} for approximately the same cost as the MLMC estimator with 4 levels needs to reach a standard deviation of 7×10^{-5} which is below the discretisation error on that grid. More than 20 times more computational work is needed with standard MC to achieve a standard deviation that is smaller than the discretisation error. Again this gain is bigger if we assume $\gamma = 1.5$ in our cost model. In the right plot in Figure 5 we show that a similar behaviour is observed for other quantities of interest, such as the horizontal flux $-k\frac{\partial p}{\partial x_1}$ at the centre of the domain. Note however that there seems to be not much gain in including a 4th level in this case. This is related to the fact discussed at the end of Section 4.1. In Figure 6 we see that indeed the graphs of $\mathbb{V}[Q_\ell]$ and $\mathbb{V}[Y_\ell]$ are very close for $m_\ell = 8$ in this case. We also observe that the rate of

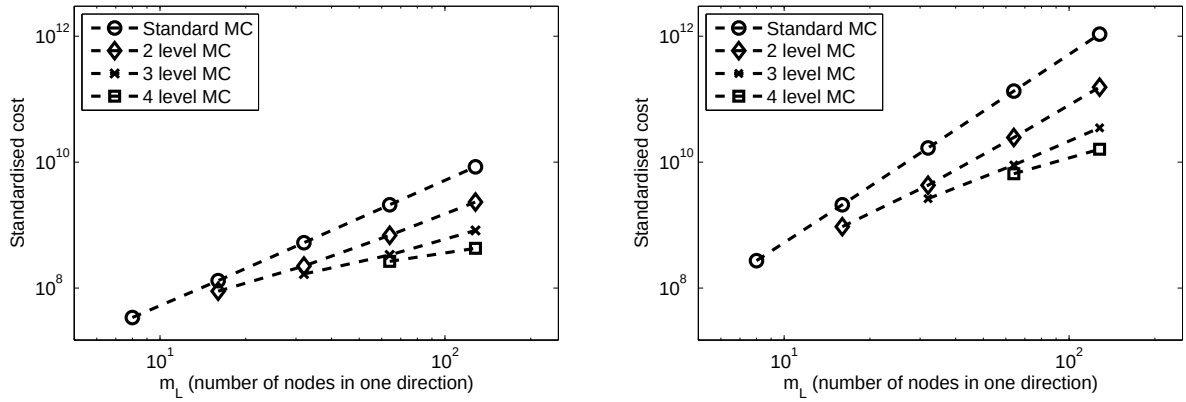

Figure 4: Plots of the standardised cost (scaled by $1/C^*$) versus m_L for a fixed tolerance of $\delta = 10^{-3}$ for the maximum standard deviation of the MLMC estimator for $\mathbb{E}[k_{\text{eff}}]$ for $\lambda = 0.3,\ \sigma^2 = 1,\ m_{\text{KL}} = 1400$ assuming $\gamma = 1$ (left plot) and $\gamma = 1.5$ (right plot).

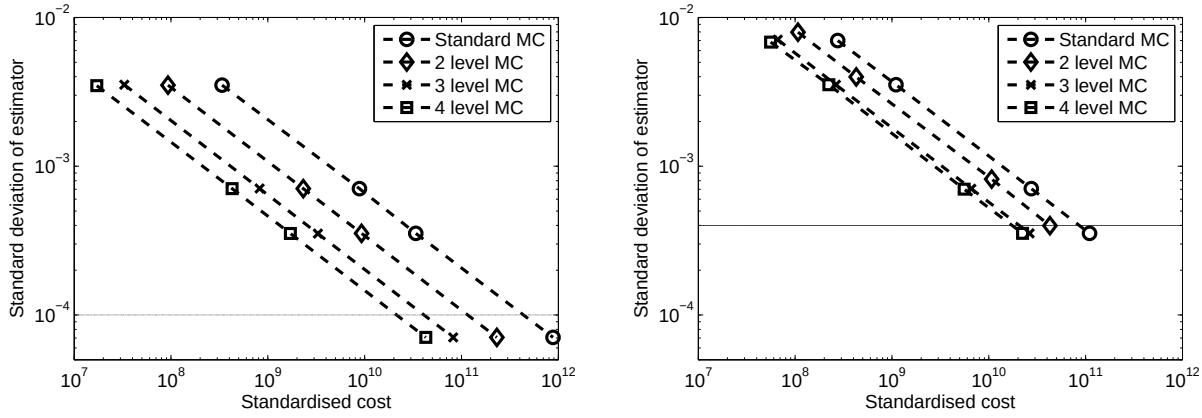

Figure 5: Same test case as in Figure 4. Standard deviation of the MLMC estimator $\widehat{Q}_L^{ML}$ versus the standardised computational cost for fixed $m_L = 128$. The horizontal line represents the estimated spatial discretisation error on this grid. Quantities of interest: $\mathbb{E}[k_{\text{eff}}]$ (left plot) and $\mathbb{E}\left[-k\frac{\partial p}{\partial x_1}\right]$ at the centre of the domain (right plot).

d	$\mathcal{C}(\widehat{Q}_M^{\mathrm{MC}})$	$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}})$	$\mathcal{C}(\widehat{Q}_1^{\mathrm{MC}})$
1	$\varepsilon^{-18/7}$	ε^{-2}	ε^{-2}
2	$\varepsilon^{-22/7}$	$\varepsilon^{-2}(\log\varepsilon)^2$	ε^{-2}
3	$\varepsilon^{-26/7}$	$\varepsilon^{-18/7}$	ε^{-2}

d	$\mathcal{C}(\widehat{Q}_M^{\mathrm{MC}})$	$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}})$	$\mathcal{C}(Q_M^{(i)})$
1	$\varepsilon^{-10/3}$	$\varepsilon^{-2}(\log\varepsilon)^2$	$\varepsilon^{-4/3}$
2	$\varepsilon^{-14/3}$	$\varepsilon^{-10/3}$	$\varepsilon^{-8/3}$
3	ε^{-6}	$\varepsilon^{-14/3}$	ε^{-4}

Table 1: Predicted asymptotic order of cost to achieve a RMSE of ε from Theorem 2.1 in the case of $\gamma = 1$, $\alpha = 1.75/d$ and $\beta = 2/d$ (left table) and $\gamma = 1$, $\alpha = 0.75/d$ and $\beta = 1/d$ (right table) for the standard MC ($\mathcal{C}(\widehat{Q}_M^{\mathrm{MC}})$) and MLMC ($\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}})$) estimators. In the left table we compare with the cost $\mathcal{C}(\widehat{Q}_1^{\mathrm{MC}})$ to obtain a RMSE of ε with the standard MC estimator for a single random variable, i.e. $M = 1$. In the right table we compare with the cost $\mathcal{C}(Q_M^{(i)})$ to obtain one sample on the finest grid.

decay for $\mathbb{V}[Y_\ell]$ and $\mathbb{E}[Y_\ell]$ is smaller for this quantity of interest, $\alpha \approx 0.375$ and $\beta \approx 0.5$ here.

Finally in Figure 7 we give some actual CPU times for a slightly harder test case, i.e. $\lambda = 0.1$, $\sigma^2 = 1$, $m_{\mathrm{KL}} = 500$ and $Q = k_{\mathrm{eff}}$. These were obtained with our `Matlab` implementation on a 3GHz Intel Core 2 Duo E8400 processor with 3.2GByte of RAM using the sparse direct solver provided in `Matlab` through the standard backslash operation to solve the linear systems for each sample. The value for γ we observed numerically in this case was 1.2 in 2D. We see that even with our non-optimised implementation it is possible to obtain a RMSE for $\widehat{Q}_M^{\mathrm{ML}}$ of less than 10^{-3} in just over 100 seconds. In the right plot in Figure 7 we further see that the advantage of the MLMC method is not restricted to computing expected values, but is just as apparent when computing the second order moment of quantities of interest.

5 Conclusions and Further Work

With the numerically observed values for α and β in the previous section (cf. Figures 2, 3 and 6) it is possible also to compare the theoretically predicted costs given by Theorem 2.1 for each of the two quantities of interest we studied and we do this in Table 1. This allows us also to project the expected gains of the MLMC method over the standard MC method to 3D. The numerical results above suggest that $\alpha \approx 1.75/d$ and $\beta \approx 2/d$ for $Q = k_{\mathrm{eff}}$, where $d = 1, 2, 3$ is the spatial dimension. For the flux at the centre of the domain $\alpha \approx 0.75/d$ and $\beta \approx 1/d$.

We see from Table 1 that asymptotically the MLMC leads to a huge improvement over standard MC for both quantities of interest. In cases where the variance of Y_ℓ decays relatively rapidly, as is the case for $Q = k_{\mathrm{eff}}$, then a relatively large portion of the computational effort is spent on the coarse grids. Indeed, if we would have $\beta > \gamma$ in $d = 1, 2, 3$, then the MLMC method would have a cost that is of asymptotic order ε^{-2}. Note that this is the same asymptotic cost as applying standard MC to a problem with only one random variable, i.e. $M = 1$. We see in the left table in Table 1 that the cost of MLMC estimator for $Q = k_{\mathrm{eff}}$ does indeed have an asymptotic order close to ε^{-2}, for $d = 1, 2, 3$.

When the variance of Y_ℓ decays more slowly, on the other hand, as is the case for the flux at the centre of the domain, then a relatively large portion of the computational effort is spent on the finest grid. We would like to point out here that if we are in the situation that $\beta < \gamma$ for $d = 1, 2, 3$ and $\beta = 2\alpha$, then the MLMC method has a cost that asymptotically is of the order $\varepsilon^{-\gamma/\alpha}$. Note

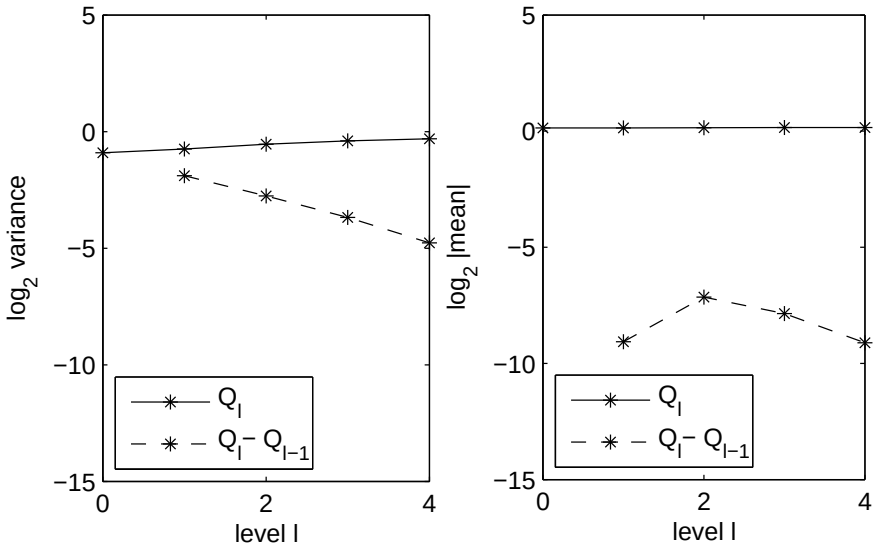

Figure 6: Plots of $\mathbb{V}[Q_\ell]$ and $\mathbb{V}[Y_\ell]$ (left plot), as well as $\mathbb{E}[Q_\ell]$ and $\mathbb{E}[Y_\ell]$ (right plot) in the case $\lambda = 0.3,\ \sigma^2 = 1,\ m_{\mathrm{KL}} = 1400$ in 2D, with $m_0 = 4$ and $Q = -k\frac{\partial p}{\partial x_1}$ at the centre of the domain.

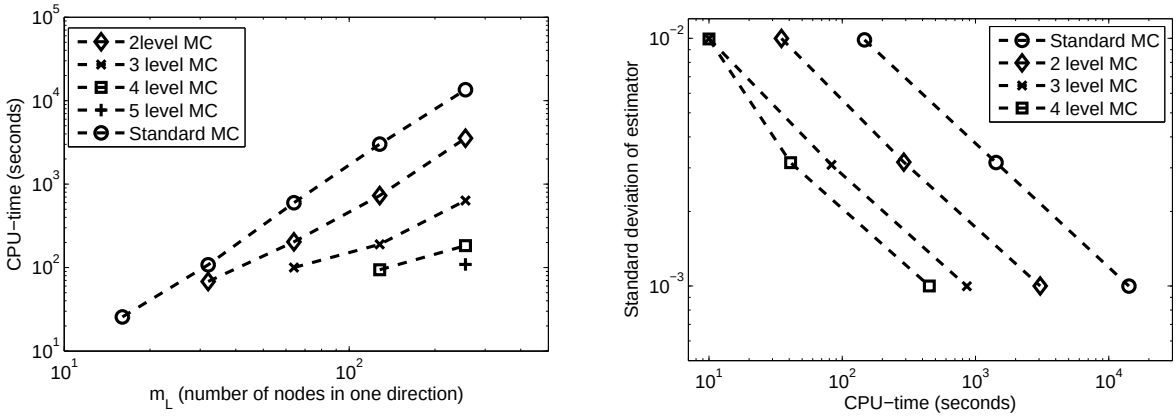

Figure 7: Same plots as in Figures 4 and 5 for $\lambda = 0.1,\ \sigma^2 = 1,\ m_{\mathrm{KL}} = 500$, but using actual CPU time in seconds to quantify the cost (`Matlab` implementation running on a 3GHz Intel Core 2 Duo E8400 processor with 3.2GByte of RAM). Left plot: $\mathbb{E}[k_{\mathrm{eff}}]$, with fixed maximum standard deviation 10^{-3}. Right plot: $\mathbb{E}[k_{\mathrm{eff}}^2]$, with $m_L = 256$.

that this is proportional to the cost of obtaining one sample on the finest grid and thus to solving a deterministic PDE to the same accuracy ε. For the horizontal flux $-k\frac{\partial p}{\partial x_1}$ at the centre of the domain, we do not quite have $\beta = 2\alpha$, but we see in the right table in Table 1 that the asymptotic order of the cost of the MLMC estimator is indeed close to that of the cost of obtaining one sample on the finest grid.

To conclude, in this paper we successfully applied the MLMC algorithm to elliptic PDEs with random coefficients. The numerical results clearly show the advantage of using the MLMC estimator over a standard MC estimator for this type of model problem for several quantities of interest. They further show that the gain of the MLMC estimator is not limited to smooth or easy problems. The improvements are in fact even more pronounced when the linear solver is not quite optimal ($\gamma > 1$), or in cases where the discretisation error is large (α and β are small).

There was nothing special about our choice of uniform grids and isotropic model problems. The MLMC estimator is expected to perform equally well on locally refined grids and for anisotropic problems provided a suitable hierarchy of grid levels can be constructed. The numerical experiments suggest ways to further improve the performance of the MLMC algorithm. Firstly, as discussed, in order to choose the coarse level M_0 independent of λ it would be better to choose smoother approximations of the random field on the coarse meshes, e.g. by truncating the KL-expansion earlier. This will not affect the asymptotic order of the cost as $\varepsilon \to 0$, but it would lead to larger gains of MLMC over standard MC for a fixed tolerance ε. A way to improve the asymptotic order of convergence of the MLMC method may be the use of a different estimator on each of the levels, such as a randomised quasi–Monte Carlo estimator [14, 15].

This paper has not addressed the challenges of numerical analysis, but the assumptions of Theorem 2.1 have recently been verified theoretically for certain quantities of interest in the context of finite element spatial discretisations in [1] and [5]. The former considers coefficient fields $k \in W^{1,\infty}$ that are bounded uniformly from above and away from zero. The latter analyses the more challenging case studied in this paper, where k is not uniformly bounded and is only in $\mathcal{C}^{0,\eta}$, with $\eta < 1/2$.

A Proof of the generalised multilevel Monte Carlo Theorem 2.1

Let us denote the hidden constants in Assumptions *i)*, *ii)* and *iii)* by c_1, c_2 and c_3, respectively. Recall that we assume that

$$M_\ell = s\, M_{\ell-1}\,, \qquad \text{for all } \ell = 1, \ldots, L,$$

for some $s \in \mathbb{N} \setminus \{1\}$, cf (2.4). Without loss of generality, we shall also assume that $M_0 = 1$. If this is not the case, this will only scale the constants c_1, c_2 and c_3.

Note also that since standard Monte Carlo estimators are unbiased, we have

$$\mathbb{E}[\widehat{Y}_\ell] = \begin{cases} \mathbb{E}[Q_{M_\ell}], & \ell = 0 \\ \mathbb{E}[Q_{M_\ell} - Q_{M_{\ell-1}}], & \ell > 0 \end{cases} \tag{A.1}$$

Then, using the notation $\lceil x \rceil$ to denote the unique integer n satisfying the inequalities $x \le n < x+1$, we start by choosing L to be

$$L = \left\lceil \alpha^{-1} \log_s(\sqrt{2}\, c_1\, \varepsilon^{-1}) \right\rceil \;<\; \alpha^{-1} \log_s(\sqrt{2}\, c_1\, \varepsilon^{-1}) + 1 \tag{A.2}$$

so that

$$s^{-\alpha}\frac{\varepsilon}{\sqrt{2}} \;<\; c_1\, s^{-\alpha\, L} \;\leq\; \frac{\varepsilon}{\sqrt{2}}, \tag{A.3}$$

and hence, due to (A.1) and assumption i), $\left(\mathbb{E}[\widehat{Q}_M^{\mathrm{ML}}] - \mathbb{E}[Q]\right)^2 \leq \frac{1}{2}\,\varepsilon^2$. This $\frac{1}{2}\varepsilon^2$ upper bound on the square of the bias error, together with the $\frac{1}{2}\varepsilon^2$ upper bound on the variance of the estimator to be proved later, gives an ε^2 upper bound on the estimator MSE.

Using the left-hand inequality in (A.3), we obtain the following inequality which will be used later,

$$\sum_{\ell=0}^{L} s^{\gamma\,\ell} \;<\; \frac{s^{\gamma L}}{1-s^{-\gamma}} \;<\; \frac{s^{\gamma}\,(\sqrt{2}\,c_1)^{\gamma/\alpha}}{1-s^{-\gamma}}\;\; \varepsilon^{-\gamma/\alpha}. \tag{A.4}$$

We now need to consider the different possible values for β.

a) If $\beta{=}\gamma$, we set $N_\ell = \left\lceil 2\,\varepsilon^{-2}\,(L{+}1)\,c_2\,s^{-\beta\,\ell}\right\rceil$ so that

$$\mathbb{V}[\widehat{Q}_M^{\mathrm{ML}}] \;=\; \sum_{\ell=0}^{L} \mathbb{V}[\widehat{Y}_\ell] \;\leq\; \sum_{\ell=0}^{L} c_2\, N_\ell^{-1} s^{-\beta\,\ell} \;\leq\; \tfrac{1}{2}\,\varepsilon^2,$$

which is the required upper bound on the variance of the estimator. Since $N_\ell \leq 2\varepsilon^{-2}\,(L{+}1)\,c_2\,s^{-\beta\,\ell} + 1$, the computational complexity is bounded by

$$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \leq\; c_3 \sum_{\ell=0}^{L} N_l\, s^{\gamma\,\ell} \;\leq\; c_3 \left(2\,\varepsilon^{-2}(L{+}1)^2\,c_2 \;+\; \sum_{\ell=0}^{L} s^{\gamma\,\ell}\right)$$

For $\varepsilon < e^{-1} < 1$ we have $1 < \log\varepsilon^{-1}$ and $\varepsilon^{-\gamma/\alpha} \leq \varepsilon^{-2} \leq \varepsilon^{-2}(\log\varepsilon)^2$ since $\alpha \geq \frac{1}{2}\gamma$. Hence, using the inequalities in (A.2) and (A.4), it follows that $\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \lesssim \varepsilon^{-2}(\log\varepsilon)^2$.

b) For $\beta{>}\gamma$, we set $N_\ell = \left\lceil 2\,\varepsilon^{-2}\,c_2\,\left(1{-}s^{-(\beta-\gamma)/2}\right)^{-1} s^{-(\beta+\gamma)\ell/2}\right\rceil$ so that

$$\sum_{\ell=0}^{L} \mathbb{V}[\widehat{Y}_\ell] \;\leq\; \tfrac{1}{2}\,\varepsilon^2\,\left(1{-}s^{-(\beta-\gamma)/2}\right)\,\sum_{\ell=0}^{L} s^{-(\beta-\gamma)\ell/2} \;<\; \tfrac{1}{2}\,\varepsilon^2.$$

Since

$$N_\ell < 2\,\varepsilon^{-2}\,c_2\,\left(1{-}s^{-(\beta-\gamma)/2}\right)^{-1} s^{-(\beta+\gamma)\ell/2} + 1,$$

the computational complexity is bounded by

$$\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \leq c_3 \left(2\,\varepsilon^{-2}\,c_2\,\left(1{-}s^{-(\beta-\gamma)/2}\right)^{-2} +\; \sum_{l=0}^{L} s^{\gamma\,l}\right).$$

Again for $\varepsilon < e^{-1} < 1$ we have $\varepsilon^{-\gamma/\alpha} \leq \varepsilon^{-2}$ and hence due to inequality (A.4) we have $\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \lesssim \varepsilon^{-2}$.

c) For $\beta{<}\gamma$, we set $N_\ell = \left\lceil 2\,\varepsilon^{-2} c_2\, s^{(\gamma-\beta)L/2}\left(1{-}s^{-(\gamma-\beta)/2}\right)^{-1} s^{-(\beta+\gamma)\ell/2}\right\rceil$, so that

$$\sum_{\ell=0}^{L} \mathbb{V}[\widehat{Y}_\ell] \;<\; \tfrac{1}{2}\,\varepsilon^2\, s^{-(\gamma-\beta)L/2}\left(1{-}s^{-(\gamma-\beta)/2}\right)\,\sum_{\ell=0}^{L} s^{(\gamma-\beta)\ell/2} \;<\; \tfrac{1}{2}\,\varepsilon^2.$$

Since $N_\ell < 2\,\varepsilon^{-2}\,c_2\,s^{(\gamma-\beta)L/2}\left(1-s^{-(\gamma-\beta)/2}\right)^{-1}s^{-(\beta+\gamma)\ell/2}+1$, the computational complexity is bounded by

$$\begin{aligned}\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) &\leq c_3\left(2\,\varepsilon^{-2}\,c_2\,s^{(\gamma-\beta)L/2}\left(1-s^{-(\gamma-\beta)/2}\right)^{-1}\sum_{\ell=0}^{L}s^{(\gamma-\beta)\ell/2}+\sum_{\ell=0}^{L}s^{\gamma\ell}\right)\\ &\leq c_3\left(2\,\varepsilon^{-2}\,c_2\,s^{(\gamma-\beta)L}\left(1-s^{-(\gamma-\beta)/2}\right)^{-2}+\sum_{\ell=0}^{L}s^{\gamma\ell}\right).\end{aligned}$$

Using the first inequality in (A.3), $s^{(\gamma-\beta)L} < \left(\sqrt{2}\,c_1\right)^{(\gamma-\beta)/\alpha}s^{\gamma-\beta}\,\varepsilon^{-(\gamma-\beta)/\alpha}$. Also, for $\varepsilon < e^{-1} < 1$ we have $\varepsilon^{-\gamma/\alpha} \leq \varepsilon^{-2-(\gamma-\beta)/\alpha}$ since $\alpha \geq \frac{1}{2}\beta$. Hence, due to inequality (A.4), we have $\mathcal{C}(\widehat{Q}_M^{\mathrm{ML}}) \lesssim \varepsilon^{-2-(\gamma-\beta)/\alpha}$.

References

[1] A. Barth, C. Schwab, and N. Zollinger. Multi–level Monte Carlo finite element method for elliptic PDE's with stochastic coefficients. *Numer. Math.*, Online First, 09 April 2011.

[2] A. Brandt, M. Galun, and D. Ron. Optimal multigrid algorithms for calculating thermodynamic limits. *J. Stat. Phys.*, 74(1-2):313–348, 1994.

[3] A. Brandt and V. Ilyin. Multilevel monte carlo methods for studying large scale phenomena in fluids. *J. Mol. Liq.*, 105(2-3):245–248, 2003.

[4] J. Charrier. Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients. Technical Report HAL:inria-00490045, Hyper Articles en Ligne, INRIA, 2010. Available at `http://hal.inria.fr/inria-00490045/en/`.

[5] J. Charrier, R. Scheichl, and A.L. Teckentrup. Finite element error analysis of elliptic pdes with random coefficients and its application to multilevel monte carlo methods. Technical report, University of Bath, 2011. BICS Preprint 02/11.

[6] K.A. Cliffe, I.G. Graham, R. Scheichl, and L. Stals. Parallel computation of flow in heterogeneous media using mixed finite elements. *J. Comput. Phys.*, 164:258–282, 2000.

[7] G. de Marsily. *Quantitative Hydrogeology.* Academic Press, 1986.

[8] G. de Marsily, F. Delay, J. Goncalves, P. Renard, V. Teles, and S. Violette. Dealing with spatial heterogeneity. *Hydrogeol. J*, 13:161–183, 2005.

[9] P. Delhomme. Spatial variability and uncertainty in groundwater flow param- eters, a geostatistical approach. *Water Resourc. Res.*, pages 269–280, 1979.

[10] O.G. Ernst, C.E. Powell, D.J. Silvester, and E. Ullmann. Efficient solvers for a linear stochastic galerkin mixed formulation of diffusion problems with random data. *SIAM J. Sci. Comput.*, 31(2):1424–1447, 1979.

[11] R.G. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach.* Springer-Verlag, New York, 1991.

[12] M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. volume 256 of *Monte Carlo and Quasi-Monte Carlo methods 2006*, pages 343–358. Springer, 2007.

[13] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 256:981–986, 2008.

[14] M.B. Giles and B.J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. In *Advanced financial modelling*, volume 8 of *Radon Ser. Comput. Appl. Math.*, pages 165–181. Walter de Gruyter, Berlin, 2009.

[15] I.G. Graham, F.Y. Kuo, D. Nuyens, R. Scheichl, and I.H. Sloan. Quasi-monte carlo methods for elliptic pdes with random coefficients and applications. *J. Comput. Phys.*, 230(10):3668–3694, 2011.

[16] S. Heinrich. Multilevel Monte Carlo methods. volume 2179 of *Lecture notes in Computer Science*, pages 3624–3651. Springer, Philadelphia, PA, 2001.

[17] R.J. Hoeksema and P.K. Kitanidis. Analysis of the spatial structure of properties of selected aquifers. *Water Resour. Res.*, 21:536–572, 1985.

[18] O.G. Ernst M. Eiermann and E. Ullmann. Computational aspects of the stochastic finite element method.

[19] O.P. Le Maître and O.M. Kino. *Spectral Methods for Uncertainty Quantification, With Applications to Fluid Dynamics*. Springer-Verlag, 2010.

[20] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[21] C. Schwab and R.A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[22] P.S. Vassilevski. *Multilevel Block Factorization Preconditioners: Matrix-based Analysis and Algorithms for Solving Finite Element Equations*. Springer-Verlag, 2008.

[23] D. Xiu. *Numerical Methods for Stochastic Computations, A Spectral Method Approach*. Princeton University Press, 2010.