MULTILEVEL MONTE CARLO METHODS FOR HIGHLY HETEROGENEOUS MEDIA

Aretha L. Teckentrup

Department of Mathematical Sciences
University of Bath
Claverton Down
Bath, BA2 7AY, UK

ABSTRACT

We discuss the application of multilevel Monte Carlo methods to elliptic partial differential equations with random coefficients. Such problems arise, for example, in uncertainty quantification in subsurface flow modeling. We give a brief review of recent advances in the numerical analysis of the multilevel algorithm under minimal assumptions on the random coefficient, and extend the analysis to cover also tensor–valued coefficients, as well as point evaluations. Our analysis includes as an example log–normal random coefficients, which are frequently used in applications.

1 INTRODUCTION

There are many situations in which modeling 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 wastes. 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 modeling 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 (de Marsily 1986). The main parameter appearing in Darcy's Law is the permeability, which characterizes how easily water can flow through the rock under a given pressure gradient. In practice it is only possible to measure the permeability at a limited number of spatial locations, and these measurements are too sparse to infer the structure of the permeability over the whole computational domain for the simulation. This fact is the primary source of uncertainty in groundwater flow calculations, 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 permeability as a random field (Delhomme 1979; de Marsily et al. 2005). A model frequently used is a log–normal random field, with a covariance function that is only Lipschitz continuous. Individual realizations of such fields have low spatial regularity and significant spatial variation, making the problem of solving for the pressure very costly. The slow rate of convergence of the standard Monte Carlo algorithm means that many such realizations are required to obtain accurate results, rendering the problem computationally unfeasible.

In this paper, we therefore employ the multilevel Monte Carlo (MLMC) method. This method was first introduced by Giles (2008) in the context of stochastic differential equations in finance, and similar ideas were also used by Heinrich (2001) and Brandt and Ilyin (2003). In the context of our groundwater flow model problem, it was shown in for example (Cliffe et al. 2011) and (Teckentrup et al. 2012), that

the multilevel method leads to a significant reduction in the computational cost required to achieve a given accuracy.

The main challenge in the numerical analysis of MLMC methods for elliptic partial differential equations (PDEs) with random coefficients, is the quantification of the numerical discretization error, or in other words the *bias* of the estimator. Models for the random coefficient frequently used in applications, such as log–normal random fields, are not uniformly coercive, making the numerical analysis challenging. A rigorous analysis of the MLMC algorithm under minimal assumptions on the random coefficient was recently carried out by Charrier, Scheichl, and Teckentrup (2011) and Teckentrup et al. (2012). In particular, uniform coercivity or boundedness were not assumed in these papers. If one does assume uniform coercivity and boundedness of the coefficient, the analysis of the discretization error is classical, and an analysis of the MLMC method for this case can be found in (Barth, Schwab, and Zollinger 2011). Other related works on numerical errors for elliptic PDEs with random coefficients are (Charrier 2012) and (Gittelson 2010).

The aim of this paper is to extend the theory in (Teckentrup et al. 2012). We here consider the case of more general, tensor–valued models of the permeability, which are often used in applications to model orthotropic media. We will also prove convergence of the MLMC algorithm for point evaluations of the pressure or the Darcy flux.

The outline of the paper is as follows. In §2, we describe the multilevel Monte Carlo algorithm applied to elliptic PDEs with random coefficients, and discuss its performance. In §3, we then prove an upper bound on the computational cost of the multilevel Monte Carlo estimator. We recall some of the main results from (Charrier, Scheichl, and Teckentrup 2011; Teckentrup et al. 2012), before extending the results to tensor–coefficients and point evaluations.

2 MULTILEVEL MONTE CARLO SIMULATION

The classical equations governing a steady state, single phase flow, are Darcy's law coupled with an incompressibility condition. These equations can be written in second order form as

$$-\operatorname{div}(\mathbf{A}\nabla u) = f, \qquad \text{in} \quad D \subset \mathbb{R}^d, \tag{1}$$

subject to appropriate boundary conditions. Here, $\mathbf{A}$ is the permeability tensor, u is the resulting pressure field, and f are the source terms. Modeling $\mathbf{A}$ as a random field, u also becomes a random field.

In applications, one is then usually interested in finding the expected value of some functional $Q = M(u)$ of the solution u to our model problem (1). This could for example be the value of the pressure u or the Darcy flux $-\mathbf{A}\nabla u$ at or around a given point in the computational domain, or outflow over parts of the boundary. Since u is not easily accessible, Q is often approximated by the quantity $Q_h := M(u_h)$, where u_h is a finite dimensional approximation to u, such as the finite element solution on a sufficiently fine spatial grid $\mathscr{T}_h$.

To estimate $\mathbb{E}[Q]$, we then compute approximations (or *estimators*) $\widehat{Q}_h$ to $\mathbb{E}[Q_h]$, and quantify the accuracy of our approximations via the root mean square error (RMSE)

$$e(\widehat{Q}_h) := \left(\mathbb{E}\big[(\widehat{Q}_h - \mathbb{E}(Q))^2 \big] \right)^{1/2}.$$

The computational cost $\mathscr{C}_\varepsilon(\widehat{Q}_h)$ of our estimator is then quantified by the number of floating point operations that are needed to achieve a RMSE of $e(\widehat{Q}_h) \leq \varepsilon$. This will be referred to as the ε–cost.

The classical Monte Carlo (MC) estimator for $\mathbb{E}[Q_h]$ is

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

where $Q_h(\omega^{(i)})$ is the ith sample of Q_h and N independent samples are computed in total.

There are two sources of error in the estimator (2), the approximation of Q by Q_h, which is related to the spatial discretization, and the sampling error due to replacing the expected value by a finite sample average. This becomes clear when expanding the mean square error (MSE) and using the fact that for Monte Carlo $\mathbb{E}[\widehat{Q}_{h,N}^{\text{MC}}] = \mathbb{E}[Q_h]$ and $\mathbb{V}[\widehat{Q}_{h,N}^{\text{MC}}] = N^{-1}\mathbb{V}[Q_h]$, where $\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$ denotes the variance of the random variable $X : \Omega \to \mathbb{R}$. We get

$$e(\widehat{Q}_{h,N}^{\text{MC}})^2 \;=\; N^{-1}\mathbb{V}[Q_h] + \big(\mathbb{E}[Q_h - Q]\big)^2. \tag{2}$$

A sufficient condition to achieve a RMSE of ε with this estimator is that both of these terms are less than $\varepsilon^2/2$. For the first term, this is achieved by choosing a large enough number of samples, $N = \mathcal{O}(\varepsilon^{-2})$. For the second term, we need to choose a fine enough finite element mesh $\mathcal{T}_h$, such that $\mathbb{E}[Q_h - Q] = \mathcal{O}(\varepsilon)$.

The main idea of the MLMC estimator is very simple. We sample not just from one approximation Q_h of Q, but from several. Linearity of the expectation operator implies that

$$\mathbb{E}[Q_h] = \mathbb{E}[Q_{h_0}] + \sum_{\ell=1}^{L} \mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}]$$

where $\{h_\ell\}_{\ell=0,\dots,L}$ are the mesh widths of a sequence of increasingly fine triangulations $\mathcal{T}_{h_\ell}$ with $\mathcal{T}_h := \mathcal{T}_{h_L}$, the finest mesh. Hence, the expectation on the finest mesh is equal to the expectation on the coarsest mesh, plus a sum of corrections adding the difference in expectation between simulations on consecutive meshes. The multilevel idea is now to independently estimate each of these terms such that the overall variance is minimized for a fixed computational cost.

Setting for convenience $Y_0 := Q_{h_0}$ and $Y_\ell := Q_{h_\ell} - Q_{h_{\ell-1}}$, for $1 \le \ell \le L$, we define the MLMC estimator simply as

$$\widehat{Q}_{h,\{N_\ell\}}^{\text{ML}} \;:=\; \sum_{\ell=0}^{L} \widehat{Y}_{\ell,N_\ell}^{\text{MC}},$$

where $\widehat{Y}_{\ell,N_\ell}^{\text{MC}}$ is the standard MC estimator for Y_ℓ,

$$\widehat{Y}_{\ell,N_\ell}^{\text{MC}} \;=\; \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} Y_\ell(\omega^{(i)}).$$

Here, it is important to note that $Y_\ell(\omega^{(i)}) = Q_{h_\ell}(\omega^{(i)}) - Q_{h_{\ell-1}}(\omega^{(i)})$, i.e. the quantity $Y_\ell(\omega^{(i)})$ is computed using the same sample on both meshes.

Since all the expectations $\mathbb{E}[Y_\ell]$ are estimated independently in (2), the variance of the MLMC estimator is $\sum_{\ell=0}^{L} N_\ell^{-1}\mathbb{V}[Y_\ell]$ and expanding as in (2) leads again to a MSE of the form

$$e(\widehat{Q}_{h,\{N_\ell\}}^{\text{ML}})^2 := \mathbb{E}\Big[\big(\widehat{Q}_{h,\{N_\ell\}}^{\text{ML}} - \mathbb{E}[Q]\big)^2\Big] \;=\; \sum_{\ell=0}^{L} N_\ell^{-1}\mathbb{V}[Y_\ell] \;+\; \big(\mathbb{E}[Q_h - Q]\big)^2.$$

As in the classical MC case before, we see that the MSE consists of two terms, the variance of the estimator and the error in mean between Q and Q_h. Note that the second term is identical to the second term for the classical MC method in (2). A sufficient condition to achieve a RMSE of ε is again to make both terms less than $\varepsilon^2/2$. This is easier to achieve with the MLMC estimator, as

- for sufficiently large h_0, samples of Q_{h_0} are much cheaper to obtain than samples of Q_h;
- the variance Y_ℓ tends to 0 as $h_\ell \to 0$, meaning we need fewer samples on $\mathcal{T}_{h_\ell}$, for $\ell > 0$.

Let now $\mathscr{C}_\ell$ denote the cost to obtain one sample of Q_{h_ℓ}. Then we have the following results on the ε–cost of the MLMC estimator (cf. Cliffe et al. (2011), Giles (2008)).

Theorem 1 Suppose there are positive constants $\alpha, \beta, \gamma, c_{M1}, c_{M2}, c_{M3} > 0$ such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and

M1. $\quad |\mathbb{E}[Q_h - Q]| \leq c_{M1}\, h^{\alpha}$,

M2. $\quad \mathbb{V}[Q_{h_\ell} - Q_{h_{\ell-1}}] \leq c_{M2}\, h_\ell^{\beta}$,

M3. $\quad \mathscr{C}_\ell \leq c_{M3}\, h_\ell^{-\gamma}$,

Then, for any $\varepsilon < e^{-1}$, there exist an L and a sequence $\{N_\ell\}_{\ell=0}^{L}$, such that $e(\widehat{Q}_{h,\{N_\ell\}}^{ML}) < \varepsilon$ and

$$
\mathscr{C}_\varepsilon(\widehat{Q}_{h,\{N_\ell\}}^{ML}) \lesssim
\begin{cases}
\varepsilon^{-2}, & \text{if } \beta > \gamma, \\
\varepsilon^{-2}(\log \varepsilon)^2, & \text{if } \beta = \gamma, \\
\varepsilon^{-2-(\gamma-\beta)/\alpha}, & \text{if } \beta < \gamma,
\end{cases}
$$

where the hidden constant depends on c_{M1}, c_{M2} and c_{M3}. For the classical MC estimator we have $\mathscr{C}_\varepsilon(\widehat{Q}_h^{MC}) \lesssim \varepsilon^{-2-\gamma/\alpha}$, where the hidden constant depends on c_{M1} and c_{M3}.

The convergence rates α and β in Theorem 1 are related to the convergence of the spatial discretization error, and have been proven for various quantities of interest in (Charrier, Scheichl, and Teckentrup 2011) and (Teckentrup et al. 2012). In §3.2, we further extend this theory to point evaluations of the pressure and the flux. Typical values of α and β for the model problem considered in this paper are $\alpha = 1$ and $\beta = 2$ for rough models of the permeability and $\alpha = 2$ and $\beta = 4$ for smoother models (this is made more precise in §3). The rate γ is related to the cost of numerically solving the PDE for one realization of the random coefficient. This involves producing a sample of the random coefficient, and solving a linear system of equations. The cost of solving the linear system will generally be dominant, and with an optimal linear solver, the cost of one such solve is proportional to h_ℓ^{-d}, the number of unknowns, and so $\gamma \approx d$.

In Table 1, we show the ε–costs as predicted by Theorem 1, for typical values of α and β. We assume an almost optimal linear solver, and take γ to be slightly larger than d. We see that the gains we can expect from using the MLMC estimators are always significant, usually in the order of two orders of magnitude. It is also worth noting that although the actual ε–costs are higher in the case of the rough model problem with $\alpha = 1$ and $\beta = 2$, the gains we can expect from MLMC are also greater in this case.

Table 1: Upper bounds for the ε-costs of classical and multilevel Monte Carlo from Theorem 1 in the cases $\alpha = 1, \beta = 2$ (left) and $\alpha = 2, \beta = 4$ (right), with $\gamma = d + \delta$ in both cases, where $\delta > 0$ is a small constant. d is the spatial dimension from (1).

d	$\alpha = 1, \beta = 2$		$\alpha = 2, \beta = 4$	
	MC	MLMC	MC	MLMC
1	ε^{-3}	ε^{-2}	$\varepsilon^{-5/2}$	ε^{-2}
2	ε^{-4}	ε^{-2}	ε^{-3}	ε^{-2}
3	ε^{-5}	ε^{-3}	$\varepsilon^{-7/2}$	ε^{-2}

The reduction in cost associated with the MLMC estimator over standard MC is largely due to the fact that the number of samples needed on the finer grids is greatly reduced. Most of the uncertainty can already be captured on the coarse grids, and so the MLMC estimator shifts some of the computational effort on to the coarse grids. Exactly how much of the computational effort can (and should) be shifted towards the coarse grids, depends on the model problem and the quantity of interest Q. The MLMC algorithm described above chooses the number of samples on each level in such a way that the computational cost of the estimator is minimized, subject to the overall variance of the estimator being less than $\varepsilon^2/2$. This can lead to three different scenarios: the computational cost could be predominantly on the coarse levels, spread evenly across the levels, or predominantly on the fine levels. This corresponds to the three upper bounds given in Theorem 1 above.

To make this more precise, note that for given $\{N_\ell\}$ and $\{\mathscr{C}_\ell\}$, the computational cost of the MLMC estimator is

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

Treating the N_ℓ as continuous variables, the cost of the MLMC estimator is minimized for a fixed variance by choosing

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

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

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

If the variance $\mathbb{V}[Y_\ell]$ decays faster with ℓ than the cost $\mathscr{C}_\ell$ increases, i.e. if $\beta > \gamma$, the dominant term will be on level 0. Similarly, if $\mathbb{V}[Y_\ell]$ decays slower than $\mathscr{C}_\ell$ increases, the dominant term will be on the finest level L, and if $\mathbb{V}[Y_\ell]$ decreases at the same rate as $\mathscr{C}_\ell$ increases, the cost is spread evenly across all levels. In the context of our model problem in subsurface flow, we are usually in the last regime, where $\beta < \gamma$. Especially in 2 or 3 space dimensions, the cost of obtaining one sample grows very rapidly, and the dominant cost will always be on the finest level. It is worth noting that if $\beta = 2\alpha$, and like here, $\beta < \gamma$, the cost of the MLMC estimator is of the order $\varepsilon^{-\gamma/\alpha}$. This is in fact the same cost as taking only one sample on the finest grid, since we have to choose $h \eqsim \varepsilon^{1/\alpha}$ to get a MSE of $\mathscr{O}(\varepsilon^2)$, and the cost of one solve is then $\mathscr{C} \lesssim h^{-\gamma} = \varepsilon^{-\gamma/\alpha}$, by assumption M3. This means that asymptotically our multilevel Monte Carlo method for the stochastic problem has the same complexity as a deterministic solver for one realization of the same problem.

Another issue which influences the cost of the MLMC estimator, is the choice of the coarsest mesh size h_0. The bigger h_0 is, the more levels we can include in the MLMC estimator, and the bigger the potential gains are with respect to standard MC. Although the choice of h_0 does not influence the asymptotic bounds on the cost given in Theorem 1, the choice of h_0 does have an effect on the absolute cost of the MLMC estimator for any fixed accuracy ε. In practical applications, h_0 must often be chosen to give a minimal level of resolution to the problem in order to get the MLMC estimator with the smallest absolute cost. For the model problem in subsurface flow, where the permeability varies on a very fine scale and is highly oscillatory, very coarse meshes do not yield a good representation of the problem, and including them in the MLMC estimator can lead to a larger absolute cost than necessary. One way to circumvent this problem, is to use smoother representations of the permeability on the coarse levels. It was shown in (Teckentrup et al. 2012) that, without introducing any additional bias in the MLMC estimator, this strategy allows for the inclusion of much coarser levels, and hence gives a significantly lower absolute cost of the MLMC estimator, even in the context of short correlation lengths.

The rest of the paper is devoted to proving theoretical convergence rates, and thus justifying assumptions M1 and M2 in Theorem 1.

3 NUMERICAL ANALYSIS

For simplicity, we consider a particular instance of model problem (1), posed on a Lipschitz–polygonal domain $D \subset \mathbb{R}^2$ and with homogeneous Dirichlet conditions: Given a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and $\omega \in \Omega$, find $u : \Omega \times D \to \mathbb{R}$ such that

$$-\mathrm{div}\,(\mathbf{A}(\omega,x)\nabla u(\omega,x)) = f(\omega,x), \qquad \text{for } x \in D, \tag{3}$$

$$u(\omega,x) = 0, \qquad \text{for } x \in \Gamma_j.$$

The differential operators div and ∇ are with respect to $x \in D$, and $\Gamma := \cup_{j=1}^{m} \overline{\Gamma}_j$ denotes the boundary of D, partitioned into straight line segments. Note that due to the tensor–valued coefficient $\mathbf{A}(\omega, x)$, this problem is more general than those studied in our earlier papers (Charrier, Scheichl, and Teckentrup 2011) and (Teckentrup et al. 2012).

We will carry out a finite element error analysis of (3), under minimal assumptions on the coefficient tensor $\mathbf{A}$ and on the source term f. In particular, we do not assume that $\mathbf{A}$ is coercive and bounded *uniformly* in ω, since this is not the case for example for log–normal random fields, which can take values arbitrarily close to zero or infinity for any given realization. The crucial observation is that for each *fixed* ω, we have a uniformly coercive and bounded problem (in x). The first step in our error analysis is therefore to derive an estimate on the finite element error for a fixed ω. However, in order to be able to compute moments (expectations) of the error, it is crucial that we keep track of how all the constants that appear in our estimates depend on ω, or in other words on $\mathbf{A}$ and f.

The coefficient tensor $\mathbf{A}(\omega, \cdot)$ is assumed to take values in the space of real–valued, symmetric 2×2 matrices. Given the usual norm $|v| := \left(\sum_{i=1}^{2} |v_i^2| \right)^{1/2}$ on $\mathbb{R}^2$, we choose the norm on $\mathbb{R}^{2 \times 2}$ as the norm induced by $|\cdot|$, or any matrix norm equivalent to it.

For all $\omega \in \Omega$, let now $\mathbf{A}_{\min}(\omega)$ be such that

$$\mathbf{A}(\omega, x)\xi \cdot \xi \gtrsim \mathbf{A}_{\min}(\omega)|\xi|^2, \quad \forall \xi \in \mathbb{R}^2, \quad \text{uniformly in } x \in D,$$

and define

$$\mathbf{A}_{\max}(\omega) := \|\mathbf{A}(\omega, \cdot)\|_{C(\overline{D}, \mathbb{R}^{2 \times 2})}.$$

We make the following assumptions on the input data:

A1. $\mathbf{A}_{\min} \geq 0$ almost surely and $1/\mathbf{A}_{\min} \in L^p(\Omega)$, for all $p \in (0, \infty)$.
A2. $\mathbf{A}_{i,j} \in C^t(\overline{D}), i, j \in \{1, 2\}$, and $\mathbf{A} \in L^p(\Omega, C^t(\overline{D}, \mathbb{R}^{2 \times 2}))$, for some $0 < t \leq 1$ and for all $p \in (0, \infty)$.
A3. $f \in L^{p_*}(\Omega, H^{t-1}(D))$, for some $p_* \in (0, \infty]$.

Here, the space $C^t(\overline{D}, \mathbb{R}^{2 \times 2})$ is the space of 2×2 matrix–valued, Hölder–continuous functions with exponent t, $H^s(D)$ is the usual fractional order Sobolev space, and $L^q(\Omega, \mathscr{B})$ denotes the space of $\mathscr{B}$-valued random fields, for which the qth moment (with respect to the measure $\mathbb{P}$) of the $\mathscr{B}$–norm is finite, see e.g (Charrier, Scheichl, and Teckentrup 2011). A space which will appear in the error analysis later is the space $L^q(\Omega, H_0^1(D))$, which denotes the space of $H_0^1(D)$–valued random fields with the norm on $H_0^1(D)$ being the usual $H^1(D)$–seminorm $|\cdot|_{H^1(D)}$. It is possible to weaken Assumptions A1 and A2 to $1/\mathbf{A}_{\min}$ and $\|\mathbf{A}\|_{C^t(\overline{D}, \mathbb{R}^{2 \times 2})}$ having finite moments of order p_a, for some $p_a \in (0, \infty)$, but we will not do this here for ease of presentation.

An example of a random tensor $\mathbf{A}(\omega, x)$ that satisfies Assumptions A1 and A2, for all $p \in (0, \infty)$, is a tensor of the form $\mathbf{A} = \exp(g_1)K_1 + \exp(g_2)K_2$, where g_1 and g_2 are real–valued Gaussian random fields with a Hölder–continuous mean and a Lipschitz continuous covariance function, and K_1 and K_2 are deterministic tensors satisfying (deterministic versions of) assumptions A1–A2. For example, $g_i, i = 1, 2$, could have constant mean and an exponential covariance function, given by

$$\mathbb{E}\Big[(g_i(\omega, x) - \mathbb{E}[g_i(\omega, x)])(g_i(\omega, y) - \mathbb{E}[g_i(\omega, y)])\Big] = \sigma^2 \exp(-\|x - y\|/\lambda)$$

where σ^2 and λ are real parameters known as the *variance* and *correlation length*, and $\|\cdot\|$ denotes a norm on $\mathbb{R}^2$. It follows from the results in (Charrier, Scheichl, and Teckentrup 2011) that the resulting random tensor satisfies assumptions A1–A2, for any $t < 1/2$. If we instead choose a smoother covariance function, like the Gaussian covariance

$$\mathbb{E}\Big[(g_i(\omega, x) - \mathbb{E}[g_i(\omega, x)])(g_i(\omega, y) - \mathbb{E}[g_i(\omega, y)])\Big] = \sigma^2 \exp(-\|x - y\|^2/\lambda^2)$$

the resulting random tensor $\mathbf{A}$ satisfies assumptions A1–A2 with $t = 1$.

To simplify the notation in the following, let $0 < C_{\mathbf{A},f} < \infty$ denote a generic constant which depends algebraically on $L^q(\Omega)$–norms of $\mathbf{A}_{\max}$, $1/\mathbf{A}_{\min}$, $\|\mathbf{A}\|_{C^t(\overline{D},\mathbb{R}^{2\times2})}$ and $\|f\|_{H^{t-1}(D)}$, with $q < p*$ in the case of $\|f\|_{H^{t-1}(D)}$. We will also use the notation $b \lesssim c$ for two positive quantities b and c, if b/c is uniformly bounded by a constant independent of $\mathbf{A}$, f and h.

We will study the PDE (3) in weak (or variational) form, for fixed $\omega \in \Omega$. This is not possible uniformly in Ω, but almost surely. In the following we will not explicitly write this each time. With $f(\omega, \cdot) \in H^{t-1}(D)$ and $0 < \mathbf{A}_{\min}(\omega) \le \mathbf{A}_{\max}(\omega) < \infty$, for all $x \in D$, the variational formulation of (3), parametrized by $\omega \in \Omega$, is

$$b_\omega\big(u(\omega,\cdot),v\big) = L_\omega(v), \quad \text{for all} \quad v \in H_0^1(D), \tag{4}$$

where the bilinear form b_ω and the linear functional L_ω (both parametrized by $\omega \in \Omega$) are defined as usual, for all $u, v \in H_0^1(D)$, by

$$b_\omega(u,v) := \int_D \mathbf{A}(\omega,x)\nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x \quad \text{and} \quad L_\omega(v) := \langle f(\omega,\cdot), v\rangle_{H^{t-1}(D),H_0^{1-t}(D)}.$$

We say that for any $\omega \in \Omega$, $u(\omega,\cdot)$ is a weak solution of (3) iff $u(\omega,\cdot) \in H_0^1(D)$ and satisfies (4). The following result is classical. It is based on the Lax-Milgram Lemma (cf Hackbusch 2010).

Lemma 1 For almost all $\omega \in \Omega$, the bilinear form $b_\omega(u,v)$ is bounded and coercive in $H_0^1(D)$ with respect to $|\cdot|_{H^1(D)}$, with constants $\mathbf{A}_{\max}(\omega)$ and $\mathbf{A}_{\min}(\omega)$, respectively. Moreover, there exists a unique solution $u(\omega,\cdot) \in H_0^1(D)$ to the variational problem (4) and

$$|u(\omega,\cdot)|_{H^1(D)} \lesssim \frac{\|f(\omega,\cdot)\|_{H^{-1}(D)}}{\mathbf{A}_{\min}(\omega)}.$$

We now consider finite element approximations of our model problem (3) using standard, continuous, piecewise linear finite elements. This is not the only possible choice, and the MLMC estimator works equally well with other spatial discretizations. See for example (Cliffe et al. 2011) for results with finite volume discretizations, and (Graham, Scheichl, and Ullmann 2012) for an error analysis in the case of mixed finite elements.

Denote by $\{\mathscr{T}_h\}_{h>0}$ a shape-regular family of simplicial triangulations of the domain D, parametrized by its mesh width $h := \max_{\tau \in \mathscr{T}_h} \mathrm{diam}(\tau)$.

Associated with each triangulation $\mathscr{T}_h$ we define the space

$$V_h := \big\{v_h \in C(\overline{D}) : v_h|_\tau \text{ linear, for all } \tau \in \mathscr{T}_h, \text{ and } v_h|_\Gamma = 0\big\}$$

of continuous, piecewise linear functions on D that vanish on the boundary.

The finite element approximation of u in V_h, denoted by u_h, is now found by solving

$$b_\omega\big(u_h(\omega,\cdot),v\big) = L_\omega(v), \quad \text{for all} \quad v \in V_h,$$

The key tools in proving convergence of the finite element method are Cea's lemma and a best approximation result (cf Hackbusch 2010; Brenner and Scott 2008):

Lemma 2 (Cea's Lemma) Let Assumptions A1–A3 hold. Then, for almost all $\omega \in \Omega$,

$$|(u - u_h)(\omega,\cdot)|_{H^1(D)} \le \left(\frac{\mathbf{A}_{\max}(\omega)}{\mathbf{A}_{\min}(\omega)}\right)^{1/2} \inf_{v_h \in V_h} |u(\omega,\cdot) - v_h|_{H^1(D)}.$$

Lemma 3 Let $v \in H^{1+s}(D) \cap H_0^1(D)$, for some $0 < s \le 1$. Then

$$\inf_{v_h \in V_h} |v - v_h|_{H^1(D)} \lesssim \|v\|_{H^{1+s}(D)}\, h^s$$

where the hidden constant is independent of v and h.

In order to conclude on the the convergence of u to u_h in the $L^p(\Omega, H_0^1(D))$–norm, or in other words on the convergence of moments of the $H^1(D)$–seminorm of the error, it is hence crucial that we can bound moments of $\|u\|_{H^{1+s}(D)}$, for some $0 < s \leq 1$. The spatial regularity of u depends both on the regularity of $\mathbf{A}$ and f, and on the geometry of the domain D, so we need the following definition in addition to assumptions A1–A3.

Definition 1 *Let $0 < \lambda_\Delta(D) \leq 1$ be such that for any $0 < s \leq \lambda_\Delta(D), s \neq \frac{1}{2}$, the Laplace operator Δ is surjective as an operator from $H^{1+s}(D) \cap H_0^1(D)$ to $H^{s-1}(D)$. In other words, let $\lambda_\Delta(D)$ be no larger than the order of the strongest singularity of the Laplace operator with homogeneous Dirichlet boundary conditions on D.*

In general, the value of $\lambda_\Delta(D)$ depends on the geometry of D, and the type of boundary conditions imposed. For the particular model problem (3), we have that $\lambda_\Delta(D) = 1$ for convex domains. For non-convex domains, we have $\lambda_\Delta(D) = \min_{j=1}^m \pi/\theta_j$, where θ_j is the angle at corner S_j, and m is the number of corners in D. Hence, $\lambda_\Delta(D) > 1/2$ for any Lipschitz polygonal domain.

In the particular case of scalar coefficients, the following result was proven in (Teckentrup et al. 2012).

Theorem 2 Suppose $\mathbf{A} = a\mathbf{I}_2$, for some $a : \Omega \times D \to \mathbb{R}$, and let Assumptions A1-A3 hold for some $0 < t \leq 1$. Then,

$$\|u(\omega, \cdot)\|_{H^{1+s}(D)} \lesssim \frac{\mathbf{A}_{\max}(\omega)\|\mathbf{A}(\omega, \cdot)\|_{C^t(\overline{D}, \mathbb{R}^{2\times 2})}^2}{\mathbf{A}_{\min}(\omega)^4} \|f\|_{H^{t-1}(D)},$$

for almost all $\omega \in \Omega$ and for all $0 < s < t$ such that $s \leq \lambda_\Delta(D)$. Hence,

$$\|u - u_h\|_{L^p(\Omega, H_0^1(D))} \leq C_{\mathbf{A}, f}\, h^s, \qquad \text{for all } p < p_*,$$

with $C_{\mathbf{A}, f} < \infty$ a constant that depends on the input data, but is independent of h. If A1-A3 hold with $t = \lambda_\Delta(D) = 1$, then $\|u - u_h\|_{L^p(\Omega, H_0^1(D))} \leq C_{\mathbf{A}, f}\, h$.

From Theorem2, one can easily deduce convergence rates α and β for Theorem 1, for $Q = |u|_{H^1(D)}$. Assume $p_* > 2$, and $0 < t < 1$. Using the reverse triangle inequality, we get

$$\mathbb{E}\left[\left||u|_{H^1(D)} - |u_{h_\ell}|_{H^1(D)}\right|\right] \leq \mathbb{E}\left[|u - u_{h_\ell}|_{H^1(D)}\right] = \|u - u_{h_\ell}\|_{L^1(\Omega, H_0^1(D))} \lesssim C_{\mathbf{A}, f}\, h_\ell^s, \tag{5}$$

and so $\alpha = s$. Similarly, using $\mathbb{V}(X) \leq \mathbb{E}[X^2]$, the reverse triangle inequality, the triangle inequality and $\mathbb{E}[X^2] = \|X\|_{L^2(\Omega)}^2$, we have

$$\mathbb{V}\left[\left||u_{h_\ell}|_{H^1(D)} - |u_{h_{\ell-1}}|_{H^1(D)}\right|\right] \leq \mathbb{E}\left[\left(|u_{h_\ell} - u_{h_{\ell-1}}|_{H^1(D)}\right)^2\right] \lesssim C_{\mathbf{A}, f}\, h_\ell^{2s}, \tag{6}$$

and so $\beta = 2s$. If $t = 1$, one can similarly show that assumptions M1–M2 are satisfied with $\alpha = 1$ and $\beta = 2$.

As in the deterministic setting, one can use Theorem2, together with a duality argument, to prove convergence of the finite element error for other quantities of interest. These quantities include $\|u\|_{L^2(D)}$, for which one can prove convergence rates twice those of the $H^1(D)$–seminorm (see Charrier, Scheichl, and Teckentrup 2011), and all functionals which are continuously Fréchet differentiable, for which one can prove convergence rates up to twice those of the $H^1(D)$–seminorm, depending on the functional (see Teckentrup et al. 2012).

The remainder of this section will be devoted to extending the theory above. In §3.1, we prove an analogue of Theorem2 in the case of more general tensor coefficients $\mathbf{A}$. In §3.2, we prove convergence of a functional that does not fit into the framework of functionals covered in (Teckentrup et al. 2012), namely point evaluations.

3.1 Regularity of the Solution for Tensor Coefficients

The main result in this section is that Theorem2 holds also for more general tensor coefficients.

Theorem 3 Let Assumptions A1-A3 hold for some $0 < t \leq 1$. Then,

$$\|u(\omega,\cdot)\|_{H^{1+s}(D)} \lesssim \frac{\mathbf{A}_{\max}(\omega)\|\mathbf{A}(\omega,\cdot)\|^2_{C^t(\overline{D},\mathbb{R}^{2\times2})}}{\mathbf{A}_{\min}(\omega)^4}\|f\|_{H^{t-1}(D)},$$

for almost all $\omega \in \Omega$ and for all $0 < s < t$ such that $s \leq \lambda_\Delta(D)$. Hence, $u \in L^p(\Omega, H^{1+s}(D))$, for any $p < p_*$. If A1-A3 hold with $t = \lambda_\Delta(D) = 1$, then the above bound holds with $s = 1$, and $u \in L^p(\Omega, H^2(D))$.

The proof of Theorem3 is very similar to the proof in the case of scalar coefficients, which can be found in full in (Teckentrup et al. 2012, §5) and (Charrier, Scheichl, and Teckentrup 2011, §A). We will therefore only give the final result, together with the main ideas of the proof, in this section. For a detailed proof, see (Teckentrup 2013).

Proof of Theorem3.   (Sketch) The proof follows closely that of (Teckentrup et al. 2012, §5.1). We denote by A_ω the differential operator $-\text{div}(\mathbf{A}\nabla\cdot)$. From a result in perturbation theory, it suffices to show that there exists a constant $C_{\text{semi}}(\omega)$ such that

$$\|v\|_{H^{1+s}(D)} \leq C_{\text{semi}}(\omega)\|A_\omega v\|_{H^{s-1}(D)}, \quad \text{for all } v \in H^{1+s}(D) \cap H^1_0(D), \tag{7}$$

in order to conclude that $u(\omega,\cdot) \in H^{1+s}(D)$. To prove the existence of such a constant $C_{\text{semi}}(\omega)$, we combine regularity results for operators with constant coefficients in polygonal domains, with regularity results for operators with variable coefficients in smooth domains.

We first choose a smooth (C^2) domain $D' \subset D$, which roughly speaking coincides with D away from the corners, and does not contain any of the corners. A slight generalization of the proof in (Charrier, Scheichl, and Teckentrup 2011, §A), establishes a bound as in (7) for all functions $w \in H^{1+s}(D') \cap H^1_0(D')$.

Secondly, in order to characterize the behavior of the function v near the corners, we let W be a polygonal subdomain of D, which includes some corner S_j. We first show that

$$\mathbf{A}_{\min}(\omega)\|w\|_{H^{1+s}(W)} \lesssim \|A^j_\omega w\|_{H^{s-1}(W)}, \quad \text{for all } w \in H^{1+s}(W) \cap H^1_0(W),$$

where A^j_ω is the operator A_ω, with coefficients frozen at the corner S_j. This is done by using the Cauchy–Schwartz and the Poincaré inequalities, together with the definition of $\mathbf{A}_{\min}(\omega)$.

Using the triangle inequality, we then have

$$\mathbf{A}_{\min}(\omega)\|w\|_{H^{1+s}(W)} \lesssim \left(\|A_\omega w\|_{H^{s-1}(W)} + \|A_\omega w - A^j_\omega w\|_{H^{s-1}(W)}\right),$$

and so the crucial step is now to bound $\|A_\omega w - A^j_\omega w\|_{H^{s-1}(W)}$. This is done by showing that this difference can be bounded in terms of $\|\mathbf{A} - \mathbf{A}(S_j)\|_{C(\overline{W},\mathbb{R}^{2\times2})}$ and $\|\mathbf{A}\|_{C^t(\overline{W},\mathbb{R}^{2\times2})}$, which, by our regularity assumption A2, can be made arbitrarily small by making W arbitrarily small. This establishes a bound as in (7) for functions $w \in H^{1+s}(W) \cap H^1_0(W)$.

The final estimate (7) can then be deduced by combining the two results with the help of suitable cut–off functions. The final result is that (7) holds with

$$C_{\text{semi}}(\omega) = \frac{\mathbf{A}_{\max}(\omega)\|\mathbf{A}(\omega,\cdot)\|^2_{C^t(\overline{D},\mathbb{R}^{2\times2})}}{\mathbf{A}_{\min}(\omega)^4}.$$

$\square$

3.2 Convergence of Point Evaluations

The aim of this section is to derive bounds on moments of $\|(u-u_h)(\omega,\cdot)\|_{L^\infty(D)}$ and $\|(u-u_h)(\omega,\cdot)\|_{W^{1,\infty}(D)}$. This will give us convergence rates of the finite element error for point evaluations of the pressure u and the Darcy flux $-\mathbf{A}\nabla u$. A classical method used to derive these estimates, is the method of weighted Sobolev spaces by Nitsche. The results presented in this section are specific to continuous, linear finite elements on triangles, but extensions to higher spatial dimensions and/or higher order elements can be proved in a similar way (see e.g. Ciarlet (1978)).

The main result is the following theorem. A detailed proof can again be found in (Teckentrup 2013).

Theorem 4 Assume $u \in H_0^1(D) \cap C^r(\overline{D})$, for some $0 < r \le 2$. Then

$$\|(u-u_h)(\omega,\cdot)\|_{L^\infty(D)} \lesssim \frac{\mathbf{A}_{\max}(\omega)}{\mathbf{A}_{\min}(\omega)} \, h^r \, |\ln h| \, \|u(\omega,\cdot)\|_{C^r(\overline{D})}.$$

If $1 < r \le 2$, we furthermore have

$$|(u-u_h)(\omega,\cdot)|_{W^{1,\infty}(D)} \lesssim \frac{\mathbf{A}_{\max}(\omega)}{\mathbf{A}_{\min}(\omega)} \, h^{r-1} \, |\ln h| \, \|u(\omega,\cdot)\|_{C^r(\overline{D})}.$$

Proof. (Sketch) Using the method of weighted norms by Nitsche, as is done in for example (Ciarlet 1978, §3.3), one can derive the quasi–optimality result

$$\|(u-u_h)(\omega,\cdot)\|_{L^\infty(D)} + h\,|(u-u_h)(\omega,\cdot)|_{W^{1,\infty}(D)} \lesssim$$
$$\frac{\mathbf{A}_{\max}(\omega)}{\mathbf{A}_{\min}(\omega)} \inf_{v_h \in V_h} \left(\|u(\omega,\cdot)-v_h\|_{L^\infty(D)} + h\,|\ln h|\,|u(\omega,\cdot)-v_h|_{W^{1,\infty}(D)} \right),$$

which holds for any h sufficiently small, and where again the dependence on $\mathbf{A}$ has been made explicit. The claim of the proposition then follows from the best approximation result

$$\inf_{v_h \in V_h} \|u(\omega,\cdot)-v_h\|_{L^\infty(D)} \lesssim h^r \|u(\omega,\cdot)\|_{C^r(\overline{D})},$$

which can be found in e.g. (Schatz 1980), and holds for all $0 < r \le 2$. $\qquad\square$

In order to conclude on the convergence of moments of $\|(u-u_h)(\omega,\cdot)\|_{L^\infty(D)}$ and $|(u-u_h)(\omega,\cdot)|_{W^{1,\infty}(D)}$, it remains to prove a bound on moments of $\|u(\omega,\cdot)\|_{C^r(\overline{D})}$, for some $0 < r \le 2$. One way to achieve this is to use the Sobolev Embedding Theorem (see e.g. Ciarlet 1978, §3.1). We know from Theorem 3 that $u(\omega,\cdot) \in H^{1+s}(D)$, for some $0 < s \le 1$, which gives the following convergence rate.

Theorem 5 Let Assumptions A1–A3 be satisfied, for some $0 < t \le 1$, and let $0 < s \le t$ be such that $u \in L^p(\Omega, H^{1+s}(D))$, for all $p < p_*$. Then

$$\|u-u_h\|_{L^p(\Omega,L^\infty(D))} \lesssim C_{\mathbf{A},f} \, h^s \, |\ln h|,$$

for all $p < p_*$, with $C_{\mathbf{A},f}$ a finite constant dependent on $\mathbf{A}$ and f, but independent of h and u.

Proof. It follows from the Sobolev Embedding Theorem that $H^{1+s}(D)$ is continuously embedded into $C^s(\overline{D})$, for any $0 < s \le 1$. The claim of the Theorem then follows directly from Theorem 4. $\qquad\square$

Alternatively, one can use regularity theory in Hölder spaces to derive a bound on $\|u(\omega,\cdot)\|_{C^r(\overline{D})}$ directly, without going through the Sobolev Embedding Theorem. Theorems 8.33 and 8.34 in (Gilbarg and Trudinger 2001) give the following.

Theorem 6 Let Assumptions A1–A3 be satisfied, for some $0 < t \leq 1$ and $p_* > d/(1-t)$, and suppose D is a C^{1+t} domain. Then $u(\omega, \cdot) \in C^{1+t}(\overline{D})$, and

$$\|u(\omega, \cdot)\|_{C^{1+t}(\overline{D})} \leq C_{\text{reg}} \left(\|u(\omega, \cdot)\|_{C(\overline{D})} + \|f(\omega, \cdot)\|_{L^{p*}(D)} \right), \tag{8}$$

where the constant C_{reg} depends on $\mathbf{A}_{\min}(\omega), \mathbf{A}_{\max}(\omega)$ and $\|\mathbf{A}(\omega, \cdot)\|_{C^t(\overline{D}, \mathbb{R}^{2 \times 2})}$.

A similar result can be proved for polygonal domains D, where the regularity of u will again depend on t and the angles in D. In particular, $u(\omega, \cdot) \in C^{1+t}(\overline{D})$ for any rectangular domain (Grisvard 1985, Theorem 6.2.10).

Theorem6 suggests that $\|u - u_h\|_{L^p(\Omega, L^\infty(D))}$ and $\|u - u_h\|_{L^p(\Omega, W^{1,\infty}(D))}$ should converge with h^{1+t} and h^t, respectively. These rates are better than those proved in Theorem5, which gave a lower convergence rate for the $L^\infty(D)$–norm and no convergence for the $W^{1,\infty}(D)$–norm. To be able to conclude rigorously on these convergence rates, we would, as in Theorem3, have to know exactly how the constant C_{reg} depends on $\mathbf{A}_{\min}(\omega), \mathbf{A}_{\max}(\omega)$ and $\|\mathbf{A}(\omega, \cdot)\|_{C^t(\overline{D}, \mathbb{R}^{2 \times 2})}$. Theorem6 does, however, allow us to conclude on these higher convergence rates path wise (i.e. for almost all $\omega \in \Omega$, as in Theorem4).

The results in this section can be used in the same way as in (5) and (6) to prove convergence rates α and β in Theorem1 for point evaluations. Using the fact that $u(\omega, \cdot) \in C^1(\overline{D})$ (cf Theorem6) for almost all ω, we for example have for evaluations of the norm of the Darcy flux $-\mathbf{A}\nabla u$ at a point $x^* \in D$

$$\mathbb{E}\left[\left| |\mathbf{A}\nabla u(x^*)| - |\mathbf{A}\nabla u_{h_\ell}(x^*)| \right| \right] \leq \mathbb{E}\left[|\mathbf{A}(x^*)| \, |(\nabla u - \nabla u_{h_\ell})(x^*)| \right] \leq \mathbb{E}\left[\mathbf{A}_{\max} |u - u_{h_\ell}|_{W^{1,\infty}(D)} \right] \lesssim C_{\mathbf{A},f} \, h_\ell^\alpha,$$

where $\alpha = t$, if we use the rates suggested by Theorem6. Similarly, we have

$$\mathbb{V}\left[\left| |\mathbf{A}\nabla u(x^*)| - |\mathbf{A}\nabla u_{h_\ell}(x^*)| \right| \right] \leq \mathbb{E}\left[\mathbf{A}_{\max}^2 |u - u_{h_\ell}|_{W^{1,\infty}(D)}^2 \right] \lesssim C_{\mathbf{A},f} \, h_\ell^{2\alpha},$$

and so $\beta = 2\alpha$, where α is as above. This can easily be generalized to point evaluations of the Darcy flux in a given coordinate direction. The proof for point evaluations of the pressure is also similar, and leads to convergence rates $\alpha = s$ and $\beta = 2s$, if we use Theorem5, and $\alpha = 1+t$ and $\beta = 2(1+t)$, if we use the rates suggested by Theorem6.

4 CONCLUSIONS

We have considered the application of multilevel Monte Carlo methods to elliptic PDEs with random coefficients, in the important case of coefficients which are not uniformly coercive and bounded with respect to the random parameter. This includes, for example, log–normal random fields. Under minimal assumptions on the random coefficient, we have proven convergence of the multilevel Monte Carlo algorithm, together with an upper bound on its computational cost. We have shown that the convergence analysis in (Charrier, Scheichl, and Teckentrup 2011) and (Teckentrup et al. 2012) holds also in the case of more general, tensor–valued coefficients, and also for point evaluations of the pressure and the flux.

Although the analysis in this paper was presented for a specific model problem posed on a two–dimensional domain, most of the analysis holds true also in three spatial dimensions. The crucial difference is that in order to obtain the required regularity result, one would also have to consider the behavior of the solution near edges of the domain. It is of course also possible to include other boundary conditions and/or lower order terms in the differential operator, provided these are regular enough.

REFERENCES

Barth, A., C. Schwab, and N. Zollinger. 2011. "Multi–Level Monte Carlo Finite Element method for elliptic PDE's with stochastic coefficients". *Numerische Mathematik* 119 (1): 123–161.

Brandt, A., and V. Ilyin. 2003. "Multilevel Monte Carlo methods for studying large scale phenomena in fluids". *Journal of Molecular Liquids* 105 (2-3): 245–248.

Brenner, S. C., and L. R. Scott. 2008. *The Mathematical Theory of Finite Element Methods*. Third ed, Volume 15 of *Texts in Applied Mathematics*. Springer.

Charrier, J. 2012. "Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients". *SIAM Journal on Numerical Analysis* 50 (1): 216–246.

Charrier, J., R. Scheichl, and A. L. Teckentrup. 2011. "Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods". Technical Report 02/11, University of Bath. Available at http://www.bath.ac.uk/math-sci/bics/papers/.

Ciarlet, P. G. 1978. *The Finite Element Method for Elliptic Problems*. North–Holland.

Cliffe, K. A., M. B. Giles, R. Scheichl, and A. L. Teckentrup. 2011. "Multilevel Monte Carlo Methods and Applications to Elliptic PDEs with Random Coefficients". *Computing and Visualization in Science* 14 (1): 3–15.

de Marsily, G. 1986. *Quantitative Hydrogeology*. Academic Press.

de Marsily, G., F. Delay, J. Goncalves, P. Renard, V. Teles, and S. Violette. 2005. "Dealing with spatial heterogeneity". *Hydrogeology Journal* 13:161–183.

Delhomme, P. 1979. "Spatial variability and uncertainty in groundwater flow param- eters, A geostatistical approach". *Water Resources Research* 15 (2): 269–280.

Gilbarg, D., and N. S. Trudinger. 2001. *Elliptic partial differential equations of second order*. Classics in Mathematics. Berlin: Springer.

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

Gittelson, C. J. 2010. "Stochastic Galerkin discretization of the log-normal isotropic diffusion problem". *Mathematical Models & Methods in Applied Sciences* 20 (2): 237–263.

Graham, I. G., R. Scheichl, and E. Ullmann. 2012. "Finite Element Error Analysis for Mixed Formulations of Elliptic PDEs with Lognormal Coefficients". In preparation.

Grisvard, P. 1985. *Elliptic problems in non–smooth domains*. Pitman.

Hackbusch, W. 2010. *Elliptic differential equations*, Volume 18 of *Springer Series in Computational Mathematics*. Springer.

Heinrich, S. 2001. "Multilevel Monte Carlo Methods". In *Large Scale Scientific Computing. 3rd international conference*, Volume 2179 of *Lecture notes in Computer Science*, 58–67. Berlin: Springer.

Schatz, A. 1980. "A Weak Discrete Maximum Principle and Stability of the Finite Element Method in L_∞ on Plane Polygonal Domains. I". *Mathematics of Computation* 34 (149): 77–91.

Teckentrup, A. L. 2013. *Multilevel Monte Carlo methods for elliptic PDEs with random coeffcients*. Ph. D. thesis, University of Bath.

Teckentrup, A. L., R. Scheichl, M. B. Giles, and E. Ullmann. 2012. "Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients". Technical Report arXiv:1204.3476v1, arXiv.org. Available at http://arxiv.org/abs/1204.3476.

AUTHOR BIOGRAPHIES

ARETHA L. TECKENTRUP is a PhD student in the Department of Mathematical Sciences at the University of Bath, UK, under the supervision of Prof Robert Scheichl. Her research interests are in the numerical analysis of elliptic PDEs with random coefficients, with applications to multilevel Monte Carlo methods in uncertainty quantification. Her email address is a.l.teckentrup@bath.ac.uk.