Bath Institute for Complex Systems

Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and its Application to Multilevel Monte Carlo Methods

Julia Charrier, Robert Scheichl and Aretha Teckentrup

Bath Institute For Complex Systems

Preprint 2/11 (2011)

http://www.bath.ac.uk/math-sci/BICS

Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and its Application to Multilevel Monte Carlo Methods

Julia Charrier[*], Robert Scheichl[†] and Aretha L. Teckentrup[†]

Abstract

We consider a finite element approximation of elliptic partial differential equations with random coefficients. Such equations arise, for example, in uncertainty quantification in subsurface flow modelling. Models for random coefficients frequently used in these applications, such as log-normal random fields with exponential covariance, have only very limited spatial regularity, and lead to variational problems that lack uniform coercivity and boundedness with respect to the random parameter. In our analysis we overcome these challenges by a careful treatment of the model problem almost surely in the random parameter, which then enables us to prove uniform bounds on the finite element error in standard Bochner spaces. These new bounds can then be used to perform a rigorous analysis of the multilevel Monte Carlo method for these elliptic problems that lack full regularity and uniform coercivity and boundedness. To conclude, we give some numerical results that confirm the new bounds.

Keywords: PDEs with stochastic data, non-uniformly elliptic, non-uniformly bounded, lack of full regularity, log-normal coefficients, truncated Karhunen-Loève expansion.

1 Introduction

Partial differential equations (PDEs) with random coefficients are commonly used as models for physical processes in which the input data are subject to uncertainty. It is often of great importance to quantify the uncertainty in the outputs of the model, based on the information that is available on the uncertainty of the input data.

In this paper, we consider elliptic PDEs with random coefficients, as they arise for example in subsurface flow modelling (see e.g. [11], [10]). The classical equations governing a steady state, single phase subsurface flow consist of Darcy's law coupled with an imcompressibility condition. Taking into account the uncertainties in the source terms f and the permeability a of the medium, this leads to a linear, second-order elliptic PDE with random coefficient $a(\omega, x)$ and random right hand side $f(\omega, x)$, subject to appropriate boundary conditions.

Solving equations like this numerically can be challenging for several reasons. Models typically used for the coefficient $a(\omega, x)$ in applications can vary on a fine scale and have relatively large variances, meaning that in many cases we only have very limited spatial regularity. In the context of subsurface flow modelling, for example, a model frequently used for the permeability $a(\omega, x)$ is a homogeneous log-normal random field. That is $a(\omega, x) = \exp[g(\omega, x)]$, where g is a Gaussian random field. We show in §2.3 that for common choices of mean and covariance functions, in particular an exponential covariance function for g, trajectories of this type of random field are

[*]Département Mathématiques, ENS Cachan, Antenne de Bretagne, Ave. Robert Schumann, 35170 Bruz, France; and INRIA Rennes Bretagne Atlantique (Équipe SAGE). Email: julia.charrier@bretagne.ens-cachan.fr

[†]Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK. Email: r.scheichl@bath.ac.uk and a.l.teckentrup@bath.ac.uk

only Hölder continuous with exponent less than 1/2. Another difficulty sometimes associated with PDEs with random coefficients, is that the coefficients cannot be bounded uniformly in the random parameter ω. In the worst case, this leads to elliptic differential operators which are not uniformly coercive or uniformly bounded. This is for example true for log-normal random fields $a(\omega, x)$. Due to the nature of Gaussian random variables, $a(\omega, x)$ can in this case not be bounded from above or away from zero uniformly in ω. It is, however, possible to bound $a(\omega, x)$ for each fixed ω.

In this paper, we consider a finite element approximation (in space) of elliptic PDEs with random coefficients as described above, with particular focus on the cases where the coefficient $a(\omega, x)$ can not be bounded from above or away from zero uniformly in ω, and trajectories of $a(\omega, x)$ are only Hölder continuous. Indeed, if one assumes that the random coefficient $a(\omega, x)$ is sufficiently regular and can be bounded from above and away from zero uniformly in ω, the resulting variational problem is uniformly coercive and bounded, and the well-posedness of the problem and the subsequent error analysis are classical, see eg. [2, 12, 1, 3]. We here derive bounds on the finite element error in the solution in both $L^p(H_0^1(D))$ and $L^p(L^2(D))$ norms. Our error estimate crucially makes use of the observation that for each *fixed* ω, we have a uniformly coercive and bounded problem (in x). The derivation of the error estimate is then based on an elliptic regularity result for coefficients supposed to be only Hölder continuous, making the dependence of all constants on $a(\omega, x)$ explicit. We emphasise that we work in standard Bochner spaces with the usual test and trial spaces as in the deterministic setting. As such, our work builds on and complements [13, 5, 17, 9] which all are concerned with the well-posedness and numerical approximation of elliptic PDEs with infinite dimensional stochastic coefficients that are not uniformly bounded from above and below (e.g. log-normal coefficients).

Finally, applying the new finite element error bounds, we quantify the error committed in the multilevel Monte Carlo (MLMC) method. The MLMC analysis is motivated by [7], where the authors recently demonstrated numerically the effectiveness of MLMC estimators for computing moments of various quantities of interest related to the solution $u(\omega, x)$ of an elliptic PDE with log-normal random coefficients. The MLMC method is a very efficient variance reduction technique for classical Monte Carlo, especially in the context of differential equations with stochastic data and stochastic differential equations. It was first introduced by Heinrich [19] for the computation of high-dimensional, parameter-dependent integrals, and has been analysed extensively by Giles [16, 15] in the context of stochastic differential equations in mathematical finance. Under the assumptions of uniform coercivity and boundedness, as well as full regularity, convergence results for the MLMC estimator for elliptic PDEs with random coefficients have recently also been proved in [3]. Here we consider the problem without these assumptions which is of particular interest in subsurface flow applications, where log-normal coefficients are most commonly used.

The outline of the rest of this paper is as follows. In §2, we present our model problem and assumptions, and recall from [5], the main results on the error resulting from truncating the Karhunen-Loéve expansion in the context of log-normal coefficients. §3 is devoted to the establishment of the finite element error bound. We first prove a regularity result for elliptic PDEs with Hölder continuous coefficients, making explicit the dependence of the bound on ω. The full proof is given in the appendix. From this regularity result we then deduce the finite element error bound in the H^1-seminorm, before also treating the case of the L^2-norm and the contribution of the quadrature error. We also improve the bound given in [5] for the finite element error in the case of truncated KL-expansions of log-normal random fields, ensuring uniformity in K (the number of terms of the truncated expansion). In §4, we use the finite element error analysis from §3 to furnish a complete error analysis of the multilevel Monte Carlo method. To begin with, we present briefly the multilevel Monte-Carlo method for elliptic PDEs with random coefficients proposed in [7], before providing a rigorous bound for the ε-cost. Finally, in §5 we present some numerical experiments, illustrating the sharpness of some of the results.

2 Preliminaries

2.1 Notation

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a bounded Lipschitz domain $D \subset \mathbb{R}^d$, we introduce the following notation. For any $k \in \mathbb{N}$, we define on the Sobolev space $H^k(D)$ the following semi-norm and norm:

$$|v|_{H^k(D)} = \left(\int_D \sum_{|\alpha|=k} |D^\alpha v|^2 \, \mathrm{d}x \right)^{1/2} \quad \text{and} \quad \|v\|_{H^k(D)} = \left(\int_D \sum_{|\alpha|\leq k} |D^\alpha v|^2 \, \mathrm{d}x \right)^{1/2}.$$

We recall that, since D is bounded, the semi-norm $|\cdot|_{H^k(D)}$ defines a norm equivalent to the norm $\|\cdot\|_{H^k(D)}$ on the subspace $H_0^k(D)$ of $H^k(D)$. For any real $r \geq 0$, with $r \notin \mathbb{N}$, set $r = k + s$ with $k \in \mathbb{N}$ and $0 < s < 1$, and denote by $|\cdot|_{H^r(D)}$ and $\|\cdot\|_{H^r(D)}$ the Sobolev–Slobodetskii semi-norm and norm, respectively, defined for $v \in H^k(D)$ by

$$|v|_{H^r(D)} = \left(\iint_{D\times D} \sum_{|\alpha|=k} \frac{[D^\alpha v(x) - D^\alpha v(y)]^2}{|x-y|^{d+2s}} \, \mathrm{d}x \, \mathrm{d}y \right)^{1/2} \quad \text{and} \quad \|v\|_{H^r(D)} = \left(\|v\|^2_{H^k(D)} + |v|^2_{H^r(D)} \right)^{1/2}.$$

The Sobolev space $H^r(D)$ is then defined as the space of functions v in $H^k(D)$ such that the integral $|v|^2_{H^r(D)}$ is finite. For $0 < s \leq 1$, the space $H^{-s}(D)$ denotes the dual space to $H_0^s(D)$ with the dual norm.

In addition to the above Sobolev spaces, we also make use of the Hölder spaces $\mathcal{C}^t(\overline{D})$, with $0 < t < 1$, on which we define the following semi-norm and norm

$$|v|_{\mathcal{C}^t(\overline{D})} = \sup_{x,y\in\overline{D}:\, x\neq y} \frac{|v(x) - v(y)|}{|x-y|^t} \quad \text{and} \quad \|v\|_{\mathcal{C}^t(\overline{D})} = \sup_{x\in\overline{D}} |v(x)| + |v|_{\mathcal{C}^t(\overline{D})}.$$

The spaces $\mathcal{C}^0(\overline{D})$ and $\mathcal{C}^1(\overline{D})$ are as usual the spaces of continuous and continuously differentiable functions with the standard norms.

Finally, we will also require spaces of Bochner integrable functions. To this end, let $\mathcal{B}$ be a Banach space with norm $\|\cdot\|_{\mathcal{B}}$, and $v : \Omega \to \mathcal{B}$ be strongly measurable. With the norm $\|\cdot\|_{L^p(\Omega,\mathcal{B})}$ defined by

$$\|v\|_{L^p(\Omega,\mathcal{B})} = \begin{cases} \left(\int_\Omega \|v\|^p_{\mathcal{B}} \, \mathrm{d}\mathbb{P} \right)^{1/p}, & \text{for} \quad p < \infty, \\ \operatorname{esssup}_{\omega\in\Omega} \|v\|_{B}, & \text{for} \quad p = \infty, \end{cases}$$

the space $L^p(\Omega, \mathcal{B})$ is defined as the space of all strongly measurable functions on which this norm is finite. In particular, we denote by $L^p(\Omega, H_0^k(D))$ the Bochner space where the norm on $H_0^k(D)$ is chosen to be the seminorm $|\cdot|_{H^k(D)}$. For simplicity we write $L^p(\Omega)$ for $L^p(\Omega, \mathbb{R})$.

The key task in this paper is keeping track of how the constants in the bounds and estimates depend on the coefficient $a(\omega, x)$ and on the mesh size h. Hence, we will almost always be stating constants explicitly. To simplify this task we shall denote constants appearing in Theorem/Proposition/Corollary x.x by $C_{x.x}$. Constants that do not depend on $a(\omega, x)$ or h will not be explicitly stated. Instead, we will write $b \lesssim c$ for two positive quantities b and c, if b/c is uniformly bounded by a constant independent of $a(\omega, x)$ and of h.

2.2 Problem Setting

We consider the following linear elliptic partial differential equation (PDE) with random coefficients:

$$\begin{aligned} -\nabla \cdot (a(\omega, x)\nabla u(\omega, x)) &= f(\omega, x), & \text{in} \quad & D, \\ u(\omega, x) &= 0, & \text{on} \quad & \partial D, \end{aligned} \tag{2.1}$$

for almost all $\omega \in \Omega$. The differential operators $\nabla\cdot$ and ∇ are with respect to $x \in D$. Let us formally define, for all $\omega \in \Omega$,

$$a_{\min}(\omega) := \min_{x\in\overline{D}} a(\omega, x) \qquad \text{and} \qquad a_{\max}(\omega) := \max_{x\in\overline{D}} a(\omega, x). \tag{2.2}$$

We make the following assumptions on the input random field a and on the source term f:

A1. $a_{\min} \geq 0$ almost surely and $1/a_{\min} \in L^p(\Omega)$, for all $p \in (0, \infty)$.

A2. $a \in L^p(\Omega, C^t(\overline{D}))$, 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 $p_* \in (0, \infty]$.

The Hölder continuity of the trajectories of a in Assumption A2 implies that both quantities in (2.2) are well defined, that $a_{\max} \in L^p(\Omega)$ and (together with Assumption A1) that $a_{\min}(\omega) > 0$ and $a_{\max}(\omega) < \infty$, for almost all $\omega \in \Omega$. We will here not make the assumption that we can bound $a_{\min}(\omega)$ away from zero and $a_{\max}(\omega)$ away from infinity, uniformly in ω, as this is not true for log-normal fields a, for example. Many authors work with such assumptions of uniform ellipticity and boundedness. We shall instead work with the quantities $a_{\min}(\omega)$ and $a_{\max}(\omega)$ directly.

Note also that our assumptions on the (spatial) regularity of the coefficient function a are significantly weaker than what is usually assumed in the literature. Most other analyses of this problem assume at least that a has a gradient that lies in $L^\infty(D)$. As we will see below, we could even weaken Assumptions A1 and A2 and only assume that $\|a\|_{C^t(\overline{D})}$ and $1/a_{\min}$ have a finite number of bounded moments, i.e. $0 < p \leq p_a$, for some fixed $p_a > 0$, but in order not to complicate the presentation we did not choose to do this.

As usual in finite element methods, we will study the PDE (2.1) 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 $0 < a_{\min}(\omega) \leq a(\omega, x) \leq a_{\max}(\omega) < \infty$, for all $x \in D$, the variational formulation of (2.1), parametrised by $\omega \in \Omega$, is

$$b_\omega\big(u(\omega, \cdot), v\big) = L_\omega(v)\,, \quad \text{for all} \quad v \in H_0^1(D), \tag{2.3}$$

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

$$b_\omega(u, v) := \int_D a(\omega, x)\nabla u(x) \cdot \nabla v(x)\, \mathrm{d}x \quad \text{and} \quad L_\omega(v) := \langle f(\omega, \cdot), v\rangle_{H^{t-1}(D), H_0^{1-t}(D)}\,. \tag{2.4}$$

We say that for any $\omega \in \Omega$, $u(\omega, \cdot)$ is a weak solution of (2.1) iff $u(\omega, \cdot) \in H_0^1(D)$ and satisfies (2.3). The following result is classical. It is based on the Lax-Milgram Lemma [18].

Lemma 2.1. *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 $a_{\max}(\omega)$ and $a_{\min}(\omega)$, respectively. Moreover, there exists a unique solution $u(\omega, \cdot) \in H_0^1(D)$ to the variational problem (2.3) and*

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

The following proposition is a direct consequence of Lemma 2.1.

Theorem 2.2. *The weak solution u of (2.1) is unique and belongs to $L^p(\Omega, H_0^1(D))$, for all $p < p_*$.*

Proof. First note that $u : \Omega \to H_0^1(D)$ is measurable, since u is a continuous function of a. The result then follows directly from Lemma 2.1, Assumptions A1 and A3 and from the Hölder inequality. □

However, as usual in finite element methods we will require more (spatial) regularity of the solution u to be able to show convergence. We will come back to this in Section 3.1. First let us look at some typical examples of input random fields used in applications.

2.3 Log-normal Random Fields

A coefficient $a(\omega, x)$ of particular interest in applications of (2.1) is a log-normal random field, where $a(\omega, x) = \exp[g(\omega, x)]$ with $g : \Omega \times \overline{D} \to \mathbb{R}$ denoting a Gaussian field. We consider only mean zero homogeneous Gaussian fields with Lipschitz continuous covariance kernel

$$C(x, y) = \mathbb{E}\Big[(g(\omega, x)-\mathbb{E}[g(\omega, x)])(g(\omega, y)-\mathbb{E}[g(\omega, y)])\Big] = k\big(\|x-y\|\big), \text{ for some } k \in \mathcal{C}^{0,1}(\mathbb{R}^+) \quad (2.5)$$

and for some norm $\|\cdot\|$ in $\mathbb{R}^d$. In particular, this may be the usual modulus $\|x\| = |x| := (x^T x)^{1/2}$ or the 1-norm $\|x\| = \|x\|_1 := \sum_{i=1}^d |x_i|$. A typical example of a covariance function used in practice that is only Lipschitz continuous is the exponential covariance function given by

$$k(r) = \sigma^2 \exp(-r/\lambda), \quad (2.6)$$

for some parameters σ^2 (*variance*) and λ (*correlation length*).

With this type of covariance function, it follows from Kolmogorov's Theorem [8] that, for all $t < 1/2$, the trajectories of g belong to $\mathcal{C}^t(\overline{D})$ almost surely. More precisely, Kolmogorov's Theorem ensures the existence of a version $\tilde{g}$ of g (i.e. for any $x \in D$, we have $g(\cdot, x) = \tilde{g}(\cdot, x)$ almost surely) such that $\tilde{g}(\omega, \cdot) \in \mathcal{C}^t(\overline{D})$, for almost all $\omega \in \Omega$. In particular, we have for almost all ω, that $g(\omega, \cdot) = \tilde{g}(\omega, \cdot)$ almost everywhere. We will identify g with $\tilde{g}$ in what follows.

Built on the Hölder continuity of the trajectories of g and using Fernique's Theorem [8], it was shown in [5] that Assumption A1 holds and that $a \in L^p(\Omega, C^0(\overline{D}))$, for all $p \in (0, \infty)$.

Lemma 2.3. *Let g be Gaussian with covariance* (2.5). *Then the trajectories of the log-normal field $a = \exp g$ belong to $\mathcal{C}^t(\overline{D})$ almost surely, for all $t < r$, and*

$$\|a(\omega)\|_{\mathcal{C}^t} \;\le\; \Big(1 + 2\,|g(\omega)|_{\mathcal{C}^t}\Big)\, a_{\max}(\omega)\,.$$

Proof. Fix $\omega \in \Omega$ and $t < 1/2$. Since the trajectories of g belong to $\mathcal{C}^t(\overline{D})$ almost surely, we have

$$|e^{g(\omega,x)} - e^{g(\omega,y)}| \;\le\; |g(\omega, x) - g(\omega, y)| \Big(e^{g(\omega,x)} + e^{g(\omega,y)}\Big) \;\le\; 2\, a_{\max}(\omega)\, |g(\omega)|_{\mathcal{C}^t}\, |x-y|^t.$$

for any $x, y \in D$. Now, $a_{\max}(\omega)\, |g(\omega)|_{\mathcal{C}^t} < \infty$ almost surely, and so the result follows by taking the supremum over all $x, y \in D$. □

Lemma 2.3 can in fact be generalised from the exponential function to any smooth function of g.

Proposition 2.4. *Let g be a mean zero Gaussian field with covariance* (2.5). *Then Assumptions A1–A2 are satisfied for the log-normal field $a = \exp g$ with any $t < \frac{1}{2}$.*

Proof. Clearly by definition $a_{\min} \ge 0$. The proof that $1/a_{\min} \in L^p(\Omega)$, for all $p \in (0, \infty)$, is based on an application of Fernique's Theorem [8] and can be found in [5, Proposition 2.2]. To prove Assumption A2 note that, for all $t < 1/2$ and $p \in (0, \infty)$, $g \in L^p(\Omega, \mathcal{C}^t(\overline{D}))$ (cf. [5, Proposition 3.5]) and $a_{\max} \in L^p(\Omega)$ (cf. [5, Proposition 2.2]). Thus the result follows from Lemma 2.3 and an application of Hölder's inequality. □

Smoother covariance functions, such as the Gaussian covariance kernel

$$k(r) = \sigma^2 \exp(-r^2/\lambda^2), \quad (2.7)$$

which is analytic on $\overline{D} \times \overline{D}$, or more generally the covariance functions in the Matérn class with $\nu > 1$, all lead to $g \in C^1(\overline{D})$ and thus Assumption A2 is satisfied for all $t \le 1$.

Remark 2.5. The results in this section can be extended to log-normal random fields for which the underlying Gaussian field $g(\omega, x)$ does not have mean zero, under the assumption that this mean is sufficiently regular. Adding a mean $c(x)$ to g, we have $a(\omega, x) = \exp[c(x)] \exp[g(\omega, x)]$, and assumptions A1-A2 can still be satisfied. In particular, the assumptions still hold if $c(x) \in \mathcal{C}^{1/2}(\overline{D})$.

2.4 Truncated Karhunen-Loève Expansions

A starting point for many numerical schemes for PDEs with random coefficients is the approximation of the random field $a(\omega, x)$ as a function of a finite number of random variables, $a(\omega, x) \approx a(\xi_1(\omega), ..., \xi_K(\omega), x)$. This is true, for example, for the stochastic collocation method discussed in Section 3.4. Sampling methods, such as Monte–Carlo type methods discussed in Section 4, do not rely on such a finite–dimensional approximation as such, but may make use of such approximations as a way of producing samples of the input random field.

A popular choice to achieve good approximations of this kind for log-normal random fields is the truncated Karhunen-Loève (KL) expansion. For the random field g (as defined in Section 2.2), the KL–expansion is an expansion in terms of a countable set of independent, standard Gaussian random variables $\{\xi_n\}_{n\in\mathbb{N}}$. It is given by

$$g(\omega, x) = \sum_{n=1}^{\infty} \sqrt{\theta_n} b_n(x) \xi_n(\omega),$$

where $\{\theta_n\}_{n\in\mathbb{N}}$ are the eigenvalues and $\{b_n\}_{n\in\mathbb{N}}$ the corresponding normalised eigenfunctions of the covariance operator with kernel function $C(x, y)$ defined in (2.5). For more details on its derivation and properties, see e.g. [14]. We will here only mention that the eigenvalues $\{\theta_n\}_{n\in\mathbb{N}}$ are all non–negative with $\sum_{n\geq 0} \theta_n < +\infty$.

We shall write $a(\omega, x)$ as

$$a(\omega, x) = \exp\left[\sum_{n=1}^{\infty} \sqrt{\theta_n} b_n(x) \xi_n(\omega)\right],$$

and denote the random fields resulting from truncated expansions by

$$g_K(\omega, x) := \sum_{n=1}^{K} \sqrt{\theta_n} b_n(x) \xi_n(\omega) \quad \text{and } a_K(\omega, x) := \exp\left[\sum_{n=1}^{K} \sqrt{\theta_n} b_n(x) \xi_n(\omega)\right], \quad \text{for some } K \in \mathbb{N}.$$

The advantage of writing $a(\omega, x)$ in this way is that it gives an expression for $a(\omega, x)$ in terms of independent, standard Gaussian random variables.

Finally, we denote by u_K the solution to our model problem with the coefficient replaced by its truncated approximation,

$$-\nabla \cdot (a_K(\omega, x) \nabla u_K(\omega, x)) = f(\omega, x), \tag{2.8}$$

subject to homogeneous Dirichlet boundary conditions $u_K = 0$ on ∂D.

Under certain additional assumptions on the eigenvalues and eigenfunctions of the covariance operator, it is possible to bound the truncation error $\|u - u_K\|_{L^p(\Omega, H_0^1(D))}$, for all $p < p_*$, as shown in [5]. We make the following assumptions on $(\theta_n, b_n)_{n\geq 0}$:

B1. The eigenfunctions are continuously differentiable, i.e. $b_n \in \mathcal{C}^1(\overline{D})$ for any $n \geq 0$.

B2. There exists an $r \in (0, 1)$ such that,

$$\sum_{n\geq 0} \theta_n \|b_n\|_{L^\infty(D)}^{2(1-r)} \|\nabla b_n\|_{L^\infty(D)}^{2r} < +\infty.$$

Proposition 2.6. *Let Assumptions B1–B2 be satisfied, for some $r \in (0, 1)$. Then Assumptions A1–A2 are satisfied also for the truncated KL–expansion a_K of a, for any $K \in \mathbb{N}$ and $t < r$. Moreover, $\|a_K\|_{L^p(\Omega, \mathcal{C}^t(\bar{D}))}$ and $\|1/a_K^{min}\|_{L^p(\Omega)}$ can be bounded independently of K. If, moreover, Assumption B2 is satisfied for $\frac{\partial b_n}{\partial x_i}$ instead of b_n, for all $i = 1, \ldots, d$, then $\|a_K\|_{L^p(\Omega, \mathcal{C}^1(\bar{D}))}$ can be bounded independently of K.*

Proof. This is essentially [5, Propositions 3.7 and 3.8]. The Hölder continuity $a_K \in L^p(\Omega, C^t(\overline{D}))$, for all $p \in (0,\infty)$, can be deduced as in Lemma 2.3 from the almost sure Hölder continuity of the trajectories of g_K proved in [5, Proposition 3.5]. □

Let us recall the main result on the strong convergence of u to u_K from [5, Theorem 4.2].

Theorem 2.7. *Let Assumptions B1–B2 be satisfied, for some $r \in (0,1)$. Then, u_K converges, for all $p < p_*$, to $u \in L^p(\Omega, H_0^1(D))$. Moreover, for any $s \in [0,1]$, we have*

$$\|u - u_K\|_{L^p(\Omega,H_0^1(D))} \lesssim \underbrace{\left(\sum_{n>K} \theta_n \|b_n\|_{L^\infty(D)}^{2(1-r)} \|\nabla b_n\|_{L^\infty(D)}^{2r}\right)^{1/2}}_{=:\, C_{2.7}(K)} \|f\|_{L^p(\Omega,H^{s-1}(D))}$$

The hidden constant depends only on D, r, p. Similarly $\|u-u_K\|_{L^p(\Omega,L^2(D))} \lesssim C_{2.7}\,\|f\|_{L^p(\Omega,H^{s-1}(D))}$.

Assumptions B1–B2 are fulfilled, among other cases, for the analytic covariance function (2.7) as well as for the exponential covariance function (2.6) with 1-norm $\|x\| = \sum_{i=1}^d |x_i|$ in (2.5), since then the eigenvalues and eigenvectors can be computed explicitly and we have explicit decay rates for the KL–eigenvalues. For details see [5, Section 7]. In the latter case, on non-rectangular domains D, we need to use a KL-expansion on a bounding box containing D to get again explicit formulae for the eigenvalues and eigenvectors. Strictly speaking this is not a KL–expansion on D.

Proposition 2.8. *We have the following bound on the constant in Theorem 2.7:*

$$C_{2.7}(K) \;\lesssim\; \begin{cases} K^{\frac{\rho-1}{2}}, & \text{for the 1-norm exponential covariance kernel,} \\ K^{\frac{d-1}{2d}} \exp\big(-c_1\, K^{1/d}\big), & \text{for the Gaussian covariance kernel,} \end{cases}$$

for some constant $c_1 > 0$ and for any $0 < \rho < 1$. The hidden constants depend only on D, ρ, p.

3 Finite Element Error Analysis

Let us now come to the central part of this paper. To carry out a finite element error analysis for (2.1) under Assumptions A1-A3, we will require first of all regularity results for trajectories of the weak solution u. However, since we did not assume uniform ellipticity or boundedness of $b_\omega(\cdot,\cdot)$, it is essential that we track exactly, how the constants in the regularity estimates depend on the input random field a. We were unable to find such a result in the literature, and we will therefore in Section 3.1 reiterate the steps in the usual regularity proof for the solution $u(\omega,\cdot)$ of (2.3) following the proof in Hackbusch [18], but highlighting explicitly where and how constants depend on a. A detailed proof is given in Appendix A. We will need to assume that $D \subset \mathbb{R}^d$ is a $\mathcal{C}^2$ bounded domain for this proof. Regularity proofs that require only Lipschitz continuity for the boundary of D do exist (cf. Hackbusch [18, Theorems 9.1.21 and 9.1.22]), but we did not consider it instructive to complicate the presentation even further. We do not see any obstacles in also getting an explicit dependence of the constants in the case of Lipschitz continuous boundaries.

Having established the regularity of trajectories of u, we then carry out a classical finite element error analysis for each trajectory in Section 3.2 and deduce error estimates for the moments of the error in u. Since we only have a regularity result for $\mathcal{C}^2$ bounded domains and since the integrals appearing in $b_\omega(\cdot,\cdot)$ can in general only be approximated by quadrature, we will also need to address variational crimes, such as the approximation of a $\mathcal{C}^2$ bounded domain by a polygonal domain as well as quadrature errors (cf. Section 3.3).

Let us assume for the rest of this section that $D \subset \mathbb{R}^d$ is a $\mathcal{C}^2$ bounded domain.

3.1 Regularity of the Solution

Proposition 3.1. *Let Assumptions A1-A3 hold with* $0 < t < 1$*. Then,* $u(\omega, \cdot) \in H^{1+s}(D)$*, for all* $0 < s < t$ *except* $s = 1/2$ *almost surely in* $\omega \in \Omega$*, and*

$$\|u(\omega,\cdot)\|_{H^{1+s}(D)} \lesssim C_{3.1}(\omega)\, \|f(\omega,\cdot)\|_{H^{s-1}(D)}, \quad \text{where} \quad C_{3.1}(\omega) := \frac{a_{\max}(\omega)\|a(\omega,\cdot)\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}(\omega)^3}.$$

If the assumptions hold with $t = 1$*, then* $u(\omega,\cdot) \in H^2(D)$ *and* $\|u(\omega,\cdot)\|_{H^2(D)} \lesssim C_{3.1}(\omega)\, \|f(\omega,\cdot)\|_{L^2(D)}$.

We give here only the main elements of the proof and consider only the case $t < 1$ in detail. It follows the proof of [18, Theorem 9.1.16] and consists in three main steps. We formulate the first two steps as separate lemmas and then give the final step following the lemmas. We fix $\omega \in \Omega$ and to simplify the notation we will not specify the dependence on ω anywhere in the proof.

In the first step of the proof we take $D = \mathbb{R}^d$ and establish the regularity of a slightly more general elliptic PDE with tensor-valued coefficients.

Lemma 3.2. *Let* $0 < t < 1$ *and* $D = \mathbb{R}^d$*, and let* $A = (A_{ij})_{i,j=1}^d \in S_d(\mathbb{R})$ *be a symmetric, uniformly positive definite* $n \times n$ *matrix-valued function from* D *to* $\mathbb{R}^{n\times n}$*, i.e. there exists* $A_{\min} > 0$ *such that* $A(x)\xi \cdot \xi \geq A_{\min}|\xi|^2$ *uniformly in* $x \in \overline{D}$ *and* $\xi \in \mathbb{R}^d$*, and let* $A_{ij} \in \mathcal{C}^t(\overline{D})$*, for all* $i, j = 1, \ldots, d$*. Consider*

$$-\mathrm{div}(A(x)\nabla w(x)) = F(x) \qquad \text{in} \;\; D \tag{3.1}$$

with $F \in H^{s-1}(\mathbb{R}^d)$*, for some* $0 < s < t$*. Any weak solution* $w \in H^1(\mathbb{R}^d)$ *of* (3.1) *is in* $H^{1+s}(\mathbb{R}^d)$ *and*

$$\|w\|_{H^{1+s}(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\Big) \;+\; \|w\|_{H^1(\mathbb{R}^d)},$$

where $|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}$ *is the Hölder seminorm on* $S_d(\mathbb{R})$ *using a suitable matrix norm.*

Proof. This is essentially [18, Theorem 9.1.8] with the dependence on A made explicit, and it can be proved using the representation of the norm on $H^{1+s}(\mathbb{R}^d)$ via Fourier coefficients, as well as a fractional difference operator R_h^i, $i = 1, \ldots, d$, on a Cartesian mesh with mesh size $h > 0$ (similar to the classical Nirenberg translation method for proving H^2 regularity).

It is shown in the proof of [18, Theorem 9.1.8] that $\sum_{i=1}^d \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)}$ is an upper bound for $\|w\|_{H^{1+s}(\mathbb{R}^d)}$ and that it can itself be bounded from above in terms of $\|w\|_{H^1(\mathbb{R}^d)}$ and $\|F\|_{H^{s-1}(\mathbb{R}^d)}$ using the weak form (2.3) of our problem. The dependence of this upper bound on $A_{\min}$ stems from the fact that in order to use (2.3), we have to switch from $|(R_h^i)^* w|_{H^1(\mathbb{R}^d)}$ to the energy norm $\int_{\mathbb{R}^d} A(x)|\nabla(R_h^i)^* w|^2\,\mathrm{d}x$. The dependence on $|A_{ij}|_{\mathcal{C}^t(\mathbb{R}^d)}$ comes from bounding the differences of A_{ij} at two consecutive grid points in the translation step.

For a definition of R_h^i and more details see [18, Theorem 9.1.8] or Section A.1 in the appendix. □

The second step consists in treating the case where $D = \mathbb{R}^d_+ := \{y = (y_1, ..., y_d) : y_d > 0\}$.

Lemma 3.3. *Let* $0 < t < 1$ *and* $D = \mathbb{R}^d_+$*, and let* $A : D \to S_d(\mathbb{R})$ *be as in Lemma 3.2. Consider now* (3.1) *on* $D = \mathbb{R}^d_+$ *subject to* $w = 0$ *on* ∂D *with* $F \in H^{s-1}(\mathbb{R}^d_+)$*, for some* $0 < s < t$*,* $s \neq 1/2$*. Then any weak solution* $w \in H^1(\mathbb{R}^d_+)$ *of this problem is in* $H^{1+s}(\mathbb{R}^d_+)$ *and*

$$\|w\|_{H^{1+s}(\mathbb{R}^d_+)} \lesssim \frac{A_{\max}}{A_{\min}^2}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) + \frac{A_{\max}}{A_{\min}}\,\|w\|_{H^1(\mathbb{R}^d_+)}\,.$$

where $A_{\max} := \|A\|_{\mathcal{C}^0(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}$.

Proof. This is essentially [18, Theorem 9.1.11] with the dependence on A made explicit. It uses the fact that for any $s > 0$, $s \neq 1/2$, the norm

$$|||v|||_s := \Big(\|v\|^2_{L^2(\mathbb{R}^d_+)} + \sum_{i=1}^d \Big\| \frac{\partial v}{\partial x_i} \Big\|^2_{H^{s-1}(\mathbb{R}^d_+)} \Big)^{1/2}$$

is equivalent to the usual norm on $H^s(\mathbb{R}^d_+)$. Then using the same approach as in the proof of Lemma 3.2, we can establish that, for $1 \leq j \leq d-1$,

$$\Big\| \frac{\partial w}{\partial x_j} \Big\|_{H^s(\mathbb{R}^d_+)} \lesssim \frac{1}{A_{\min}} \Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))} \, |w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)} \Big) + \|w\|_{L^2(\mathbb{R}^d_+)} . \tag{3.2}$$

To establish a similar bound for $j = d$ is technical. We use (3.1) and the following inequality, for any $D \subset \mathbb{R}^d$ and $0 < s < t < 1$:

$$\|bv\|_{H^s(D)} \lesssim |b|_{\mathcal{C}^t(\overline{D})} \|v\|_{L^2(D)} + \|b\|_{\mathcal{C}^0(\overline{D})} \|v\|_{H^s(D)}, \quad \text{for all } b \in \mathcal{C}^t(\overline{D}),\ v \in H^s(D). \tag{3.3}$$

We use (3.3) to bound the H^s–norm of $A_{ij} \frac{\partial w}{\partial x_j}$, for $(i,j) \neq (d,d)$, and so by rearranging the weak form of (3.1) we can also bound the H^s–norm of $A_{dd} \frac{\partial w}{\partial x_d}$. This leads to the additional factor $A_{\max}$ in the bound. The final result can then be deduced by applying (3.3) once more, with $b = 1/A_{dd}$ and $v = A_{dd} \frac{\partial w}{\partial x_d}$, leading to an additional factor $1/A_{\min}$. For details see [18, Theorem 9.1.11] or Section A.2 in the appendix. □

Proof of Proposition 3.1. We are now ready to prove Proposition 3.1 using Lemmas 3.2 and 3.3. The third and last step consists in using a covering of D by $m+1$ bounded regions $(D_i)_{0 \leq i \leq m}$, such that

$$\overline{D}_0 \subset D, \quad \overline{D} \subset \bigcup_{i=0}^m D_i \quad \text{and} \quad \partial D = \bigcup_{i=1}^m (D_i \cap \partial D).$$

Using a (non-negative) partition of unity $\{\chi_i\}_{0 \leq i \leq m} \subset \mathcal{C}^\infty(\mathbb{R}^d)$ subordinate to this cover, it is possible to reduce the proof to bounding $\|\chi_i u\|_{H^{1+s}(D)}$, for all $0 \leq i \leq m$.

For $i = 0$ this reduces to an application of Lemma 3.2 with w and F chosen to be extensions by 0 from D to $\mathbb{R}^d$ of $\chi_0 u$ and of $f\chi_0 + a\nabla u \cdot \nabla \chi_0 + \mathrm{div}(au\nabla\chi_0)$, respectively. The tensor A degenerates to $\overline{a}(x) I_d$, where $\overline{a}$ is a smooth extension of $a(x)$ on D_0 to $a_{\min}$ on $\mathbb{R}^d \backslash D$, and so $A_{\min} = a_{\min}$ and $|A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))} \leq |a|_{\mathcal{C}^t(D)}$.

For $1 \leq i \leq m$, the proof reduces to an application of Lemma 3.3. As for $i = 0$, we can see that $\chi_i u \in H^1_0(D \cap D_i)$ is the weak solution of the problem $-\mathrm{div}(a\nabla u_i) = f_i$ on $D \cap D_i$ with $f_i := f\chi_i + a\nabla u \cdot \nabla \chi_i + \mathrm{div}(au\nabla\chi_i)$. To be able to apply Lemma 3.3 to the weak form of this PDE, we define now a twice continuously differentiable bijection α_i (with α_i^{-1} also in $\mathcal{C}^2$) from D_i to the cylinder

$$Q_i := \{y = (y_1, \ldots, y_d) : |(y_1, \ldots, y_{d-1})| < 1 \ \text{ and } \ |y_d| < 1\},$$

such that $D_i \cap D$ is mapped to $Q_i \cap \mathbb{R}^d_+$, and $D_i \cap \partial D$ is mapped to $Q_i \cap \{y : y_d = 0\}$. We use α_i^{-1} to map all the functions defined above on $D_i \cap D$ to $Q_i \cap \mathbb{R}^d_+$, and then extend them suitably to functions on $\mathbb{R}^d_+$ to finally apply Lemma 3.3. The tensor A in this case is a genuine tensor depending on the mapping α_i. However, since ∂D was assumed to be $\mathcal{C}^2$, we get $A_{\min} \lesssim a_{\min}$, $A_{\max} \lesssim a_{\max}$ and $|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))} \lesssim |a|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}$, with hidden constants that only depend on α_i, α_i^{-1} and their Jacobians. For details see [18, Theorem 9.1.16] or Section A.3 in the appendix. □

Using Proposition 3.1 and Assumptions A1-A3, we can now conclude on the regularity of u.

Theorem 3.4. *Let Assumptions A1-A3 hold with $0 < t \le 1$. Then $u \in L^p(\Omega, H^{1+s}(D))$, for all $p < p_*$ and for all $0 < s < t$ except $s = 1/2$. If $t = 1$, then $u \in L^p(\Omega, H^2(D))$.*

Proof. Since $a_{\max}(\omega) \le \|a(\omega,\cdot)\|_{\mathcal{C}^t(\overline{D})}$ and $H^{s-1}(D) \subset H^{t-1}(D)$, for all $s < t$, the result follows directly from Proposition 3.1 and Assumptions A1–A3 via Hölder's inequality. □

Remark 3.5. Note that in order to establish $u \in L^p(\Omega, H^{1+s}(D))$, for some fixed $p > 0$, it would have been sufficient to assume that the constant $C_{3.1}$ in Proposition 3.1 is in $L^q(\Omega)$ for $q = \frac{p_* p}{p_* - p}$. In the case $p_* = \infty$, $q = p$ is sufficient. This in turn implies that we can weaken Assumption A1 to $1/a_{\min} \in L^q(\Omega)$ with $q > 3p$, or Assumption A2 to $a \in L^q(\Omega, C^t(\overline{D}))$ with $q > 2p$, or both assumptions to L^q with $q > 5p$.

However, in the case of a log-normal field a and $p_* = \infty$, we do have bounds on all moments $p \in (0, \infty)$, but in general we only have the limited spatial regularity of $1 + s < 3/2$.

3.2 Finite Element Approximation

We consider finite element approximations of our model problem (2.1) using standard, continuous, piecewise linear finite elements. The aim is to derive estimates of the finite element error in the $L^p(\Omega, H^1_0(D))$ and $L^p(\Omega, L^2(D))$ norms. To remain completely rigorous, we keep the assumption that boundary of D is $\mathcal{C}^2$, so that we can apply the explicit regularity results from the previous section. However, this means that we will have to approximate our domain D by polygonal domains D_h in dimensions $d \ge 2$.

We denote by $\{\mathcal{T}_h\}_{h>0}$ a shape-regular family of simplicial triangulations of the domain D, parametrised by its mesh width $h := \max_{\tau \in \mathcal{T}_h} \mathrm{diam}(\tau)$, such that, for any $h > 0$,

- $\overline{D} \subset \bigcup_{\tau \in \mathcal{T}_h} \tau$, i.e. the triangulation covers all of $\overline{D}$, and
- the vertices $x_1^\tau, \ldots, x_{d+1}^\tau$ of any $\tau \in \mathcal{T}_h$ lie either all in $\overline{D}$ or all in $\mathbb{R}^d \backslash D$.

Let $\overline{D}_h$ denote the union of all simplices that are interior to $\overline{D}$ and D_h its interior, so that $D_h \subset D$.

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

$$V_h := \Big\{ v_h \in C(\overline{D}) : v_h|_\tau \text{ linear, for all } \tau \in \mathcal{T}_h \text{ with } \tau \subset \overline{D}_h\,, \ \text{ and } \ v_h|_{\overline{D}\backslash D_h} = 0 \Big\} \tag{3.4}$$

of continuous, piecewise linear functions on D_h that vanish on the boundary of D_h and in $D \backslash D_h$. Let us recall the following standard interpolation result (see e.g. [4, Section 4.4]).

Lemma 3.6. *Let $v \in H^{1+s}(D_h)$, for some $0 < s \le 1$. Then*

$$\inf_{v_h \in V_h} |v - v_h|_{H^1(D_h)} \lesssim \|v\|_{H^{1+s}(D_h)}\, h^s. \tag{3.5}$$

The hidden constant is independent of h and v.

This can easily be extended to an interpolation result for functions $v \in H^{1+s}(D) \cap H^1_0(D)$, by estimating the residual over $D \backslash D_h$. However, when D is not convex it requires local mesh refinement in the vicinity of any non-convex parts of the boundary. We make the following assumption on $\mathcal{T}_h$:

A4 For all $\tau \in \mathcal{T}_h$ with $\tau \cap D_h = \emptyset$ and $x_1^\tau, \ldots, x_{d+1}^\tau \in \overline{D}$, we assume $\mathrm{diam}(\tau) \lesssim h^2$.

Lemma 3.7. *Let $v \in H^{1+s}(D) \cap H^1_0(D)$, for some $0 < s \le 1$, and let Assumption A4 hold. Then*

$$\inf_{v_h \in V_h} |v - v_h|_{H^1(D)} \lesssim \|v\|_{H^{1+s}(D)}\, h^s. \tag{3.6}$$

Proof. This result is classical (for parts of the proof see [18, Section 8.6] or [20]). Set $D_\delta := D\backslash\overline{D}_h$ where δ denotes the maximum width of D_δ, and let first $s=1$. Since $v_h = 0$ on D_δ it suffices to show that

$$|v|_{H^1(D_\delta)} \lesssim \|v\|_{H^2(D)}\, h\,. \tag{3.7}$$

The result then follows for $s=1$ with Lemma 3.6. The result for $s<1$ follows by interpolation, since trivially, $|v|_{H^1(D_\delta)} \le \|v\|_{H^1(D)}$,

To show (3.7), let $w \in H^1(D)$. Using a trace result we get

$$\|w\|_{L^2(D_\delta)} \le \|w\|_{L^2(S_\delta)} \lesssim \delta^{1/2}\|w\|_{H^1(D)},$$

where $S_\delta = \{x \in D : \mathrm{dist}(x, \partial D) \le \delta\} \subset D$ is the boundary layer of width δ. It follows from Assumption A4 that $\mathrm{diam}(\tau) \lesssim h^2$ wherever the boundary is not convex. In regions where D is convex it follows from the smoothness assumption on ∂D that the width of D_δ is $\mathcal{O}(h^2)$. Hence $\delta \lesssim h^2$, which completes the proof of (3.7). □

Now, for almost all $\omega \in \Omega$, the finite element approximation to (2.3), denoted by $u_h(\omega,\cdot)$, is the unique function in V_h that satisfies

$$b_\omega(u_h(\omega,\cdot), v_h) = L_\omega(v_h), \qquad \text{for all} \quad v_h \in V_h. \tag{3.8}$$

Since $b_\omega(\cdot,\cdot)$ is coercive and bounded in $H^1_0(D)$ (cf. Lemma 2.1), we have

$$|u_h(\omega,\cdot)|_{H^1(D)} \lesssim \|f(\omega,\cdot)\|_{H^{t-1}(D)}/a_{\min}(\omega). \tag{3.9}$$

as well as the following classical quasi optimality result (cf. [4, 18]).

Lemma 3.8 (Cea's Lemma). *Let Assumptions A1–A3 hold. Then, for almost all* $\omega \in \Omega$,

$$|u(\omega,\cdot) - u_h(\omega,\cdot)|_{H^1(D)} \le C_{3.8}(\omega) \inf_{v_h\in V_h} |u(\omega,\cdot) - v_h|_{H^1(D)}, \quad \textit{where} \quad C_{3.8}(\omega) := \left(\frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right)^{1/2}.$$

Combining this with the interpolation result above we get the following error estimates.

Theorem 3.9. *Let Assumptions A1–A4 hold, for some* $0<t<1$ *and* $p_* \in (0,\infty]$. *Then, for all* $p<p_*$, $s<t$ *with* $s \ne 1/2$ *and* $h>0$, *we have*

$$\|u-u_h\|_{L^p(\Omega,H^1_0(D))} \;\lesssim\; C_{3.9}\, \|f\|_{L^{p_*}(\Omega,H^{s-1}(D))}\, h^s, \quad \textit{where} \quad C_{3.9} := \left\| \frac{a_{\max}^{3/2}\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^{7/2}} \right\|_{L^q(\Omega)}$$

with $q = \frac{p_* p}{p_* - p}$. *If Assumptions A2 and A3 hold with* $t=1$, *then*

$$\|u-u_h\|_{L^p(\Omega,H^1_0(D))} \;\lesssim\; C_{3.9}\, \|f\|_{L^{p_*}(\Omega,L^2(D))}\, h\,.$$

Proof. Let $0<t<1$ in Assumptions A2 and A3. It follows from Proposition 3.1 and Lemmas 3.7 and 3.8 that, for almost all $\omega \in \Omega$,

$$|u(\omega,\cdot) - u_h(\omega,\cdot)|_{H^1(D)} \;\lesssim\; C_{3.1}(\omega)\, C_{3.8}(\omega)\, \|f(\omega,\cdot)\|_{H^{s-1}(D)}\, h^s. \tag{3.10}$$

The result now follows by Hölder's inequality. The proof for $t=1$ is analogous. □

The usual duality (or Aubin–Nitsche) "trick" leads to a bound on the L^2–error.

Corollary 3.10. *Let Assumptions A1–A4 hold, for some $0 < t < 1$ and $p_* \in (0,\infty]$. Then, for all $p < p_*$, $s < t$ with $s \neq 1/2$ and $h > 0$, we have*

$$\|u - u_h\|_{L^p(\Omega,L^2(D))} \lesssim C_{3.10} \, \|f\|_{L^{p_*}(\Omega,H^{s-1}(D))} \, h^{2s}, \quad \text{where} \quad C_{3.10} := \left\| \frac{a_{\max}^{7/2} \|a\|^2_{C^t(\overline{D})}}{a_{\min}^{13/2}} \right\|_{L^q(\Omega)}$$

with $q = \frac{p_* p}{p_* - p}$. If Assumptions A2 and A3 hold with $t = 1$, then

$$\|u - u_h\|_{L^p(\Omega,L^2(D))} \lesssim C_{3.10} \, \|f\|_{L^{p_*}(\Omega,L^2(D))} \, h^2.$$

Proof. We will use a duality argument. For almost all $\omega \in \Omega$, let $e_\omega := u(\omega,\cdot) - u_h(\omega,\cdot)$ and denote by w_ω the solution to the adjoint problem

$$b_\omega(v, w_\omega) = (e_\omega, v) \qquad \text{for all } v \in H_0^1(D),$$

which, by Proposition 3.1 with $f(\omega,\cdot) = e_\omega$, is also in $H^{1+s}(D)$. By Galerkin orthogonality

$$\|e_\omega\|^2_{L^2(D)} = (e_\omega, e_\omega)_{L^2(D)} = b_\omega(e_\omega, w_\omega - z_h), \quad \text{for any} \quad z_h \in V_h.$$

Using the boundedness of $b_\omega(\cdot,\cdot)$ and Lemma 3.7, we then get

$$\|e_\omega\|^2_{L^2(D)} \lesssim a_{\max}(\omega) \, |u(\omega,\cdot) - u_h(\omega,\cdot)|_{H^1(D)} \, \|w_\omega\|_{H^{1+s}(D)} \, h^s.$$

Now it follows from Proposition 3.1 and Theorem 3.9 that

$$\|e_\omega\|^2_{L^2(D)} \leq a_{\max}(\omega) \, C_{3.1}(\omega)^2 \, C_{3.8}(\omega) \, \|f(\omega,\cdot)\|_{H^{s-1}(D)} \|e_\omega\|_{L^2(D)} \, h^{2s}.$$

Dividing by $\|e_\omega\|_{L^2(D)}$, the result follows again by an application of Hölder's inequality. □

Remark 3.11. As usual, the analysis can be extended in a straightforward way to other boundary conditions, to tensor coefficients, or to more complicated PDEs including low–order terms, provided the boundary data and the coefficients are sufficiently regular, see [18] for details.

3.3 Quadrature Error

The integrals appearing in the bilinear form

$$b_\omega(w_h, v_h) \; = \sum_{\tau \in \mathcal{T}_h \colon \tau \in \overline{D}_h} \int_\tau a(\omega, x) \nabla w_h \cdot \nabla v_h \; \mathrm{d}x$$

and in the linear functional $L_\omega(v_h)$ involve realisations of random fields. It will in general be impossible to evaluate these integrals exactly, and so we will use quadrature instead. We will only explicitly analyse the quadrature error in b_ω, but the quadrature error in approximating $L_\omega(v_h)$ can be analysed analogously.

In our analysis, we use the midpoint rule, approximating the integrand by its value at the midpoint x_τ of each simplex $\tau \in \mathcal{T}_h$. The trapezoidal rule which we use in the numerical section can be analysed analogously. Let us denote the resulting bilinear form that approximates b_ω on the grid $\mathcal{T}_h$ by

$$\tilde{b}_\omega(w_h, v_h) = \sum_{\tau \in \mathcal{T}_h \colon \tau \in \overline{D}_h} a(\omega, x_\tau) \int_\tau \nabla w_h(x) \cdot \nabla v_h(x) \, \mathrm{d}x \,,$$

and let $\tilde{u}_h(\omega,\cdot)$ denote the corresponding solution to

$$\tilde{b}_\omega(\tilde{u}_h(\omega,\cdot), v_h) = L_\omega(v_h) \,, \quad \text{for all} \quad v_h \in V_h.$$

Clearly the bilinear form $\tilde{b}_\omega$ is bounded and coercive, with the same constants as the exact bilinear form b_ω and so we can apply the following classical result [6] (with explicit dependence of the bound on the coefficients).

Lemma 3.12 (First Strang Lemma)**.** *Let Assumptions A1-A3 hold. Then, for almost all* $\omega \in \Omega$*,*

$$|u(\omega,\cdot) - \tilde{u}_h(\omega,\cdot)|_{H^1(D)} \leq \inf_{v_h \in V_h} \left\{ \left(1 + \frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right) |u(\omega,\cdot) - v_h|_{H^1(D)} + \frac{1}{a_{\min}(\omega)} \sup_{w_h \in V_h} \frac{|b_\omega(v_h, w_h) - \tilde{b}_\omega(v_h, w_h)|}{|w_h|_{H^1(D)}} \right\}.$$

Proposition 3.13. *Let Assumptions A1–A4 hold, for some* $0 < t < 1$ *and* $p_* \in (0, \infty]$*. Then, for all* $p < p_*$*,* $s < t$ *with* $s \neq 1/2$ *and* $0 < h < 1$*, we have*

$$\|u - \tilde{u}_h\|_{L^p(\Omega, H_0^1(D))} \lesssim C_{3.13} \, \|f\|_{L^{p_*}(\Omega, H^{s-1}(D))} \, h^s, \quad \text{where} \quad C_{3.13} := \left\| \frac{a_{\max}^{5/2} \|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^{9/2}} \right\|_{L^q(\Omega)}$$

with $q = \frac{p_* p}{p_* - p}$*. If Assumptions A2 and A3 hold with* $t = 1$*, then*

$$\|u - \tilde{u}_h\|_{L^p(\Omega, H_0^1(D))} \lesssim C_{3.13} \, \|f\|_{L^{p_*}(\Omega, L^2(D))} \, h \,.$$

Proof. We first note that, for all $w_h \in V_h$,

$$\begin{aligned}
\left| b_\omega(v_h, w_h) - \tilde{b}_\omega(v_h, w_h) \right| &= \left| \sum_{\tau \in \mathcal{T}_h} \int_\tau (a(\omega, x) - a(\omega, x_\tau)) \, \nabla v_h \cdot \nabla w_h \, \mathrm{d}x \right| \\
&\leq \sum_{\tau \in \mathcal{T}_h} \int_\tau \frac{|a(\omega, x) - a(\omega, x_\tau)|}{|x - x_\tau|^t} \, |x - x_\tau|^t \, |\nabla v_h(\omega) \cdot \nabla w_h| \, \mathrm{d}x \\
&\leq |a(\omega)|_{\mathcal{C}^t(\overline{D})} \, h^t \, |v_h|_{H^1(D)} \, |w_h|_{H^1(D)} \,.
\end{aligned}$$

Hence, it follows from Lemma 3.12 that, for almost all $\omega \in \Omega$,

$$|u(\omega,\cdot) - \tilde{u}_h(\omega,\cdot)|_{H^1(D)} \leq \inf_{v_h \in V_h} \left\{ \left(1 + \frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right) |u(\omega,\cdot) - v_h|_{H^1(D)} + h^t \, \frac{|a(\omega)|_{\mathcal{C}^t(\overline{D})}}{a_{\min}(\omega)} \, |v_h|_{H^1(D)} \right\}.$$

Let us now make the particular choice $v_h := u_h(\omega, \cdot) \in V_h$, i.e. the solution of (3.8). Then it follows from (3.9) and (3.10) and the fact that $h^t < h^s$, for any $s < t \leq 1$ and $h < 1$, that

$$|u(\omega,\cdot) - \tilde{u}_h(\omega,\cdot)|_{H^1(D)} \lesssim \left(\left(1 + \frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right) C_{3.1}(\omega) \, C_{3.8}(\omega) \; + \; \frac{|a(\omega)|_{\mathcal{C}^t(\overline{D})}}{a_{\min}(\omega)^2} \right) \|f(\omega, \cdot)\|_{H^{s-1}(D)} \, h^s.$$

The result follows again via an application of Hölder's inequality. □

Remark 3.14. To recover the $\mathcal{O}(h^{2s})$ convergence for the L^2-error $\|u - \tilde{u}_h\|_{L^p(\Omega, L^2(D))}$ in the case of quadrature, we require additional regularity of the coefficient function a. If $a(\omega, \cdot)$ is at least $\mathcal{C}^{2s}$, then we can again obtain $\mathcal{O}(h^{2s})$ convergence even with quadrature, using duality as in Corollary 3.10. This is for example the case in the context of lognormal random fields with Gaussian covariance kernel, where $a(\omega, \cdot) \in \mathcal{C}^\infty(\overline{D})$. In the context of the exponential covariance kernel the L^2-convergence rate is always bounded by $\mathcal{O}(h^{1/2-\delta})$, due to the lack of regularity in a.

3.4 Truncated Fields

Combining the results from Sections 2.4 and 3.2 we can also estimate the error in the case of truncated KL-expansions of log-normal coefficients which we will use in the numerical experiments in Section 5. These results give an improvement of the error estimate given in [5] for log-normal coefficients that do not admit full regularity, e.g. those with exponential covariance kernel. The novelty with respect to [5] is the improvement of the bound for the finite element error, leading to a uniform bound with respect to the number of terms in the KL–expansion. We will assume here that $f(\omega,\cdot) \in L^2(D)$.

Let a be a log-normal field as defined in Section 2.4 via a KL expansion and let a_K be the expansion truncated after K terms. Now, for almost all $\omega \in \Omega$, let $u_{K,h}(\omega,\cdot)$ denote the unique function in V_h that satisfies

$$b_{K,\omega}(u_{K,h}(\omega,\cdot), v_h) = L_\omega(v_h), \qquad \text{for all} \quad v_h \in V_h\,, \tag{3.11}$$

where $b_{K,\omega}(u,v) := \int_D a_K(\omega,x)\nabla u \cdot \nabla v \,\mathrm{d}x$. Then we have the following result.

Theorem 3.15. *Let $f \in L^{p^*}(\Omega, L^2(D))$, for some $p^* \in (0,\infty]$. Then, in the Gaussian covariance case, there exists a $c_1 > 0$ such that, for any $p < p^*$,*

$$\|u - u_{K,h}\|_{L^p(\Omega,H_0^1(D))} \lesssim \left(h \;+\; K^{\frac{d-1}{2d}} \exp\big(-c_1\, K^{1/d}\big)\right) \|f\|_{L^{p^*}(\Omega,L^2(D))}\,.$$

In the case of the 1-norm exponential covariance, for any $p < p^*$, $0 < \rho < 1$ and $0 < s < 1/2$,

$$\|u - u_{K,h}\|_{L^p(\Omega,H_0^1(D))} \lesssim \left(C_{3.15}\, h^s \;+\; K^{\frac{\rho-1}{2}}\right) \|f\|_{L^{p^*}(\Omega,L^2(D))}\,,$$

where $C_{3.15} := \min\{\eta(K)\, h^{1-s}, C_{3.9}\}$ and $\eta(K)$ is an exponential function of K given in [5, Proposition 6.3].

Proof. It follows from Proposition 2.6 that a_K satisfies assumptions A1–A2. Therefore we can apply Theorem 3.9 to bound $\|u_K - u_{K,h}\|_{L^2(\Omega,H_0^1(D))}$. The result then follows from Theorem 2.7, Proposition 2.8 and the triangle inequality. □

In contrast to [5, Proposition 6.3] this bound is uniform in K. Note that for small values of K, the constant $C_{3.15}$ will actually be of order $\mathcal{O}(h^{1-s})$ leading to a linear dependence on h also in the exponential covariance case (as stated in [5, Proposition 6.3]). For larger values of K this will not be the case and the lower regularity of u_K will affect the convergence rate with respect to h.

Remark 3.16. As we will see in Section 5, in the exponential covariance case for correlation lengths that are smaller than the diameter of the domain, a relatively large number K of KL-modes is necessary to get even 10% accuracy. The new uniform error bound in Theorem 3.15 is therefore crucial also for the analysis of stochastic Galerkin and stochastic collocation methods, such as the one given in [5].

4 Convergence Analysis of Multilevel Monte Carlo Methods

We will now apply this new finite element error analysis in Section 4, to give a rigorous bound on the cost of the multilevel Monte Carlo method applied to (2.1) for general random fields satisfying Assumptions A1–A3, and to establish its superiority over the classical Monte Carlo method. This builds on the recent paper [7]. We start by briefly recalling the classical Monte Carlo (MC) and multilevel Monte Carlo (MLMC) algorithms for PDEs with random coefficients, together with the main results on their performance. For a more detailed description of the methods, we refer the reader to [7] and the references therein.

In the Monte Carlo framework, we are usually interested in finding the expected value of some functional $Q = \mathcal{G}(u)$ of the solution u to our model problem (2.1). Since u is not easily accessible, Q is often approximated by the quantity $Q_h := \mathcal{G}(u_h)$, where u_h denotes the finite element solution on a sufficiently fine spatial grid $\mathcal{T}_h$. Thus, to estimate $\mathbb{E}[Q] := \|Q\|_{L^1(\Omega)}$, we 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)

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

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

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

$$\widehat{Q}^{\mathrm{MC}}_{h,N} := \frac{1}{N}\sum_{i=1}^{N} Q_h(\omega^{(i)}), \tag{4.1}$$

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 (4.1), the approximation of Q by Q_h, which is related to the spatial discretisation, 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 $\mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{h,N}] = \mathbb{E}[Q_h]$ and $\mathbb{V}[\widehat{Q}^{\mathrm{MC}}_{h,N}] = N^{-1}\,\mathbb{V}[Q_h]$, 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

$$e(\widehat{Q}^{\mathrm{MC}}_{h,N})^2 \;=\; N^{-1}\mathbb{V}[Q_h] + \big(\mathbb{E}[Q_h - Q]\big)^2. \tag{4.2}$$

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

$$\mathbb{E}[Q_h] = \mathbb{E}[Q_{h_0}] + \sum_{\ell=1}^{L} \mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}] \tag{4.3}$$

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, and $h_{\ell-1}/h_\ell \le M^*$, for all $\ell = 1, \dots, L$. 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 minimised 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

$$\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}} \;:=\; \sum_{\ell=0}^{L} \widehat{Y}^{\mathrm{MC}}_{\ell,N_\ell} \;=\; \sum_{\ell=0}^{L} \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} Y_\ell(\omega^{(i)}), \tag{4.4}$$

where importantly $Y_\ell(\omega^{(i)}) = Q_{h_\ell}(\omega^{(i)}) - Q_{h_{\ell-1}}(\omega^{(i)})$, i.e. using the same sample on both meshes.

Since all the expectations $\mathbb{E}[Y_\ell]$ are estimated independently in (4.3), the variance of the MLMC estimator is $\sum_{\ell=0}^{L} N_\ell^{-1}\,\mathbb{V}[Y_\ell]$ and expanding as in (4.2) leads again to

$$e(\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}})^2 \;:=\; \mathbb{E}\Big[\big(\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}} - \mathbb{E}[Q]\big)^2\Big] \;=\; \sum_{\ell=0}^{L} N_\ell^{-1}\,\mathbb{V}[Y_\ell] \;+\; \big(\mathbb{E}[Q_h - Q]\big)^2. \tag{4.5}$$

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 (4.2). 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 $\mathcal{C}_\ell$ 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. [7, 16]).

Theorem 4.1. *Suppose that there are positive constants $\alpha, \beta, \gamma > 0$ such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and*

M1. $|\mathbb{E}[Q_h - Q]| \;=\; O(h^\alpha)$

M2. $\mathbb{V}[Q_{h_\ell} - Q_{h_{\ell-1}}] \;=\; O(h_\ell^\beta)$

M2. $\mathcal{C}_\ell \;=\; O(h_\ell^{-\gamma})$,

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

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

For the classical MC estimator we have $\mathcal{C}_\varepsilon(\widehat{Q}^{\mathrm{MC}}_h) \;=\; O(\varepsilon^{-2-\gamma/\alpha})$.

We can now use the results from Section 3.2 to verify Assumptions M1–M2 in Theorem 4.1 for some simple functionals $\mathcal{G}(u)$. More complicated functionals could then be tackled by duality arguments in a similar way.

Proposition 4.2. *Let $Q = \mathcal{G}(u) := |u|^q_{H^1(D)}$, for some $1 \leq q < p_*/2$. Suppose that D is a $\mathcal{C}^2$ bounded domain and that Assumptions A1–A4 hold, for some $0 < t < 1$. Then Assumptions M1–M2 in Theorem 4.1 hold for any $\alpha < t$ and $\beta < 2t$. For $t = 1$, we get $\alpha = 1$ and $\beta = 2$.*

Proof. We only give the proof for $t < 1$. The proof for $t = 1$ is analogous. Let $Q_h := |u_h|^q_{H^1(D)}$. Using the expansion $a^q - b^q = (a-b)\sum_{j=0}^{q-1} a^j b^{q-1-j}$, for $a, b \in \mathbb{R}$ and $q \in \mathbb{N}$, we get

$$|Q(\omega) - Q_h(\omega)| \;\lesssim\; |u(\omega,\cdot) - u_h(\omega,\cdot)|_{H^1(D)} \max\Big\{|u(\omega,\cdot)|^{q-1}_{H^1(D)}, |u_h(\omega,\cdot)|^{q-1}_{H^1(D)}, 1\Big\}, \quad \text{almost surely.}$$

This also holds for non-integer values of $q > 1$. Now, it follows from Lemma 2.1 and Theorem 3.9 that, for any $\alpha < t$,

$$|Q(\omega) - Q_h(\omega)| \;\lesssim\; \frac{C_{3.1}(\omega)\, C_{3.8}(\omega)}{a_{\min}^{q-1}(\omega)}\, \|f(\omega,\cdot)\|^q_{H^{\alpha-1}(D)}\, h^\alpha, \quad \text{almost surely.}$$

d	$\lvert u\rvert_{H^1(D)}$ MC	$\lvert u\rvert_{H^1(D)}$ MLMC	$\lVert u\rVert_{L^2(D)}$ MC	$\lVert u\rVert_{L^2(D)}$ 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}

d	$\lvert u\rvert_{H^1(D)}$ MC	$\lvert u\rvert_{H^1(D)}$ MLMC	$\lVert u\rVert_{L^2(D)}$ MC	$\lVert u\rVert_{L^2(D)}$ MLMC
1	ε^{-4}	ε^{-2}	ε^{-3}	ε^{-2}
2	ε^{-6}	ε^{-4}	ε^{-4}	ε^{-2}
3	ε^{-8}	ε^{-6}	ε^{-5}	ε^{-3}

Table 1: Theoretical upper bounds for the ε-costs of classical and multilevel Monte Carlo from Theorem 4.1 in the case of log-normal fields a with Gaussian covariance function (left) and exponential covariance function (right). (For simplicity we wrote ε^{-p}, instead of $\varepsilon^{-p-\delta}$ with $\delta > 0$.)

Taking expectation on both sides and applying Hölder's inequality, since $q < p_*$, it follows from Assumptions A1–A3 that Assumption M1 holds with $\alpha < t$.

To prove Assumption M2, let us consider $Y_{h_\ell} := Q_{h_\ell} - Q_{h_{\ell-1}}$. As above, it follows from Lemma 2.1 and Theorem 3.9 together with the triangle inequality that, for any $s < t$,

$$\begin{aligned} |Y_{h_\ell}(\omega)| &\lesssim |u_{h_\ell}(\omega,\cdot) - u_{h_{\ell-1}}(\omega,\cdot)|_{H^1(D)} \max\left\{|u_{h_\ell}(\omega,\cdot)|^{q-1}_{H^1(D)}, |u_{h_{\ell-1}}(\omega,\cdot)|^{q-1}_{H^1(D)}, 1\right\} \\ &\lesssim \frac{C_{3.1}(\omega)\, C_{3.8}(\omega)}{a_{\min}^{q-1}(\omega)} \|f(\omega,\cdot)\|^q_{H^{s-1}(D)}\, h_\ell^s, \qquad \text{almost surely,} \end{aligned}$$

where the hidden constant depends on M^*. It follows again from Assumptions A1-A3 and Hölder's inequality, since $q < p_*/2$, that

$$\mathbb{V}[Y_{h_\ell}] = \mathbb{E}[Y_{h_\ell}^2] - (\mathbb{E}[Y_{h_\ell}])^2 \leq \mathbb{E}[Y_{h_\ell}^2] \lesssim h_\ell^\beta, \quad \text{where} \quad \beta < 2t.$$

□

Proposition 4.3. *Let $Q := \|u\|^q_{L^2(D)}$, for some $1 \leq q < p_*/2$. Suppose that D is a $\mathcal{C}^2$ bounded domain and that Assumptions A1–A4 hold, for some $0 < t < 1$. Then Assumptions M1–M2 in Theorem 4.1 hold for any $\alpha < 2t$ and $\beta < 4t$. For $t = 1$, we get $\alpha = 2$ and $\beta = 4$.*

Proof. This can be shown in the same way as Proposition 4.2 using Corollary 3.10 instead of Theorem 3.9. □

Substituting these values into Theorem 4.1 we can get theoretical upper bounds for the ε-costs of classical and multilevel Monte Carlo in the case of log-normal fields a, as shown in Table 1. We assume here that we can obtain individual samples in optimal cost $\mathcal{C}_\ell \lesssim h_\ell^{-d}\log(h_\ell^{-1})$ via a multigrid solver, i.e. $\gamma = d+\delta$ for any $\delta > 0$. We clearly see the advantages of the multilevel Monte Carlo method. More importantly, note that in the exponential covariance case in dimensions $d > 1$, the cost of MLMC is proportional to the cost of obtaining one sample on the finest grid, i.e. solving one deterministic PDE with the same regularity properties to accuracy ε. This implies that the method is optimal.

5 Numerical Results

In this section, we want to confirm numerically some of the results proved in earlier sections. All numerical results shown are for the case of a log-normal random coefficient $a(\omega, x)$ with exponential covariance function (2.6). All results are calculated for a model problem in 1D. The domain D is taken to be $(0, 1)$, and $f \equiv 1$. The sampling from the random coefficient $a(\omega, x)$ is done using a truncated KL-expansion, and as in the analysis in Section 3.3, we use the midpoint rule to approximate the integrals in the stiffness matrix. As the quantities of interest we will just study the simple functionals $Q := \|u\|_{L^2(D)}$ and $Q := |u|_{H^1(D)}$.

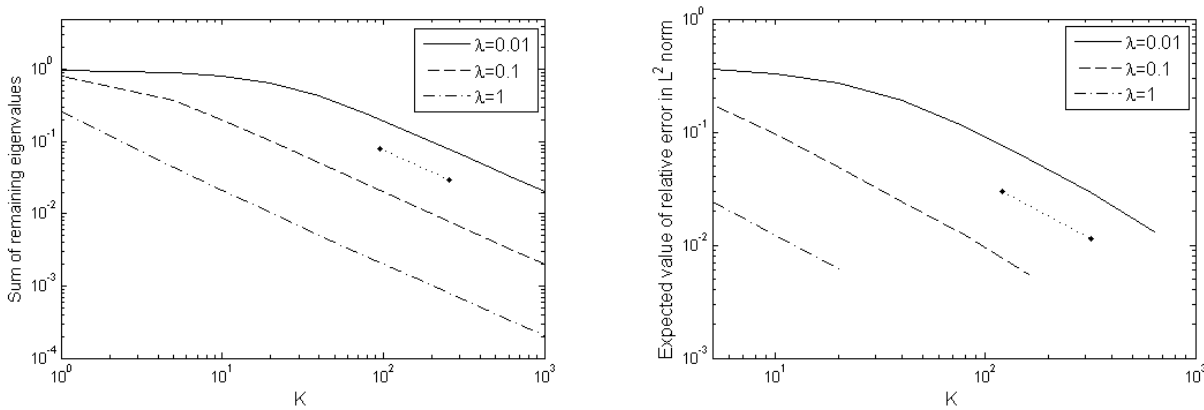

Figure 1: Let $d = 1$ and $\sigma^2 = 1$ in the exponential covariance case for different choices of the correlation length λ. Left: Plot of $\sum_{n>K} \theta_n$ as a function of K, i.e. the sum of the remaining KL eigenvalues when truncating after K terms. Right: The corresponding relative error $\left|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K,h^*}\|_{L^2(D)}\right]\right| / \mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)}\right]$ for problem (2.8) with $K^* = 5000$ and $h^* = 1/2048$. The dotted line has a gradient of -1.

5.1 Convergence with Respect to K

We start with the results in Section 2.4. We want to show that in the exponential covariance case a relatively large number of KL-modes are necessary (even in 1D) to obtain acceptable accuracies especially when the correlation length is smaller than the diameter of the domain. In Figure 1 we study the decay of the eigenvalues corresponding to the KL-expansion, as well as the convergence with respect to K of $\mathbb{E}\left[\|u_K\|_{L^2(D)}\right]$ to $\mathbb{E}\left[\|u\|_{L^2(D)}\right]$. Since we do not know the exact solution and cannot solve the differential equation explicitly, we approximate u by u_{K^*,h^*} with $K^* = 5000$ and $h^* = 1/2048$ and u_K by u_{K,h^*} with $h^* = 1/2048$.

In the left figure we plot $\sum_{n>K} \theta_n$, i.e. the sum of the remaining eigenvalues when truncating after K terms. We see that after a short pre-asymptotic phase (depending on the size of λ) this sum decays linearly in K^{-1}. On the right we plot $\left|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K,h^*}\|_{L^2(D)}\right]\right| / \mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)}\right]$, the relative error. We see that the error decays roughly like $\sum_{n>K} \theta_n$ and thus also linearly in K^{-1}. This is better than the bound for $\mathbb{E}\left[\|u_{K^*,h^*} - u_{K,h^*}\|_{L^2(D)}\right] \geq \left|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K,h^*}\|_{L^2(D)}\right]\right|$ in Proposition 2.8, which predicts at most $\mathcal{O}(K^{-1/2})$ convergence, and it is related to weak convergence. As shown in [5, Section 5], the weak convergence order of the PDE solution with truncated KL-expansion can be $\mathcal{O}(\sum_{n>K} \theta_n)$ for certain functionals. We believe that the faster convergence of $\mathbb{E}\left[\|u_K\|_{L^2(D)}\right]$ to $\mathbb{E}\left[\|u\|_{L^2(D)}\right]$ can be shown using similar techniques.

5.2 Convergence with Respect to h

We now move on to the results in the main Section 3. For the reference solution we choose again $h^* = 1/2048$ and $K^* = 5000$. Figure 2 shows the convergence of $\mathbb{E}\left[|u_{K^*,h^*} - u_{K^*,h}|_{H^1(D)}\right]$ and $\mathbb{E}\left[\|u_{K^*,h^*} - u_{K^*,h}\|_{L^2(D)}\right]$ for (2.8). In the plots, a dash-dotted line indicates $\mathcal{O}(h^{1/2})$ convergence and a dotted line indicates $\mathcal{O}(h)$ convergence.

In both cases, we can see a short pre-asymptotic phase, which is again related to the correlation length λ. We can see in the right plot that the H^1–error converges with $\mathcal{O}(h^{1/2})$, confirming the sharpness of the bound proven in Theorem 3.13. The L^2–error converges linearly in h, and so the quadrature error does not seem to be dominant here.

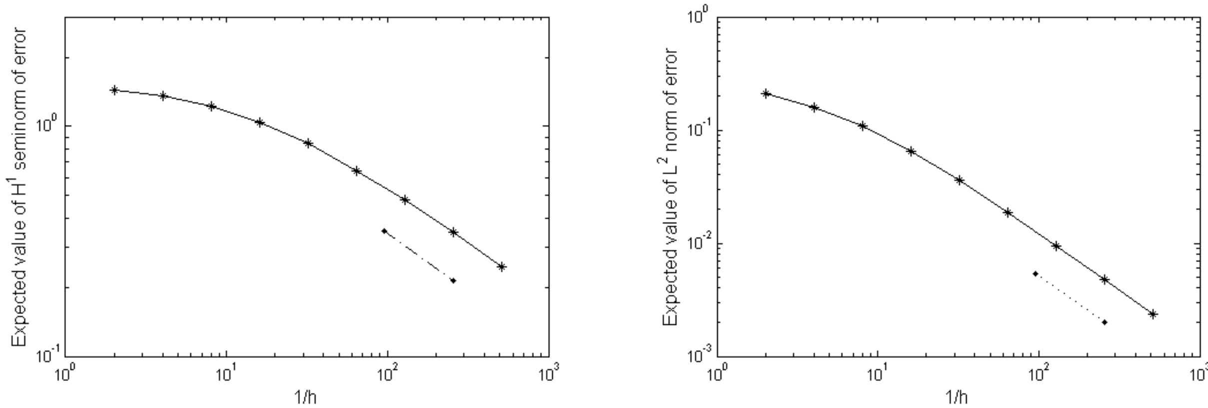

Figure 2: Left: Plot of $\mathbb{E}\left[|u_{K^*,h^*} - u_{K^*,h}|_{H^1(D)}\right]$ versus $1/h$ for model problem (2.8) with $d = 1$, $\lambda = 0.1$ and $\sigma^2 = 3$. Right: Corresponding L^2-error $\mathbb{E}\left[\|u_{K^*,h^*} - u_{K^*,h}\|_{L^2(D)}\right]$. The gradient of the dash-dotted (resp. dotted) line is $-1/2$ (resp. -1).

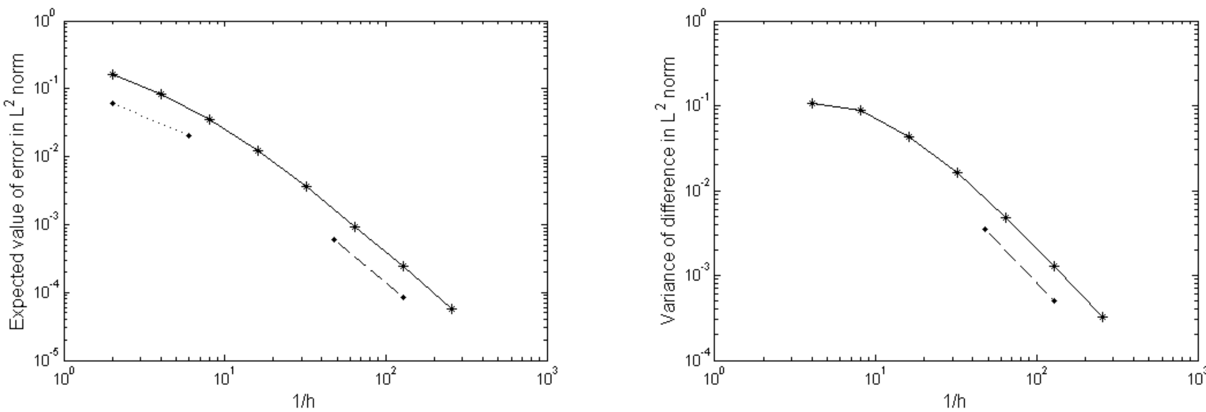

Figure 3: Left: Plot of $|\mathbb{E}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,h^*}\|_{L^2(D)}\right]|$, for model problem (2.8) with $d = 1$, $\lambda = 0.1$, $\sigma^2 = 3$, $K^* = 5000$ and $h^* = 1/2048$. Right: Corresponding plot of the variance $\mathbb{V}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,2h}\|_{L^2(D)}\right]$. The gradient of the dotted (resp. dashed) line is -1 (resp. -2).

5.3 Multilevel Monte Carlo Convergence

We first want to confirm Assumptions M1 and M2 in Theorem 4.1, for $Q = \|u\|_{L^2(D)}$. In other words, we want to confirm the rate of decay for the quantities $|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K^*,h}\|_{L^2(D)}\right]|$ and $\mathbb{V}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,2h}\|_{L^2(D)}\right]$. Again, we choose $K^* = 5000$ and $h^* = 1/2048$ and look at the model problem (2.8) for $d = 1$. The results are shown in Figure 3. A dotted line indicates linear convergence, and a dashed line indicates quadratic convergence.

For the variance $\mathbb{V}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,2h}\|_{L^2(D)}\right]$ (right plot), we observe the quadratic convergence predicted by Proposition 4.3. The expected value in the left plot seems to converge slightly faster than the predicted linear convergence. In our experience, this faster convergence depends on the choice of model problem and quantity of interest and will need to be investigated further, since it directly affects the cost of the multilevel Monte Carlo method through the size of the ratio β/α in the bound in Theorem 4.1.

The actual performance of the standard MC and MLMC estimators in estimating $Q = \|u\|_{L^2(D)}$ is shown in Figure 4. In the right plot the accuracy is scaled by the expected value of the quantity

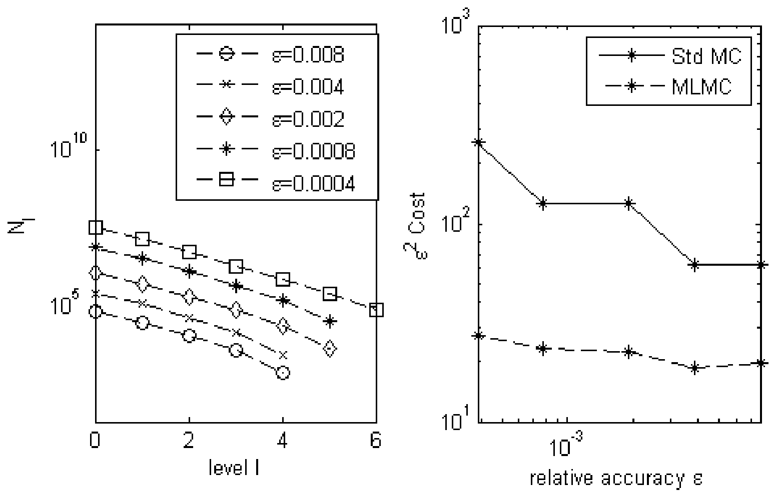

Figure 4: Left: Number of samples N_ℓ per level. Right: Plot of the cost scaled by ε^{-2} of the MLMC and standard MC estimators for $d = 1$, with $\lambda = 0.1$ and $\sigma^2 = 3$. The coarsest mesh size in all tests is $h_0 = 1/16$.

of interest. We see a clear advantage of the multilevel Monte Carlo method. For more numerical results, in particular results in 2D, we refer the reader to [7].

6 Conclusions and Further Work

One of the major bottlenecks in probabilistic uncertainty quantification is the efficient numerical solution of PDEs with random coefficients. A typical model problem that arises in subsurface flow is an elliptic PDE with log-normal coefficients. The resulting PDE is not uniformly elliptic or bounded with respect to the random variable ω and usually lacks also full regularity. In this paper, we carry out for the first time a careful finite element error analysis under these weaker assumptions and provide optimal error estimates — with respect to the given regularity of the coefficients — for all finite moments of the finite element error (including quadrature errors). This then allows us to rigorously analyse the convergence and the cost of multilevel Monte Carlo methods, recently proposed for elliptic PDEs with random coefficients, in particular in the practically important case of log-normal coefficients with exponential covariance which has hitherto been unproved.

Via duality arguments, the results in this paper are readily applicable also to other functionals of the solution of more practical relevance. Other important aspects which we want to investigate in the future are the influence of the correlation length λ on the MLMC convergence, as well as the superconvergence in the expected value of certain model problems and quantities of interest (as highlighted in the discussion of Figure 2). Another issue which will require further research is the choice of quadrature rule. So far we have only used simple midpoint or trapezoidal rules. In future work we would like to also explore multiscale methods and model reduction on coarser grids.

A Detailed Regularity Proof

In this appendix, we give more detailed proofs of Proposition 3.1 and Lemmas 3.2 and 3.3. The proofs follow those of Hackbusch [18, Theorems 9.1.8, 9.1.11 and 9.1.16], making explicit the dependence of all the constants that appear on the PDE coefficient a. The proof follows the classical Nirenberg translation method and consists in three main steps.

A.1 Step 1 – The Case $D = \mathbb{R}^d$

Proof of Lemma 3.2. In this proof we will use the norm on $H^s(\mathbb{R}^d)$ provided by the Fourier transform, $\|u\|^2_{H^s(\mathbb{R}^d)} := \|\hat{u}(\xi)(1+|\xi|^2)^{s/2}\|_{L^2(\mathbb{R}^d)}$, which is equivalent to the norm defined previously and defines the same space.

For any $h > 0$, we define the fractional difference operator (in direction $i = 1, \dots, d$) by

$$R_h^i(v)(x) := h^{-s} \sum_{\mu=0}^{+\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} v(x + \mu h e_i),$$

where e_i is the ith unit vector in $\mathbb{R}^d$,

$$\binom{s}{0} = 1 \quad \text{and} \quad \binom{s}{\mu} (-1)^\mu = \frac{-s(1-s)(2-s)...(\mu-1-s)}{\mu!}.$$

Let us recall here some properties of R_h^i from [18, Proof of Theorem 9.1.8]:

- $(R_h^i)^*(v)(x) = h^{-s} \sum_{\mu=0}^{+\infty} e^{-\mu h}(-1)^\mu \binom{s}{\mu} v(x - \mu h e_i)$.
- For any $\tau \in \mathbb{R}$ and $v \in H^{\tau+s}(\mathbb{R}^d)$,
$$\|R_h^i v\|_{H^\tau(\mathbb{R}^d)} \le \|v\|_{H^{\tau+s}(\mathbb{R}^d)} \text{ and } \|(R_h^i)^* v\|_{H^\tau(\mathbb{R}^d)} \le \|v\|_{H^{\tau+s}(\mathbb{R}^d)}. \tag{A.1}$$
- $\widehat{R_h^i v}(\xi) = [(1 - e^{-h + i\xi_j h})/h]^s \hat{v}(\xi) \quad$ and $\quad \widehat{(R_h^i)^* v}(\xi) = [(1 - e^{-h - i\xi_j h})/h]^s \hat{v}(\xi)$.

We define for $u, v \in H^1(\mathbb{R}^d)$ the bilinear form

$$\begin{aligned} d(u,v) &:= \int_{\mathbb{R}^d} A \nabla u \nabla R_h^i(v) \,\mathrm{d}x - \int_{\mathbb{R}^d} A \nabla (R_h^i)^* u \nabla v \,\mathrm{d}x \\ &= \sum_{\mu=1}^{\infty} h^{-s} e^{-\mu h} (-1)^\mu \binom{s}{\mu} \int_{\mathbb{R}^d} (A(x - \mu h e_i) - A(x)) \nabla u(x - \mu h e_i) \nabla v \,\mathrm{d}x. \end{aligned}$$

Hence,

$$|d(u,v)| \lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))} |u|_{H^1(\mathbb{R}^d)} |v|_{H^1(\mathbb{R}^d)},$$

where the hidden constant is proportional to

$$\sum_{\mu=1}^{\infty} h^{-s} e^{-\mu h} (-1)^\mu \binom{s}{\mu} (\mu h)^t,$$

which is finite, since $\binom{s}{\mu} = \mathcal{O}(\mu^{-s-1})$ and thus $\sum_{\mu=1}^{\infty} e^{-\mu h} \mu^{t-s-1} = \mathcal{O}(h^{s-t})$. The three spaces $H^{1-s}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d) \subset H^{s-1}(\mathbb{R}^d)$ form a Gelfand triple, so that we can deduce, using $(A.1)$, that

$$\begin{aligned} A_{\min} |(R_h^i)^* w|^2_{H^1(\mathbb{R}^d)} &\le \int_{\mathbb{R}^d} A \nabla (R_h^i)^* w \nabla (R_h^i)^* w \,\mathrm{d}x \\ &= -d(w, (R_h^i)^* w) + \langle F, R_h^i (R_h^i)^* w \rangle_{H^{s-1}(\mathbb{R}^d), H^{1-s}(\mathbb{R}^d)} \\ &\le |d(w, (R_h^i)^* w)| + \|F\|_{H^{s-1}(\mathbb{R}^d)} \|R_h^i (R_h^i)^* w\|_{H^{1-s}(\mathbb{R}^d)} \\ &\lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))} |w|_{H^1(\mathbb{R}^d)} |(R_h^i)^* w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)} \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)}, \end{aligned}$$

therefore we get

$$
\begin{aligned}
A_{\min}\|(R_h^i)^*w\|^2_{H^1(\mathbb{R}^d)} &\lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)}\,|(R_h^i)^*w|_{H^1(\mathbb{R}^d)} + \\
&\qquad \|F\|_{H^{s-1}(\mathbb{R}^d)}\,\|(R_h^i)^*w\|_{H^1(\mathbb{R}^d)} + A_{\min}\|(R_h^i)^*w\|^2_{L^2(\mathbb{R}^d)} \\
&\lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)}\,|(R_h^i)^*w|_{H^1(\mathbb{R}^d)} + \\
&\qquad \|F\|_{H^{s-1}(\mathbb{R}^d)}\,\|(R_h^i)^*w\|_{H^1(\mathbb{R}^d)} + A_{\min}\|(R_h^i)^*w\|_{H^{-1}(\mathbb{R}^d)}\|(R_h^i)^*w\|_{H^1(\mathbb{R}^d)},
\end{aligned}
$$

and finally, using $(A.1)$ once more,

$$
\|(R_h^i)^*w\|_{H^1(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\Big) + \|w\|_{L^2(\mathbb{R}^d)}.
$$

For any $1 \geq h > 0$, since $|1-e^{-h-i\xi_i h}|^2 \geq |\mathrm{Im}(1-e^{-h-i\xi_i h})|^2 = e^{-2h}\sin(\xi_i h)^2 \geq e^{-2}\sin(\xi_i h)^2$, and since $\sin^2(\xi h) \geq \left(\frac{2}{\pi}\xi h\right)^2$, for all $|\xi| \leq 1/h$, we have conversely that

$$
\begin{aligned}
\sum_{i=1}^d \|(R_h^i)^*w\|^2_{H^1(\mathbb{R}^d)} &\geq \int_{|\xi|\leq 1/h}(1+|\xi|^2)\sum_{i=1}^d |\widehat{(R_h^i)^*w}(\xi)|^2 d\xi \\
&= \int_{|\xi|\leq 1/h}(1+|\xi|^2)\sum_{i=1}^d \left|\frac{1-e^{-h-i\xi_i h}}{h}\right|^{2s} |\hat{w}(\xi)|^2 d\xi \\
&\geq e^{-2}\int_{|\xi|\leq 1/h}(1+|\xi|^2)\sum_{i=1}^d \left|\frac{\sin(\xi_i h)}{h}\right|^{2s} |\hat{w}(\xi)|^2 d\xi \\
&\gtrsim \int_{|\xi|\leq 1/h}(1+|\xi|^2)|\xi|^{2s}|\hat{w}(\xi)|^2 d\xi.
\end{aligned}
$$

Hence, for any $0 < h \leq 1$, we obtain

$$
\begin{aligned}
\|w\|^2_{H^{1+s}(\mathbb{R}^d)} &\leq \int_{\mathbb{R}^d}(1+|\xi|^2)|\xi|^{2s}|\hat{w}(\xi)|^2 d\xi + \int_{\mathbb{R}^d}(1+|\xi|^2)|\hat{w}(\xi)|^2 d\xi \\
&\leq \sum_{i=1}^d \|(R_h^i)^*w\|^2_{H^1(\mathbb{R}^d)} + \|w\|^2_{H^1(\mathbb{R}^d)}
\end{aligned}
$$

and so

$$
\|w\|_{H^{1+s}(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\Big) + \|w\|_{H^1(\mathbb{R}^d)} < +\infty.
$$

□

A.2 Step 2 – The Case $D = \mathbb{R}^d_+$

In order to proof Lemma 3.3 we will need the following two lemmas.

Lemma A.1 ([18, Lemma 9.1.12])**.** *Let $s > 0$, $s \neq 1/2$, then*

$$
|||v|||_s := \left(\|v\|^2_{L^2(\mathbb{R}^d_+)} + \sum_{i=1}^d \left\|\frac{\partial v}{\partial x_i}\right\|^2_{H^{s-1}(\mathbb{R}^d_+)}\right)^{1/2}
$$

defines a norm on $H^s(\mathbb{R}^d_+)$ that is equivalent to $\|v\|_{H^s(\mathbb{R}^d_+)}$.

Lemma A.2. *Let $D \subset \mathbb{R}^d$ and $0 < s < t < 1$. If $b \in \mathcal{C}^t(\overline{D})$ and $v \in H^s(D)$, then $bv \in H^s(D)$ and*

$$\|bv\|_{H^s(D)} \lesssim |b|_{\mathcal{C}^t(\overline{D})}\|v\|_{L^2(D)} + |b|_{\mathcal{C}^0(\overline{D})}\|v\|_{H^s(D)}.$$

The hidden constant depends only on t, s and d.

Proof. This is a classical result, but we require again the exact dependence of the bound on b. First note that trivially $\|bv\|_{L^2(D)} \leq b_{\max}\|v\|_{L^2(D)}$ where $b_{\max} := |b|_{\mathcal{C}^0(\overline{D})}$. Now, for any $x, y \in D$, we have

$$|b(x)v(x) - b(y)v(y)|^2 \leq 2(b(x)^2|v(x) - v(y)|^2 + v(y)^2|b(x) - b(y)|^2).$$

Continuing v by 0 on $\mathbb{R}^d$ and denoting this extension by $\tilde{v}$, this implies

$$\begin{aligned}
\iint_{D^2} \frac{|b(x)v(x) - b(y)v(y)|^2}{\|x-y\|^{d+2s}}\,\mathrm{d}x\,\mathrm{d}y \;&\leq\; 2b_{\max}^2\|v\|_{H^s(D)}^2 + 2\iint_{D^2} \frac{v(y)^2|b(x) - b(y)|^2}{\|x-y\|^{d+2s}} \\
&\leq\; 2b_{\max}^2\|v\|_{H^s(D)}^2 + \iint_{\substack{x,y\in D\\ \|x-y\|\geq 1}} 8b_{\max}^2\frac{v(y)^2}{\|x-y\|^{d+2s}} \;+\; 2|b|_{\mathcal{C}^t(\overline{D})}^2\frac{v(y)^2}{\|x-y\|^{d+2(s-t)}}\,\mathrm{d}x\,\mathrm{d}y \\
&\leq\; 2b_{\max}^2\|v\|_{H^s(D)}^2 + \left(8b_{\max}^2\left\|\frac{\mathbb{1}_{\|z\|\geq 1}}{\|z\|^{d+2s}}\right\|_{L^1(\mathbb{R}^d)} \;+\; 2|b|_{\mathcal{C}^t(\overline{D})}^2\left\|\frac{\mathbb{1}_{\|z\|\leq 1}}{\|z\|^{d+2(s-t)}}\right\|_{L^1(\mathbb{R}^d)}\right)\|\tilde{v}\|_{L^2(\mathbb{R}^d)}^2 \\
&\lesssim\; b_{\max}^2\|v\|_{H^s(D)}^2 \;+\; |b|_{\mathcal{C}^t(\overline{D})}^2\|v\|_{L^2(D)}^2.
\end{aligned}$$

□

Proof of Lemma 3.3. First we continue the solution w by 0 on $\mathbb{R}^d \setminus \mathbb{R}^d_+$ and denote the extension $\tilde{w} \in H^1(\mathbb{R}^d)$. Take $1 \leq i \leq d-1$. Similarly to the previous section, we define for $u, v \in H^1(\mathbb{R}^d)$

$$d(u,v) \;:=\; \int_{\mathbb{R}^d_+} A\nabla u \nabla R_h^i(v)\,\mathrm{d}x - \int_{\mathbb{R}^d_+} A\nabla(R_h^i)^* u \nabla v\,\mathrm{d}x$$

and deduce again that

$$|d(u,v)| \;\lesssim\; |A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))}\,|u|_{H^1(\mathbb{R}^d_+)}\,|v|_{H^1(\mathbb{R}^d_+)}\,.$$

We now note that, since $i \neq d$, $(R_h^i)^* w \in H^1_0(\mathbb{R}^d_+)$ and $(R_h^i)^*\tilde{w} \in H^1(\mathbb{R}^d)$ is equal to the continuation by 0 on $\mathbb{R}^d \setminus \mathbb{R}^d_+$ of $(R_h^i)^* w$. We deduce, similarly to the proof in Section A.1 using $(A.1)$, that

$$\|(R_h^i)^*\tilde{w}\|_{H^1(\mathbb{R}^d)} \;\lesssim\; \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) + \|w\|_{H^1(\mathbb{R}^d_+)} \;=:\; \mathcal{B}(w).$$

(Note that we added $|w|_{H^1(\mathbb{R}^d_+)}$ to the bound to simplify the notation later.) Hence, by the same token as in the previous section, we get

$$\int_{\mathbb{R}^d} (1 + |\xi|^2)(|\xi_1|^2 + ... + |\xi_{d-1}|^2)^s|\widehat{\tilde{w}}(\xi)|^2 d\xi \;\lesssim\; \mathcal{B}(w)^2. \tag{A.2}$$

In particular, this implies that, for $1 \leq i \leq d$ and $1 \leq j \leq d-1$, we have

$$\int_{\mathbb{R}^d} \left|\widehat{\frac{\partial^2 \tilde{w}}{\partial x_i \partial x_j}}(\xi)\right|^2 (1 + |\xi|^2)^{s-1} d\xi \;=\; \int_{\mathbb{R}^d} |\xi_i|^2|\xi_j|^2|\widehat{\tilde{w}}(\xi)|^2(1 + |\xi|^2)^{s-1} d\xi \;\lesssim\; \mathcal{B}(w)^2,$$

which means that $\frac{\partial^2 \tilde{w}}{\partial x_i \partial x_j} \in H^{s-1}(\mathbb{R}^d)$ and $\left\| \frac{\partial^2 \tilde{w}}{\partial x_i \partial x_j} \right\|_{H^{s-1}(\mathbb{R}^d)} \lesssim \mathcal{B}(w)$. In particular, for all $(i,j) \neq (d,d)$, this further implies that $\frac{\partial^2 w}{\partial x_i \partial x_j} \in H^{s-1}(\mathbb{R}^d_+)$ and that $\left\| \frac{\partial^2 w}{\partial x_i \partial x_j} \right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim \mathcal{B}(w)$. Using Lemma A.1 we deduce that $\frac{\partial w}{\partial x_j} \in H^s(\mathbb{R}^d_+)$ and that

$$\left\| \frac{\partial w}{\partial x_j} \right\|_{H^s(\mathbb{R}^d_+)} \lesssim \mathcal{B}(w), \qquad \text{for all } 1 \leq j \leq d-1. \tag{A.3}$$

It remains to bound $\left\| \frac{\partial w}{\partial x_d} \right\|_{H^s(\mathbb{R}^d_+)}$, which is rather technical. To achieve it we will use the PDE (3.1), Lemma A.2 and the following result.

Lemma A.3. *For almost all $x_d \in \mathbb{R}$, we have $\frac{\partial \tilde{w}}{\partial x_d}(.,x_d) \in H^s(\mathbb{R}^{d-1})$ and*

$$\int_{\mathbb{R}} \left\| \frac{\partial \tilde{w}}{\partial x_d}(.,x_d) \right\|^2_{H^s(\mathbb{R}^{d-1})} \mathrm{d}x_d = \int_{\mathbb{R}^d} (1+|\xi'|^2)^s |\xi_d|^2 \left| \widehat{\tilde{w}}(\xi) \right|^2 d\xi \lesssim \mathcal{B}(w)^2.$$

Proof. This follows from Fubini's theorem and Plancherel's formula, together with (A.2). □

From this we deduce that $\frac{\partial w}{\partial x_d}(.,x_d) \in H^s(\mathbb{R}^{d-1})$, for almost all $x_d \in \mathbb{R}_+$, and that

$$\int_{\mathbb{R}_+} \left\| \frac{\partial w}{\partial x_d}(.,x_d) \right\|^2_{H^s(\mathbb{R}^{d-1})} \mathrm{d}x_d \lesssim \mathcal{B}(w)^2.$$

Let $1 \leq i \leq d-1$. Using Lemma A.2 we deduce that $A_{id}\frac{\partial w}{\partial x_d}(.,x_d) \in H^s(\mathbb{R}^{d-1})$, for almost all $x_d \in \mathbb{R}_+$, and that

$$\begin{aligned} \left\| \left(A_{id} \frac{\partial w}{\partial x_d} \right)(.,x_d) \right\|_{H^s(\mathbb{R}^{d-1})} &\lesssim |A_{id}(.,x_d)|_{\mathcal{C}^t(\mathbb{R}^{d-1})} \left\| \frac{\partial w}{\partial x_d}(.,x_d) \right\|_{L^2(\mathbb{R}^{d-1})} \\ &+ \|A_{id}(.,x_d)\|_{\mathcal{C}^0(\mathbb{R}^{d-1})} \left\| \frac{\partial w}{\partial x_d}(.,x_d) \right\|_{H^s(\mathbb{R}^{d-1})}. \end{aligned}$$

Therefore, since by definition $\|A_{id}\|_{\mathcal{C}^0(\mathbb{R}^d)} \leq A_{\max}$, we get

$$\int_{\mathbb{R}_+} \left\| A_{id} \frac{\partial w}{\partial x_d} \right\|^2_{H^s(\mathbb{R}^{d-1})} \mathrm{d}x_d \lesssim |A_{id}|^2_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})} |w|^2_{H^1(\mathbb{R}^d_+)} + A^2_{\max} \mathcal{B}(w)^2 \lesssim A^2_{\max} \mathcal{B}(w)^2.$$

Since $\frac{\partial}{\partial x_i}$ is linear continuous from $H^{1-s}(\mathbb{R}^{d-1})$ to $H^{-s}(\mathbb{R}^{d-1})$ (cf. [18, Remark 6.3.14(b)]) we can deduce from this that $\frac{\partial}{\partial x_i}\left(A_{id} \frac{\partial w}{\partial x_d} \right) \in H^{s-1}(\mathbb{R}^d_+)$ and that

$$\left\| \frac{\partial}{\partial x_i} \left(A_{id} \frac{\partial w}{\partial x_d} \right) \right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max} \mathcal{B}(w), \qquad \text{for all } 1 \leq i \leq d-1. \tag{A.4}$$

To see this take $\varphi \in \mathcal{D}(\mathbb{R}^d_+)$. Then

$$\begin{aligned} \left| \left\langle \frac{\partial}{\partial x_i} \left(A_{id} \frac{\partial w}{\partial x_d} \right), \varphi \right\rangle_{\mathcal{D}'(\mathbb{R}^d_+), \mathcal{D}(\mathbb{R}^d_+)} \right| &= \left\| A_{id} \frac{\partial w}{\partial x_d}(x',x_d) \frac{\partial \varphi}{\partial x_i}(x',x_d) \right\|_{L^2(\mathbb{R}_+, H^s(\mathbb{R}^{d-1}))} \\ &\leq \left\| A_{id} \frac{\partial w}{\partial x_d} \right\|_{L^2(\mathbb{R}_+, H^s(\mathbb{R}^{d-1}))} \left\| \frac{\partial \varphi}{\partial x_i}(x',x_d) \right\|_{L^2(\mathbb{R}_+, H^{-s}(\mathbb{R}^{d-1}))} \\ &\leq \left\| A_{id} \frac{\partial w}{\partial x_d} \right\|_{L^2(\mathbb{R}_+, H^s(\mathbb{R}^{d-1}))} \|\varphi\|_{L^2(\mathbb{R}_+, H^{1-s}(\mathbb{R}^{d-1}))}. \end{aligned}$$

Using $(A.3)$ and Lemma $A.2$, we deduce in a similar way that $\frac{\partial}{\partial x_i}\left(A_{ij}\frac{\partial w}{\partial x_j}\right) \in H^{s-1}(\mathbb{R}^d_+)$ and that

$$\left\|\frac{\partial}{\partial x_i}\left(A_{ij}\frac{\partial w}{\partial x_j}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\mathcal{B}(w), \quad \text{for all } \ 1 \le i \le d \text{ and } 1 \le j \le d-1. \tag{A.5}$$

We can now use the PDE (3.1) to get a similar bound for $(i,j) = (d,d)$. Since $F \in H^{s-1}(\mathbb{R}^d_+)$, it follows from $(A.4)$ and $(A.5)$ that $\frac{\partial}{\partial x_d}\left(A_{dd}\frac{\partial w}{\partial x_d}\right) \in H^{s-1}(\mathbb{R}^d_+)$ and that

$$\left\|\frac{\partial}{\partial x_d}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\mathcal{B}(w) \ + \ \|F\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\mathcal{B}(w)\,.$$

Analogously to $(A.4)$ we can prove that

$$\left\|\frac{\partial}{\partial x_i}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\mathcal{B}(w), \qquad \text{for all } \ 1 \le i \le d-1.$$

Hence, we can finally apply Lemma $A.1$ to get that $A_{dd}\frac{\partial w}{\partial x_d} \in H^s(\mathbb{R}^d_+)$ and that

$$\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|^2_{H^s(\mathbb{R}^d_+)} \lesssim \sum_{i=1}^{d}\left\|\frac{\partial}{\partial x_i}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\right\|^2_{H^{s-1}(\mathbb{R}^d_+)} + \left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|^2_{L^2(\mathbb{R}^d_+)} \lesssim A^2_{\max}\mathcal{B}(w)^2\,.$$

By applying Lemma $A.2$ again, this time with $b := 1/A_{dd}$ and $v := A_{dd}\frac{\partial w}{\partial x_d}$, we deduce that $\frac{\partial w}{\partial x_d} \in H^s(\mathbb{R}^d_+)$ and that

$$\begin{aligned}\left\|\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)} &\lesssim \left|\frac{1}{A_{dd}}\right|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|_{L^2(\mathbb{R}^d_+)} + \frac{1}{A_{\min}}\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)} \\ &\lesssim \frac{|A_{dd}|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}}{A^2_{\min}}A_{\max}\,|w|_{H^1(\mathbb{R}^d_+)} + \frac{1}{A_{\min}}\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)} \lesssim \frac{A_{\max}}{A_{\min}}\mathcal{B}(w)\,.\end{aligned}$$

To finish the proof we use this bound together with $(A.3)$ and apply once more Lemma $A.1$ to show that $w \in H^{1+s}(\mathbb{R}^d_+)$ and

$$\|w\|_{H^{1+s}(\mathbb{R}^d_+)} \lesssim \frac{A_{\max}}{A^2_{\min}}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) \ + \ \frac{A_{\max}}{A_{\min}}\|w\|_{H^1(\mathbb{R}^d_+)}.$$

□

A.3 Step 3 – The Case D Bounded

We can now prove Proposition 3.1 using Lemmas 3.2 and 3.3 in two successive steps. We recall that D was assumed to be $\mathcal{C}^2$. Let $(D_i)_{0\le i\le m}$ be a covering of D such that the $(D_i)_{0\le i\le m}$ are open and bounded, $\overline{D} \subset \cup_{i=0}^m D_i$, $\cup_{i=1}^m(D_i \cap \partial D) = \partial D$, $\overline{D}_0 \subset D$.

Let $(\chi_i)_{0\le i\le m}$ be a partition of unity subordinate to this cover, i.e. we have $\chi_i \in \mathcal{C}^\infty(\mathbb{R}^d,\mathbb{R}_+)$ with compact support $\mathrm{supp}(\chi_i) \subset D_i$, such that $\sum_{i=0}^p \chi_i = 1$ on $\overline{D}$. We denote by u the solution of (2.3) and split it into $u = \sum_{i=0}^p u_i$, with $u_i = u\chi_i$. We treat now separately u_0 and then u_i, $1 \le i \le p$, using Section A.1 and A.2, respectively.

Lemma A.4. *u_0 belongs to $H^{1+s}(D)$ and*

$$\|u_0\|_{H^{1+s}(D)} \lesssim \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a^2_{\min}}\|f\|_{H^{s-1}(D)}\,.$$

Proof. Since $\mathrm{supp}(u_0) \subset D_0$, we have that $u_0 \in H^1_0(D)$ and it is the weak solution of the new equation $-\mathrm{div}(a\nabla u_0) = F$ on D, where

$$F \quad := \quad f\chi_0 + a\nabla u \cdot \nabla\chi_0 + \mathrm{div}(au\nabla\chi_0) \quad \text{on } D.$$

To apply Lemma 3.2 we will now extend all terms to $\mathbb{R}^d$, but continue to denote them the same. The terms u_0 and $f\chi_0 + a\nabla u \cdot \nabla\chi_0$ can both be extended by 0. Their extensions will belong to $H^1(\mathbb{R}^d)$ and $H^{s-1}(\mathbb{R}^d)$, respectively. It follows from Lemma $A.2$ that $au \in H^s(D)$ and so if we continue $au\nabla\chi_0$ by 0 on $\mathbb{R}^d$, the extension belongs to $H^s(\mathbb{R}^d)$, since $\mathrm{supp}(\chi_0)$ is compact in D. Using the fact that div is linear and continuous from $H^s(\mathbb{R}^d)$ to $H^{s-1}(\mathbb{R}^d)$ (cf. [18, Remark 6.3.14(b)] and the proof of (A.4) above), we can deduce that the divergence of the extension of $au\nabla\chi_0$ is in $H^{s-1}(\mathbb{R}^d)$, leading to an extension of F on $\mathbb{R}^d$, which belongs to $H^{s-1}(\mathbb{R}^d)$.

Let $\psi \in \mathcal{C}^\infty(\mathbb{R}^d, [0,1])$ such that $\psi = 0$ on D_0 and $\psi = 1$ on $\tilde{D}^c$, where $\tilde{D}$ is an open set such that $\overline{D}_0 \subset \tilde{D}$ and $\overline{\tilde{D}} \subset D$. We use the following extension of a from D_0 to all of $\mathbb{R}^d$:

$$\overline{a}(x) := \begin{cases} a(x)(1-\psi(x)) + a_{\min}\psi(x), & \text{if } x \in D, \\ a_{\min}\psi(x), & \text{otherwise.} \end{cases}$$

This implies that $\overline{a} \in \mathcal{C}^t(\mathbb{R}^d)$, and for any $x \in \mathbb{R}^d$, $a_{\min} \le \overline{a}(x) \le a(x)$ and $|\overline{a}|_{\mathcal{C}^t(\mathbb{R}^d)} \lesssim \|a\|_{\mathcal{C}^t(\overline{D})}$.

Using these extensions, we have that $-\mathrm{div}(\overline{a}\nabla u_0) = F$ in $\mathcal{D}'(\mathbb{R}^d)$. Indeed, for any $v \in \mathcal{D}(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \overline{a}(x)\nabla u_0(x)\nabla v(x) \quad = \quad \int_D a(x)\nabla u_0(x)\nabla v(x) \qquad \text{for any} \quad v \in \mathcal{D}(\mathbb{R}^d),$$

since $\mathrm{supp}(u_0)$ is included in the open bounded set D_0, which implies that $\nabla u_0 = 0$ on D_0^c and $a = \overline{a}$ on D_0. Since $u \in H^1_0(D)$, we have by Poincaré's inequality that $\|u\|_{L^2(D)} \lesssim |u|_{H^1(D)}$. Therefore it follows from Lemma 2.1 that

$$|u_0|_{H^1(\mathbb{R}^d)} \quad \le \quad |u|_{H^1(D)}\|\chi_0\|_\infty + \|u\|_{L^2(D)}\|\nabla\chi_0\|_\infty \;\lesssim\; \frac{\|f\|_{H^{s-1}(D)}}{a_{\min}}.$$

Since $\chi_0 \in \mathcal{C}^\infty(\mathbb{R}^d)$, using Lemma $A.2$ and the linearity and continuity of div from $H^s(\mathbb{R}^d)$ to $H^{1-s}(\mathbb{R}^d)$ we further get

$$\begin{aligned} \|F\|_{H^{s-1}(\mathbb{R}^d)} \quad &\le \quad \|f\chi_0\|_{H^{s-1}(\mathbb{R}^d)} + \|a\nabla u \cdot \nabla\chi_0\|_{H^{s-1}(\mathbb{R}^d)} + \|\mathrm{div}(au\nabla\chi_0)\|_{H^{s-1}(\mathbb{R}^d)} \\ &\lesssim \quad \|f\|_{H^{s-1}(D)} + a_{\max}|u|_{H^1(D)} + |a|_{\mathcal{C}^t(D)}\|u\|_{L^2(D)} + a_{\max}\|u\|_{H^s(D)} \\ &\lesssim \quad \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}}\|f\|_{H^{s-1}(D)}. \end{aligned}$$

We can now apply Lemma 3.2 with $A = \overline{a}I_d$ and $w = u_0$ to show that $u_0 \in H^{1+s}(\mathbb{R}^d)$ and

$$\|u_0\|_{H^{1+s}(\mathbb{R}^d)} \;\lesssim\; \frac{1}{a_{\min}}\Big(|\overline{a}|_{\mathcal{C}^t(\overline{D})}|u_0|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\Big) + \|u_0\|_{H^1(\mathbb{R}^d)} \;\lesssim\; \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a^2_{\min}}\,\|f\|_{H^{s-1}(D)}\,.$$

The hidden constant depends on the choices of χ_0 and ψ and on the constant in Poincaré's inequality, which depends on the shape and size of D, but not on a. $\square$

Let us now treat the case of u_i, $1 \le i \le m$.

Lemma A.5. *For* $1 \le i \le m$, $u_i \in H^{1+s}(D)$ *and*

$$\|u_i\|_{H^{1+s}(D)} \;\lesssim\; \frac{a_{\max}\|a\|_{\mathcal{C}^t(\overline{D})}}{a^3_{\min}}\|f\|_{H^{s-1}(D)}.$$

Proof. Similarly to the proof of the previous Lemma, $u_i \in H_0^1(D \cap D_i)$ is the weak solution of a new problem $-\mathrm{div}(a\nabla u_i) = g_i$ in $\mathcal{D}'(D \cap D_i)$ with

$$g_i \;:=\; f\chi_i + a\nabla u \cdot \nabla \chi_i + \mathrm{div}(au\nabla\chi_i).$$

As in Lemma A.4, since again div is linear and continuous from H^s to H^{s-1}, we can establish that $g_i \in H^{s-1}(D \cap D_i)$ and

$$\|g_i\|_{H^{s-1}(D\cap D_i)} \;\lesssim\; \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}} \|f\|_{H^{s-1}(D)}.$$

Now let $Q = \{(y', y_d) \in \mathbb{R}^{d-1} \times \mathbb{R} : |y'| < 1 \text{ and } |y_d| < 1\}$, $Q_0 = \{(y', y_d) \in \mathbb{R}^{d-1} \times \{0\} |\, \|y'\| < 1\}$ and $Q_+ = Q \cap \mathbb{R}^d_+$. For $1 \leq i \leq p$, let α_i be a bijection from D_i to Q such that $\alpha_i \in \mathcal{C}^2(\overline{D}_i)$, $\alpha_i^{-1} \in \mathcal{C}^2(\overline{Q})$, $\alpha_i(D_i \cap D) = Q_+$ and $\alpha_i(D_i \cap \partial D) = Q_0$.

For all $y \in Q_+$, we define $w_i(y) := u_i(\alpha_i^{-1}(y)) \in H_0^1(Q_+)$ with $\nabla w_i(y) = J_i^{-t}(y)\nabla u_i(\alpha_i^{-1}(y))$, where $J_i(y) := D\alpha_i(\alpha_i^{-1}(y)))$ is the Jacobian of α_i. Furthermore, for $x \in D_i \cap D$ and $\varphi \in H_0^1(Q_+)$, we define $v(x) := \varphi(\alpha_i(x))$. Then $v \in H_0^1(D_i \cap D)$ and $\nabla v(\alpha_i^{-1}(y)) = J_i^t(y)\nabla\varphi(y)$, for all $y \in Q_+$, so that

$$\begin{aligned}\int_{D_i\cap D} a(x)\nabla u_i(x) \cdot \nabla v(x)\,\mathrm{d}x &= \int_{Q_+} a(\alpha_i^{-1}(y))\,\nabla u_i(\alpha_i^{-1}(y)) \cdot \nabla v(\alpha_i^{-1}(y))\,|\mathrm{det}J_i(y))|^{-1}\,\mathrm{d}y \\ &= \int_{Q_+} A_i(y)\,\nabla w_i(y) \cdot \nabla\varphi(y)\;\mathrm{d}y\,,\end{aligned}$$

where

$$A_i(y) := a(\alpha_i^{-1}(y))\,|\mathrm{det}J_i(y))|^{-1}(J_i\,J_i^t)(y) \in S_d(\mathbb{R}).$$

We define $F_i \in H^{s-1}(Q_+)$ by

$$\langle F_i, \varphi\rangle_{H^{s-1}(Q_+),H_0^{1-s}(Q_+)} \;:=\; \langle g_i, \varphi \circ \alpha_i\rangle_{H^{s-1}(D_i\cap D),H_0^{1-s}(D_i\cap D)}, \qquad \text{for all } \;\varphi \in H_0^{1-s}.$$

Indeed, since we assumed that α_i and α_i^{-1} are in $\mathcal{C}^2$, we have $\varphi \circ \alpha_i \in H_0^{1-s}(D_i \cap D)$ and moreover $\|\varphi \circ \alpha_i\|_{H^{1-s}(D_i\cap D)} \lesssim \|\varphi\|_{H^{1-s}(Q_+)}$ (cf. [18, Theorems 6.2.17 and 6.2.25(g)]), which implies that $F_i \in H^{s-1}(Q_+)$ and

$$\|F_i\|_{H^{s-1}(Q_+)} \;\lesssim\; \|g_i\|_{H^{s-1}(D\cap D_i)}.$$

We finally get that $v_i \in H_0^1(Q_+)$ solves

$$\int_{Q_+} A_i\nabla v_i \cdot \nabla\varphi\;\mathrm{d}y \;=\; \langle F_i, \varphi\rangle_{H^{s-1}(Q_+),H_0^{1-s}(Q_+)} \qquad \text{for all } \;\varphi \in H_0^1(Q_+).$$

In order to apply Lemma 3.3 we check first that $A_i \in \mathcal{C}^t(\overline{Q_+}, S_d(\mathbb{R}))$ and that it is coercive, and then define an extension of A_i to $\mathbb{R}^d_+$. Recalling that α_i is a $\mathcal{C}^2$–diffeomorphism from $D_i \cap D$ to Q_+, with $\alpha_i^{-1} \in \mathcal{C}^2(\overline{Q_+})$, we have for any $y \in Q_+$ and $\xi \in \mathbb{R}^d$:

- Coercivity: $A_i(y)\xi \cdot \xi \;=\; a(\alpha_i^{-1}(y))|\mathrm{det}J_i(y)|^{-1}\,|J_i^t(y)\xi|^2 \;\gtrsim\; a_{\min}|\xi|^2$. Hence $A_{\min} \gtrsim a_{\min}$.
- Boundedness:

$$A_{\max} \;:=\; \|A_i\|_{\mathcal{C}^0(\overline{Q_+},S_d(\mathbb{R}))} \;=\; \max_{x\in\overline{D_i\cap D}} a(x)\,\||\mathrm{det}J_i|^{-1}J_iJ_i^t\|_{\mathcal{C}^0(\overline{Q_+},S_d(\mathbb{R}))} \;\lesssim\; a_{\max}\,.$$

- Regularity: $A_i \in \mathcal{C}^t(\overline{Q_+}, S_d(\mathbb{R}))$ and

$$\begin{aligned}\|A_i\|_{\mathcal{C}^t(\overline{Q_+},S_d(\mathbb{R}))} \;&\leq\; a_{\max}\,\||\mathrm{det}J_i|^{-1}J_iJ_i^t\|_{\mathcal{C}^t(\overline{Q_+},S_d(\mathbb{R}))} \\ &\quad+\; |a|_{\mathcal{C}^t(\overline{D})}\,\||\mathrm{det}J_i|^{-1}J_iJ_i^t\|_{\mathcal{C}^0(\overline{Q_+},S_d(\mathbb{R}))} \;\lesssim\; \|a\|_{\mathcal{C}^t(\overline{D})}\,.\end{aligned}$$

We now extend A_i to $\mathbb{R}^d_+$. Since we assumed that $\operatorname{supp}(\chi_i)$ is compact in D_i, we can choose Q_i and $\tilde{Q}_i$ such that $\operatorname{supp}(v_i) \subset Q_i \subset \overline{Q_i} \subset \widetilde{Q}_i \subset \overline{\widetilde{Q}_i} \subset Q$ and consider $\psi \in \mathcal{C}^\infty(\mathbb{R}^d, [0,1])$ such that $\psi = 0$ on Q_i and $\psi = 1$ on $\overline{\widetilde{Q}_i}^c$. We define the extension $\overline{A}_i$ of A_i on $\mathbb{R}^d_+$ by

$$\overline{A}_i(x) := \begin{cases} A_i(x)(1-\psi(x)) + a_{\min}\psi(x)I_d & \text{if } x \in Q_+ \\ a_{\min}\psi(x)I_d & \text{if } x \in Q_+^c \end{cases}$$

Analogously to the case of $\overline{a}$ in Lemma A.4, we can deduce that, for any $y \in \mathbb{R}^d_+$ and $\xi \in \mathbb{R}^d$,

$$\overline{A}_i(y)\xi\cdot\xi \gtrsim a_{\min}|\xi|^2, \quad A_{\max} = \|\overline{A}_i\|_{\mathcal{C}^0(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))} \lesssim a_{\max} \text{ and } \|\overline{A}_i(y)\|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))} \lesssim \|a\|_{\mathcal{C}^t(\overline{D})}. \quad (A.6)$$

We now define an extension of F_i on $\mathbb{R}^d_+$. Note again that we can choose an open set G_i such that $\operatorname{supp}(F_i) \subset G_i \subset \overline{G}_i \subset Q$ and extend F_i to all of $\mathbb{R}^d_+$ such that $\|F_i\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim \|F_i\|_{H^{s-1}(Q_+)}$. Finally we continue v_i by 0 on $\mathbb{R}^d_+$, which yields $v_i \in H^1_0(\mathbb{R}^d_+)$. Moreover, since $\overline{A}_i = A_i$ on $\operatorname{supp}(v_i) \subset Q_i$, v_i is then the weak solution on $\mathbb{R}^d_+$ of

$$-\operatorname{div}(\overline{A}_i(x)\nabla v_i(x)) = F_i(x),$$

which enables us to apply Lemma 3.3 and to obtain that $v_i \in H^{1+s}(\mathbb{R}^d_+)$ and

$$\|v_i\|_{H^{1+s}(\mathbb{R}^d_+)} \quad \lesssim \quad \frac{A_{\max}}{A^2_{\min}}\Big(|\overline{A}_i|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))}\,|v_i|_{H^1(\mathbb{R}^d_+)} + \|F_i\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) \;+\; \frac{A_{\max}}{A_{\min}}\|v_i\|_{H^1(\mathbb{R}^d_+)}.$$

Recalling that $u_i(x) = v_i(\alpha_i(x))$ for any $x \in D \cap D_i$ and using the bounds in (A.6), as well as the transformation theorem [18, Theorem 6.2.17], we finally get

$$\begin{aligned} \|u_i\|_{H^{1+s}(D)} \quad &\lesssim \quad \frac{a_{\max}}{a^2_{\min}}\Big(\|a\|_{\mathcal{C}^t(\overline{D})}\,\|u\|_{H^1(D\cap D_i)} + \|g_i\|_{H^{s-1}(D\cap D_i)}\Big) \;+\; \frac{a_{\max}}{a_{\min}}\|u\|_{H^1(D)} \\ &\lesssim \quad \frac{a_{\max}\,\|a\|_{\mathcal{C}^t(\overline{D})}}{a^3_{\min}}\,\|f\|_{H^{s-1}(D)}. \end{aligned}$$

□

The result in Proposition 3.1 follows directly from Lemmas A.4 and A.5, if we recall that $u = \sum_{i=0}^{m} u_i$.

References

[1] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3):1005–1034, 2007.

[2] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.

[3] A. Barth, C. Schwab, and N. Zollinger. Multi–level Monte Carlo finite element method for elliptic PDE's with stochastic coefficients. SAM Research Reports 2010-18, ETH Zürich, 2010.

[4] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[5] 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/`.

[6] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[7] K.A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 2011. accepted for publication.

[8] 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.

[9] M. Dashti and A. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. Technical Report arXiv:1102.0143, `arXiv.org`, 2010. Submitted to SIAM Journal on Numerical Analysis.

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

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

[12] P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):205–228, 2005.

[13] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity. *SIAM Journal on Numerical Analysis*, 47:3624–3651, 2009.

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

[15] 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.

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

[17] C. J. Gittelson. Stochastic Galerkin discretization of the log-normal isotropic diffusion problem. *Math. Models Methods Appl. Sci.*, 20(2):237–263, 2010.

[18] W. Hackbusch. *Elliptic differential equations: Theorey and numerical treatment*, volume 18 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010.

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

[20] S. Zhang. Analysis of finite element domain embedding methods for curved domains using uniform grids. *SIAM Journal on Numerical Analysis*, 46:2843–2866, 2008.