Eidgenössische Technische Hochschule Zürich

Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Multi-Level Monte Carlo Finite Element method for elliptic PDE's with stochastic coefficients[1]

A. Barth, C. Schwab and N. Zollinger

Research Report No. 2010-18
June 2010

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

[1]The authors wishes to express their thanks to Roman Andreev and Claude J. Gittelson for fruitful discussions and helpful comments. The research partially supported under ERC AdG Grant STAHDPDE No. 247277

MULTI–LEVEL MONTE CARLO FINITE ELEMENT METHOD FOR ELLIPTIC PDES WITH STOCHASTIC COEFFICIENTS

ANDREA BARTH, CHRISTOPH SCHWAB, AND NATHANIEL ZOLLINGER

Abstract. It is a well–known property of Monte Carlo methods that quadrupling the sample size halves the error. In the case of simulations of a stochastic partial differential equations, this implies that the total work is the sample size times the discretization costs of the equation. This leads to a convergence rate which is impractical for many simulations, namely in finance, physics and geosciences. With the Multi–level Monte Carlo method introduced herein, the overall work can be reduced to that of the discretization of the equation, which results in the same convergence rate as for the standard Monte Carlo method. The model problem is an elliptic equation with stochastic coefficients. Multi-level Monte Carlo errors and work estimates are given both for the mean of the solutions and for higher moments. Numerical examples complete the theoretical analysis.

1. Introduction

Monte Carlo methods are widely used in statistical simulation. In the case of partial differential equations with random inputs, "sampling" entails the numerical solution of a deterministic partial differential equation (PDE). For time dependent, parabolic problems driven by noise (see, e.g. [3, 4, 5, 22, 15]), numerous paths must be simulated. Here, we are concerned with Monte Carlo methods (MC methods) for elliptic problems where the source of randomness lies in the coefficients. Such problems arise prominently in the numerical simulation of subsurface flow problems (see, e.g., [25, 26] and the references therein). Some key characteristics of elliptic problems with stochastic coefficients, which arise in computational geosciences, are the low spatial regularity of the permeability samples, the small spatial correlation lengths (this implies slow convergence of Karhúnen-Loève expansions), and, more challenging, the possible nonstationarity of realistic stochastic models. All these factors hinder the efficient numerical simulation of such problems. In order to deal with these difficulties, we propose a Multi–level Monte Carlo method (MLMC method). This family of methods was introduced, to the authors' knowledge, by M. Giles in [18, 17] for Itô stochastic ordinary differential equations after earlier work by A. Heinrich on numerical quadrature (see [21]).

As a model problem of the class of partial differential equations described above, we consider the following elliptic model problem

$$-\mathrm{div}(a\nabla u) = f \qquad \text{in } D,$$

Date: August 24, 2010.

Key words and phrases. Multi–level Monte Carlo, Stochastic partial differential equations, Stochastic finite element methods, Multi–level approximations.

The authors wish to express their thanks to Roman Andreev and Claude J. Gittelson for fruitful discussions and helpful comments. The research partially supported under ERC AdG Grant STAHDPDE No. 247277.

where $D \subset \mathbb{R}^d$, $d = 1, 2, \ldots$. We impose mixed Dirichlet and Neumann boundary conditions and assume f to be in $L^2(D)$. The coefficient a is a correlated random field over the spatial domain D. The random field solution u of Problem (1) not exclusively depends on $x \in \mathbb{R}^d$, but also on a stochastic parameter $\omega \in \Omega$, where Ω is the stochastic state space, which we specify later. For a fixed ω, we face an elliptic PDE with a space dependent coefficient. This can be solved by various numerical methods, for instance by a Galerkin approximation, which leads to a Finite Element method (FE method). We shall be interest in the computation of moments of the stochastic solution, which can be numerically calculated by a Monte Carlo method (MC method). When the sample size increases, convergence is ensured by the law of large numbers: A quadrupling of the sample size halves the error of the approximation. To estimate the moments of the solution of Equation (1), we must solve the deterministic equation for each sample. The cost of this algorithm is the number of samples times the cost for the approximation in the space domain (in the case of a FE method this corresponds to the degrees of freedom).

One of the aims of the Multi–level Monte Carlo Finite Element method (MLMC–FE method), proposed here, is to decrease the cost of this computation. In order to do so, we introduce a hierarchical system of Finite Element spaces (FE spaces), in each of which we calculate a certain number of samples of the approximation of the solution. Dealing with a large number of samples on a very coarse grid is computationally cheap, but convergence in space is rather poor. On the other hand, on a fine grid the convergence in the space variable x is fast, but solving the system of equations for each sample is expensive. The MLMC method exploits this fact. The number of samples solved is inversely proportional to the fineness of the grid. This strategy allows us to have the same overall convergence as the MC method on the finest grid, but the computational costs are only a fraction of the latter. With the use of a full Multigrid solver, the computational costs of the MLMC method are log–linear in $\mathbb{R}^d$, for $d > 1$. Since the error of the MLMC method balances the error of the space approximation and the error of the MC approximation, the sample size should be increased if the space approximation is "too good" or the solution "too smooth". Depending on the dimension of the physical domain, the approximation with linear Finite Elements becomes, in terms of degrees of freedom, less efficient.

We further our calculations to approximate higher moments of the solution, with a similar approach as above. We introduce a Wavelet transformation, since the tensor product of the solution may then be estimated with linear complexity. This leads to a sparse tensor MLMC–FE method, which exhibits once more log–linear complexity. We show that the solution exhibits a certain "mix" regularity, which takes the form of r–summability of the stochastic solution as a Bochner function in a scale $\{X_s\}_{s\geq 0}$ of Sobolev spaces on the domain D.

This paper is structured as follows. In the second chapter we present all the preliminaries. This is followed by the formulation of our model problem, where we also study the well posedness and certain regularity conditions of the solution. In the fourth chapter we analyze the rate of convergence of the Multi–level Monte Carlo method. We give convergence rates of the Monte Carlo approximations for the continuous solution and its Galerkin Finite Element approximation. Chapter 5 contains the extension of our previous results to the approximation of higher order moments of the solution. Here we derive rates of convergence for the sparse tensor Multi–level Monte Carlo method for the k–th moment, $1 \leq k \in \mathbb{N}$, of the solution. Subsequently we present the numerical analysis of some examples in one and two space dimensions.

2. Preliminaries

For the variational formulation as well as for our error analysis of the MLMC–FE method for the problem at hand, given by Equations (1), we shall require Bochner spaces of r-summable functions in $D \subset \mathbb{R}^d$, for $d = 1, 2, \ldots$. To this end, for any Banach space B of real-valued functions on the domain D with norm $\|\cdot\|_B$, we denote the set of strongly measurable, r–summable mappings $v : \Omega \to B$ by

$$L^r(\Omega, \mathcal{A}, \mathbb{P}; B) := \left\{ v : \Omega \to B \quad | \quad v \text{ strongly measurable}, \quad \|v\|_{L^r(\Omega;B)} < \infty \right\},$$

where, for $0 < r \leq \infty$,

$$\|v\|_{L^r(\Omega;B)} := \begin{cases} \left(\int_\Omega \|v(\omega, \cdot)\|_B^r d\mathbb{P}(\omega) \right)^{1/r} & \text{if} \quad 0 < r < \infty, \\ \\ \operatorname{esssup}_{\omega \in \Omega} \|v(\omega, \cdot)\|_B & \text{if} \quad r = \infty \, . \end{cases}$$

Here we introduced a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where, as usual, Ω denotes a set elementary events, $\mathcal{A} \subset 2^\Omega$ the σ-algebra of all possible events and where $\mathbb{P} : \mathcal{A} \to [0, 1]$ is a probability measure. Let $\mathcal{B} \in \mathcal{L}(X, Y)$ denote a continuous linear mapping from X to another separable Hilbert space Y. For a random field $x \in L^r(\Omega; X)$ this mapping defines a random variable $y(\omega) = \mathcal{B}x(\omega)$, and we have that $y \in L^r(\Omega; Y)$ and

$$\|\mathcal{B}x\|_{L^r(\Omega;Y)} \leq C \|x\|_{L^r(\Omega;X)}.$$

Furthermore, there holds

$$\mathcal{B} \int_\Omega x \, dP(\omega) = \int_\Omega \mathcal{B}x \, dP(\omega).$$

We refer to Chapter 1 of [13] for a synopsis of these and further results of Banach space valued random variables.

3. Model elliptic problem with stochastic coefficients

In the bounded Lipschitz polyhedron $D \subset \mathbb{R}^d$, $d = 1, 2, 3, \ldots$, we consider the elliptic diffusion problem with stochastic diffusion coefficient a

$$-\operatorname{div}(a \nabla u) = f \qquad \text{in } D. \tag{3.1}$$

Here, $f \in L^2(D)$ is a given source term. We assume that the Lipschitz boundary $\Gamma = \partial D$ is partitioned into a finite union of $d-1$–dimensional planes, which in turn are grouped into a Dirichlet part Γ_D and a Neumann part Γ_N. We assume in addition that

$$|\Gamma_N| \geq 0 \text{ and } |\Gamma_D| > 0. \tag{3.2}$$

Furthermore, the exterior unit normal vector $\vec{n}$ to Γ exists almost everywhere on Γ. Equation (3.1) is completed by the boundary conditions

$$\gamma_0 u := u|_{\Gamma_D} = 0, \qquad \gamma_{1,a} u := (a \, \vec{n} \cdot \nabla u)|_{\Gamma_N} = g, \tag{3.3}$$

where g is a given normal flux on Γ_N (specific assumptions on g will be given below). In the case of the Laplacean, i.e. when $a = 1$, we write γ_1 in place of $\gamma_{1,a}$.

To ensure well–posedness of our problem, we require that the following assumption on the stochastic diffusion coefficient a is fulfilled:

Assumption 3.1. The stochastic diffusion coefficients $a(\omega, x)$ in Equation (3.1), and Equation (3.3) is assumed to be a strongly measurable mapping from Ω into $L^\infty(D)$.

There exist constants $0 < a_- < a_+ < \infty$ such that the random coefficient $a(\omega, x)$ in Equation (3.1) is uniformly elliptic, i.e. for *every* $\omega \in \Omega$ holds[1]

$$0 < a_- \leq \mathrm{essinf}_{x\in D} a(\omega, x) \leq \|a(\omega, \cdot)\|_{L^\infty(D)} \leq a_+ < \infty. \tag{3.4}$$

We remark that for Lipschitz domains D the trace operator γ_0 in Equation (3.3) is well-defined and continuous from $H^1(D)$ onto $H^{1/2}(\Gamma_D)$.

For the normal derivative operator $\gamma_{1,a}$, we have

Lemma 3.2. *Under Assumption 3.1, for $f \in L^2(D)$ and every $0 < r \leq \infty$, the co–normal derivative operator $\gamma_{1,a}$ in Equation* (3.3) *is a well-defined and surjective linear operator from $L^r(\Omega; H^1_{\Gamma_D}(D, \Delta))$ onto $L^r(\Omega; H^{-1/2}(\Gamma_N))$ where*

$$H^1_{\Gamma_D}(D, \Delta) := \{v \in H^1(D) : \gamma_0 v = 0, \quad \Delta v \in L^2(D)\}$$

and $H^{-1/2}(\Gamma_N) := (H^{1/2}_{00}(\Gamma_D))^*$ (with duality being understood with respect to the "pivot" space $L^2(\Gamma)$; see [24] *for the definition of $H^{1/2}_{00}(\Gamma_D)$).*

3.1. **Variational Formulation and Well-Posedness.** To present the variational formulation of Equation (3.1) we introduce the Hilbert space

$$V = H^1_{\Gamma_D}(D) = \{v \in H^1(D) : \gamma_0 v = 0\}. \tag{3.5}$$

Due to the assumption $|\Gamma_D| > 0$, by the continuity of the trace operator γ_0 the space V is a closed, linear subspace of $H^1(D)$ and by the Poincaré inequality the expression

$$V \ni v \to \|v\| := \left(\int_D |\nabla v|^2 dx\right)^{1/2}$$

is a norm on V. We identify $L^2(D)$ with its dual and denote by V^* the dual of V with respect to the "pivot" space $L^2(D)$, i.e. we work in the triplet $V \subset L^2(D) \simeq L^2(D)^* \subset V^*$.

To derive the variational formulation of the stochastic elliptic boundary value problem, given by Equation (3.1) – (3.3) we fix $\omega \in \Omega$ for the moment. We then multiply Equation (3.1) by a test function $v \in L^2(\Omega; V)$ and integrate by parts in D to obtain (for fixed $\omega \in \Omega$) the (formal) integral identity

$$\int_D a\nabla v \cdot \nabla u dx = \int_D f v dx + \int_{\Gamma_N} g\gamma_0 v ds.$$

Taking expectations on both sides of this expression, we arrive at the *weak formulation* of the stochastic elliptic boundary value problem (Equations (3.1) – (3.3)):

given a satisfies Assumption 3.1, $f \in L^2(\Omega; V^*)$ and $g \in L^2(\Omega; H^{-1/2}(\Gamma_N))$, which are mutually independent, find $u \in L^2(\Omega; V)$ such that

$$B(u, v) = F(v) \quad \forall v \in L^2(\Omega; V), \tag{3.6}$$

where the bilinear form $B(\cdot, \cdot) : L^2(\Omega; V) \times L^2(\Omega; V) \to \mathbb{R}$ is given by

$$B(u, v) = \mathbb{E}\left[\int_D a(\cdot, x)\nabla u(\cdot, x) \cdot \nabla v(\cdot, x) dx\right],$$

[1]We assume that the random coefficient a is, possibly after modification of a given a on a null–set, well-defined and computationally accessible for *every* $\omega \in \Omega$

and

$$F(v) = \mathbb{E}\left[\int_D f(\cdot,x)v(\cdot,x)dx\right] + \mathbb{E}\left[\int_{\Gamma_N} g\gamma_0 v(\cdot,x)ds_x\right],$$

where the 'integrals' $\int_D f(\cdot,x)v(\cdot,x)dx$ and $\int_{x\in\Gamma_N}\ldots ds_x$ understood as $L^2(\Omega;V)\times L^2(\Omega;V^*)$ respectively as $L^2(\Omega;H^{1/2}_{00}(\Gamma_N))\times L^2(\Omega;H^{-1/2}(\Gamma_N))$ duality pairings obtained by extending the corresponding L^2 inner products by continuity. By Riesz Representation Theorem there exists a linear operator $A(\omega)\in\mathcal{L}(V,V^*)$ such that for all $v,w\in V$

$$B(v,w) = {}_V\langle w, A(\omega)v\rangle_{V^*}. \tag{3.7}$$

Theorem 3.3. *Under Assumption 3.1, for every $f\in L^2(\Omega;V^*)$ and $g\in L^2(\Omega;H^{-1/2}(\Gamma_N))$, the weak formulation, Equation* (3.6)*, of the stochastic elliptic boundary value problem, given by equation* (3.1) – (3.3)*, admits a unique solution $u\in L^2(\Omega;V)$.*

Proof. By Assumption 3.1, we have for every $v,w\in V$

$$|B(v,w)| \le \operatorname{esssup}_{\omega\in\Omega}\|a(\cdot,x)\|_{L^\infty(D)}\|v\|_{L^2(\Omega;V)}\|w\|_{L^2(\Omega;V)} \le a_+\|v\|_{L^2(\Omega;V)}\|w\|_{L^2(\Omega;V)} \tag{3.8}$$

and

$$B(v,v) \ge a_-\|v\|^2_{L^2(\Omega;V)}. \tag{3.9}$$

Moreover, for given $g\in L^2(\Omega;H^{-1/2}(\Gamma_N))$ and $f\in L^2(\Omega;V^*)$, we have by the Cauchy-Schwarz and Poincaré inequalities and the continuity of the trace operator γ_0 that for every $w\in V$

$$\begin{aligned} |F(w)| &\le \|f\|_{L^2(\Omega;V^*)}\|w\|_{L^2(\Omega;V)} + \|g\|_{L^2(\Omega;H^{-1/2}(\Gamma_N))}\|\gamma_0 w\|_{L^2(\Omega;H^{1/2}(\Gamma_D))} \\ &\le C(D)\left(\|f\|^2_{L^2(\Omega;V^*)} + \|g\|^2_{L^2(\Omega;H^{-1/2}(\Gamma_N))}\right)^{1/2}\|w\|_{L^2(\Omega;V)}. \end{aligned} \tag{3.10}$$

The assertion now follows from the Lax-Milgram Lemma. □

Remark 3.4. The variational formulation, Equation (3.6), requires in Assumption 3.1 only the definition and boundedness of the random coefficient a $\mathbb{P}$-a.s.. The (stronger) Assumption 3.1 implies in particular the unique solvability of the stochastic diffusion problem, defined in Equation (3.1) for *every* sample ω; this is required for the MLMC–FE simulation. Assumption 3.1 in addition also implies

$$\forall\omega\in\Omega:\quad \|u(\omega,\cdot)\|_V \le \frac{1}{a_-}\left(\|f(\omega,\cdot)\|^2_{V^*} + \|g(\omega,\cdot)\|^2_{H^{-1/2}(\Gamma_N)}\right)^{1/2}. \tag{3.11}$$

3.2. **Regularity of Solutions.** To ensure local $H^2(D)$ regularity and the existence of higher moments of the stochastic solution $u\in L^2(\Omega;V)$ we impose additional assumptions on the data f and g:

Assumption 3.5. We assume that $f\in L^r(\Omega;L^2(D))$, $g\in L^r(\Omega;H^{1/2}(\Gamma_N))$, for some $2\le r\le\infty$, and that the mapping $\Omega\ni\omega\to a(\omega,\cdot)$ takes values in $W^{1,\infty}(D)$ for *every* $\omega\in\Omega$. Moreover, we assume that the sources of randomness, i.e. a, f and (if $|\Gamma_N|>0$) g are independent and strongly measurable as mappings taking values in the respective Banach spaces $W^{1,\infty}(D)$, $L^2(D)$ and in $H^{1/2}(\Gamma_N)$.

By the usual elliptic regularity theory (see, e.g. [16]), Assumption 3.5 ensures in particular that $u\in H^2_{loc}(D)$, $\mathbb{P}$-a.s.. We have the following

Proposition 3.6. *Under Assumption 3.5 and by Equation* (3.4)*, the elliptic problem, given by Equation* (3.1) – (3.3)*, admits a unique solution* $u \in L^r(\Omega; W)$*. Here, the space* W *is defined by*

$$W := \{w \in V : \Delta w \in L^2(D), \quad \gamma_0 w = 0, \quad \gamma_1 w \in H^{1/2}(\Gamma_N)\},$$

equipped with the norm $\|\cdot\|_W$ *given by*

$$\|w\|_W := \|\Delta w\|_{L^2(D)} + \|w\|_{L^2(D)}.$$

Further, with $2 \le r \le \infty$ *as in Assumption 3.5, there holds the* a-priori estimate

$$\|u\|_{L^r(\Omega;W)} \le C(a) \left(\|f\|_{L^r(\Omega;L^2(D))} + \|g\|_{L^r(\Omega;H^{1/2}(\Gamma_N))} \right) . \tag{3.12}$$

Here, $C(a)$ *depends on* a_- *and* a_+*, resp. on* $\|a\|_{L^\infty(\Omega;W^{1,\infty}(D))}$*.*

Proof. The proof is a consequence of the $W^{1,\infty}(D)$–regularity of all realizations of the stochastic coefficient a which implies that the stochastic solution $u \in L^2(\Omega; V)$ satisfies the identity

$$-\Delta u(\omega, \cdot) = f(\omega, \cdot) + \nabla a(\omega, \cdot) \cdot \nabla u(\omega, \cdot) \quad \text{in} \quad L^2(D), \forall \omega \in \Omega.$$

Therefore we may estimate for every $\omega \in \Omega$

$$\|\Delta u(\omega, \cdot)\|_{L^2(D)} \le C(a) \left(\|f(\omega, \cdot)\|_{L^2(D)} + \|g(\omega, \cdot)\|_{H^{1/2}(\Gamma_N)} \right), \qquad \omega \in \Omega.$$

Adding the corresponding $L^2(D)$ bound (which results from Equation (3.11) and the Poincaré-inequality), raising both sides of the resulting bound on the $\|\cdot\|_W$ norm of u to the power r and taking expectations implies the assertion. □

We remark that the space W can be characterized as a weighted Sobolev space with weights vanishing at vertices and (in case $d = 3$) at edges of the polyhedron D; see, e.g. [20].

In the following section we introduce the Galerkin projections our Finite Element method will be based on. We prove convergence of the resulting discrete problem by a Monte Carlo method, before we proceed with the convergence and a work estimate for the MLMC method for the discrete equation.

4. Multi–level Monte Carlo Finite Element method

A key ingredient in MLMC–FE method are pathwise, *hierarchic* Finite Element discretizations of the stochastic elliptic problem (Equation (3.1)) which we present next. Followed by an error estimate for the Monte Carlo method of the (non discrete) solution of the problem at hand. From this result we derive a convergence rate for the MC method of the discrete solution (full tensor MC–FE method) and finally for the Multi–level MC–FE method.

4.1. Mean Square Stability of the Galerkin Projection.

The Finite Element method which we consider is based on sequences of regular simplicial meshes of quasi–uniform triangles or tetrahedra $\{\mathcal{T}_l\}_{l=0}^\infty$ of the polygonal respectively polyhedral domain D. For any $l \ge 0$, we denote the meshwidth of $\mathcal{T}_l$ by

$$h_l = \max_{K \in \mathcal{T}_l} \{\mathrm{diam}(K)\} =: \max_{K \in \mathcal{T}_l} \{h_K\}.$$

We recall (see, e.g., [8, 9]) that the nested family $\{\mathcal{T}_l\}_{l=0}^\infty$ of regular, simplicial meshes is called *κ-shape regular* if and only if there exists a $\kappa < \infty$ such that $\kappa := \sup_l \kappa_l = \sup_l \max_{K \in \mathcal{T}_l} \frac{h_K}{\rho_K}$. Here ρ_K is the radius of the largest ball that can be inscribed into any $K \in \mathcal{T}_l$.

The uniform refinement of the mesh is achieved by regular subdivision. This results in the meshwidth $h_l = 2^{-l}h_0$, since $h_{l+1} = \frac{1}{2}h_l$, where h_0 is the maximal diameter of the coarsest mesh. Constructed like this the nested family $\{\mathcal{T}_l\}_{l=0}^{\infty}$ is κ –shape regular, since $\kappa_l = \kappa_0 = \kappa$.

For $p \geq n \geq 0$ we define the Finite Element spaces as

$$\mathcal{S}^{p,n}(D,\mathcal{T}) = \{v \in H^n(D) : v|_K \in \mathcal{P}_p, \quad \forall K \in \mathcal{T}\}\,, \tag{4.1}$$

where we denote by $\mathcal{P}_p(K) = \text{span}\{x^\alpha : |\alpha| \leq p\}$ the space of polynomials of total degree $\leq p$ on a set K. Equivalent we denote by

$$\mathcal{S}^{p,n}_{\Gamma_D}(D,\mathcal{T}) = \{v \in H^n_{\Gamma_D}(D) : v|_K \in \mathcal{P}_p, \quad \forall K \in \mathcal{T}\}\,,$$

the corresponding Finite Element space respecting boundary conditions.

The family of FE spaces that we employ is $\mathcal{V} = \{\mathcal{S}^{1,1}_{\Gamma_D}(D,\mathcal{T}_l)\}_{l=0}^{\infty}$, which is the family of spaces of continuous, piecewise linear functions on the regular, simplicial triangulation $\{\mathcal{T}_l\}_{l=0}^{\infty}$ that satisfy the homogeneous essential boundary conditions on the Dirichlet boundary Γ_D (whose closure coincides, by assumption, with the union of all closed edges of elements $K \in \mathcal{T}_l$ abutting at Γ_D).

The Galerkin approximation is based on the weak formulation in Equation (3.6) of the stochastic elliptic boundary value problem, presented in Equation (3.1) – (3.3).

Since for each level l of mesh refinement, $V_l = \mathcal{S}^{1,1}_{\Gamma_D}(D,\mathcal{T}_l) \subset H^1_{\Gamma_D}(D)$ the corresponding discrete problem reads: find $u_l \in L^2(\Omega;\mathcal{S}^{1,1}_{\Gamma_D}(D,\mathcal{T}_l))$ such that

$$B(u_l, v_l) = F(v_l) \qquad \forall v_l \in L^2(\Omega;\mathcal{S}^{1,1}_{\Gamma_D}(D,\mathcal{T}_l)), \tag{4.2}$$

where $F(v)$ is defined as in Equation (3.6).

By Equation (3.8), Equation (3.9) and Equation (3.10), for each $l = 0, 1, 2, \ldots$ exists a unique stochastic FE solution $u_l \in L^2(\Omega;\mathcal{S}^{1,1}_{\Gamma_D}(D,\mathcal{T}_l))$.

The operator G_l projecting the variational solution $u \in L^2(\Omega;V)$ into the stochastic Finite Element solution $u_l \in L^2(\Omega;\mathcal{S}^{1,1}_{\Gamma_D}(D,\mathcal{T}_l))$ is an $L^2(\Omega;V)$-stable projection: by Equation (4.2) and Equation (3.9), we have for every l

$$a_-\|u_l\|^2_{L^2(\Omega;V)} \leq B(u_l,u_l) = B(u,u_l) \leq a_+\|u\|_{L^2(\Omega;V)}\|u_l\|_{L^2(\Omega;V)},$$

which implies

$$\|G_l v\|_{L^2(\Omega;V)} \leq \frac{1}{a_-}\|v\|_{L^2(\Omega;V)} \qquad \forall v \in L^2(\Omega;V).$$

Under Assumption 3.1, the Galerkin projection G_l is well-known to be quasioptimal (see, e.g. [8, 9]), i.e.

$$\|u - u_l\|_{L^2(\Omega;V)} \leq C_a \inf_{v_l \in V_l}\|u - v_l\|_{L^2(\Omega;V)}, \tag{4.3}$$

where $C_a = \sqrt{\frac{a_+}{a_-}}$.

Assumptions 3.1 and 3.5 imply local $H^2(D)$ –regularity of the solution. More precisely, with the space $W \subset V$ as defined in Equation (3.6), there exists a unique weak solution $u \in L^2(\Omega;W)$ and Equation (3.12) holds. Then, $C(a)$ depends on a_- and a_+ resp. on $\|a\|_{L^\infty(\Omega;W^{1,\infty}(D))}$. With the same assumptions and by well known results (see for example [8, 9]) we have for all $w \in W$

$$\inf_{v_l \in \mathcal{S}^1_{\Gamma_D}(D,\mathcal{T}_l)}\|w - v_l\|_{H^1_{\Gamma_D}(D)} \leq C_I 2^{-l}h_0\|w\|_W, \tag{4.4}$$

where $C_I > 0$ is some constant, independent of l.

We proceed with an analysis of the rate of convergence of the Monte Carlo method for the solution of the stochastic elliptic problem at hand. First we derive the estimate for the solution which is not discretized in space and then generalize this result to the Finite Element solution.

4.2. **Rate of convergence of the Monte Carlo method.** The stochastic solution is characterized by its moments. We estimate the expectation $\mathbb{E}[u] \in V$ by the mean over solution samples $\hat{u}^i \in V$, $i = 1, \ldots, M$ corresponding to M independent, identically distributed realizations of the random input data a, f and g:

$$E_M[u] := \frac{1}{M} \sum_{i=1}^{M} \hat{u}^i \in V. \tag{4.5}$$

The following result is a bound on the *statistical error* resulting from this Monte Carlo estimator.

Lemma 4.1. *For any $M \in \mathbb{N}$ and for $u \in L^2(\Omega; V)$ holds*

$$\|\mathbb{E}[u] - E_M[u]\|_{L^2(\Omega;V)} \leq M^{-1/2} \|u\|_{L^2(\Omega;V)}.$$

Proof. Let us denote by $\hat{u}_M$ the sample average over M samples. Defined as such $\hat{u}_M$ is a random variable that maps Ω into V. With the independence of the identically distributed samples it follows

$$\begin{aligned}
\|\mathbb{E}[u] - E_M[u]\|^2_{L^2(\Omega;V)} &= \mathbb{E}\left[\|\mathbb{E}[u] - \frac{1}{M} \sum_{i=1}^{M} \hat{u}^i\|_V^2 \right] = \frac{1}{M^2} \sum_{i=1}^{M} \mathbb{E}\left[\|\mathbb{E}[u] - \hat{u}^i\|_V^2 \right] \\
&= \frac{1}{M} \mathbb{E}\left[\|\mathbb{E}[u] - u\|_V^2 \right] = \frac{1}{M} (\mathbb{E}\|u\|_V^2 - \|\mathbb{E}[u]\|_V^2) \leq \frac{1}{M} \|u\|^2_{L^2(\Omega;V)}.
\end{aligned}$$

□

4.3. **Single–level Monte Carlo Finite Element method.** The implementation of the estimator $E_M[u]$ in Equation (4.5) requires a Finite Element approximation of the 'samples' $\hat{u}^i$ which we choose from a continuous, piecewise linear Finite Element space on a family of shape regular, affine and simplicial triangulations $\{\mathcal{T}_l\}_{l=0}^{\infty}$.

The key question which arises naturally here is which is the optimal choice of the sample size in dependence of the grid size to achieve a prescribed error level with minimal work.

We shall address this question under the following assumptions on the Finite Element method.

Assumption 4.2. For a given Finite Element mesh $\mathcal{T}_l$ from the family of meshes the FE solution for a given realization $\hat{a}^i \in W^{1,\infty}(D)$ of the stochastic coefficient which satisfies Assumption 3.1, the Galerkin projection $u_l = G_l u \in V_l$ on the Finite Element subspace $V_l = \mathcal{S}^{1,1}_{\Gamma_D}(D, \mathcal{T}_l)$ of dimension $N_l = \dim(\mathcal{S}^{1,1}_{\Gamma_D}(D, \mathcal{T}_l))$ can be realized in $\mathrm{O}(N_l)$ work and memory. The approximation has the accuracy

$$\|w - G_l w\|_V \leq C_a C_I h_l \|w\|_W,$$

where $h_l = 2^{-l} h_0 = \max_{K \in \mathcal{T}_l} \mathrm{diam}(K)$ denotes the meshwidth of $\mathcal{T}_l$.

We remark that for polygonal domains $D \subset \mathbb{R}^2$, Assumption 4.2 on the space W can be satisfied by standard Multilevel solvers for the Finite Element equations on families of meshes with suitable refinement towards the vertices of D. We now establish a first error estimate for the MC–FE method in the case when *the same Finite Element mesh $\mathcal{T}_l$ is used for all samples*: we estimate the expectation of the solution, $\mathbb{E}[u]$, by

$$E_M[u_l] := \frac{1}{M}\sum_{i=1}^{M} G_l \hat{u}^i \in \mathcal{S}^{1,1}(D, \mathcal{T}_l) \tag{4.6}$$

Theorem 4.3. *Under Assumptions 3.1 and 3.5 holds the error bound*

$$\|\mathbb{E}[u] - E_M[u_l]\|_{L^2(\Omega;V)} \leq C(a)\left(\frac{1}{\sqrt{M}} + h_l\right)\left(\|f\|_{L^2(\Omega;L^2(D))} + \|g\|_{L^2(\Omega;H^{3/2}(\Gamma_N))}\right). \tag{4.7}$$

Proof. We split the left hand side of the Equation above as follows

$$\begin{aligned}\|\mathbb{E}[u] - E_M[u_l]\|_{L^2(\Omega;V)} &\leq \|\mathbb{E}[u] - \mathbb{E}[u_l]\|_{L^2(\Omega;V)} + \|\mathbb{E}[u_l] - E_M[u_l]\|_{L^2(\Omega;V)} \\ &\leq \mathbb{E}\left[\|[u] - [u_l]\|_V\right] + \|\mathbb{E}[u_l] - E_M[u_l]\|_{L^2(\Omega;V)}.\end{aligned}$$

The first term on the right hand side is bounded by Assumption 4.2 and Proposition 3.6. The assertion follows with Lemma 4.1 for the second term. □

The optimal choice of sample size versus grid size for a fixed error is reached when the statistical and the discretization errors are equilibrated, i.e. when $M^{-\frac{1}{2}} = \mathrm{O}(h_l) = \mathrm{O}(2^{-l})$. In terms of the degrees of freedom of the Finite Element method, N_l, therefore, we obtain from Equation (4.7) the basic relation

$$M^{-\frac{1}{2}} = O(h_l) = O(N_l^{-\frac{1}{d}}). \tag{4.8}$$

We have a closer look at the computational cost of the Monte Carlo method. We work under Assumption 4.2 and take the same estimate as before, i.e. we estimate $\mathbb{E}[u]$ by the mean of $\hat{u}_l^i$, $i = 1, \ldots, M$ for M independent samples with the fixed discretization level l:

$$\mathbb{E}[u_l] \simeq E_M[u_l] = \frac{1}{M}\sum_{i=1}^{M} \hat{u}_l^i = \frac{1}{M}\sum_{i=1}^{M} G_l \hat{u}^i.$$

Under Assumption 4.2, the computational cost of this estimate is $\mathrm{O}(M \cdot N_l)$ work and memory, i.e. the number of samples times the cost for each Finite Element solution with $N_l = 2^{ld}$ degrees of freedom. With the previous calculation on the optimal sample size, i.e. Equation (4.8), which implies $M = N_l^{\frac{2}{d}} = \mathrm{O}(2^{2l})$, we may write for the computational cost $\mathrm{O}(2^{l(2+d)})$.

Subsequently we generalize these calculations to the case of a Multi-level approximation of the Monte Carlo method.

4.4. **Multi–level Monte Carlo Finite Element method.** For the MLMC method we discretize the variational formulation, given by Equation (3.6), by Galerkin projection onto a hierarchic sequence of finite dimensional sub–spaces

$$V_0(D) \subset V_1(D) \subset \ldots \subset V_l(D) \subset \ldots \subset V(D),$$

where $V_l(D) := \mathcal{S}^{1,1}_{\Gamma_D}(D, \mathcal{T}_l)$, here l denotes the level of mesh refinement. With the notation $u_0 := 0$ we may write

$$u_L = \sum_{l=1}^{L} (u_l - u_{l-1})$$

and, by linearity of the expectation operator $\mathbb{E}[\cdot]$,

$$\mathbb{E}[u_L] = \mathbb{E}\left[\sum_{l=1}^{L} (u_l - u_{l-1})\right] = \sum_{l=1}^{L} \mathbb{E}[u_l - u_{l-1}] = \sum_{l=1}^{L} \left(\mathbb{E}[u_l] - \mathbb{E}[u_{l-1}]\right).$$

In the MLMC–FE method, we estimate $\mathbb{E}[u_l - u_{l-1}]$ by a level dependent number M_l of samples, which implies that we may estimate $\mathbb{E}[u]$ by

$$E^L[u] := \sum_{l=1}^{L} E_{M_l}[G_l u - G_{l-1} u] = \sum_{l=1}^{L} \left(E_{M_l}[u_l] - E_{M_l}[u_{l-1}]\right). \tag{4.9}$$

Convergence of the MLMC–FE method is guaranteed by the following

Lemma 4.4. *Under Assumptions 3.1, 3.5 and 4.2, the MLMC–FE approximation, Equation* (4.9) *of the expectation* $\mathbb{E}[u]$ *of the solution* $u \in L^2(\Omega; W)$ *to the stochastic elliptic boundary value problem, presented in Equation* (3.1) – (3.3)*, in the polyhedral domain* $D \subset \mathbb{R}^d$ *admits the error bound*

$$\|\mathbb{E}[u] - E^L[u]\|_{L^2(\Omega;V)} \leq C \left(h_L + \sum_{l=1}^{L} h_l M_l^{-1/2}\right) \left(\|f\|_{L^2(\Omega;L^2(D))} + \|g\|_{L^2(\Omega;H^{1/2}(\Gamma_N))}\right). \tag{4.10}$$

Here, the constant C *depends only on* d, a_- *and on the bound* $\|a\|_{L^\infty(\Omega;W^{1,\infty}(D))}$ *in Assumption 3.5.*

Proof. We rewrite the error to be estimated as in the proof of Theorem 4.3 as

$$\begin{aligned}
\|\mathbb{E}[u] - E^L[u]\|_{L^2(\Omega;V)} &= \|\mathbb{E}[u] - \mathbb{E}[u_L] + \mathbb{E}[u_L] - \sum_{l=1}^{L} E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;V)} \\
&\leq \|\mathbb{E}[u] - \mathbb{E}[u_L]\|_{L^2(\Omega;V)} + \|\sum_{l=1}^{L} \left(\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\right)\|_{L^2(\Omega;V)} \\
&=: I + II.
\end{aligned}$$

We calculate the error bounds for the terms I and II separately.

Term I: By Jensen's and the Cauchy-Schwarz inequality, for every $l = 1, ..., L$, we get

$$I \leq \left([\|\mathbb{E}(u - G_l u)\|^2_{L^2(\Omega;V)}]\right)^{1/2} = \|u - u_l\|_{L^2(\Omega;V)} \leq C_I C_a \, h_l \|u\|_{L^2(\Omega;W)}.$$

In particular for $l = L$ we obtain the asserted bound for Term I.

Term II: by the triangle inequality, we must consider for each $l = 1, ..., L$ the term

$$\|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;V)}.$$

Each of these terms is estimated as follows:

$$\begin{aligned}\|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;V)} &= \|(\mathbb{E} - E_{M_l})[u_l - u_{l-1}]\|_{L^2(\Omega;V)} \\ &\leq M_l^{-1/2}\|u_l - u_{l-1}\|_{L^2(\Omega;V)} \\ &\leq M_l^{-1/2}\left(\|u - u_l\|_{L^2(\Omega;V)} + \|u - u_{l-1}\|_{L^2(\Omega;V)}\right) \\ &\leq C_a C_I\, M_l^{-1/2}(h_l + h_{l-1})\|u\|_{L^2(\Omega;W)} \\ &= 3\, C_a C_I\, h_l M_l^{-1/2}\|u\|_{L^2(\Omega;W)}.\end{aligned}$$

Here we used Lemma 4.1, Equation (4.3) and Equation (4.4). Summing these estimates from $l = 1, ..., L$ completes the proof. □

The preceding result gives an error bound for the MLMC–FE approximation, for *any* distribution $\{M_l\}_{l=1}^L$ of samples over the mesh levels. Like in the single–level Monte Carlo approximation one is interested in the optimal ratio of sample size versus grid size in every level, i.e. how M_l relates to h_l to achieve an overall convergence rate of $\mathrm{O}(h_L)$.

Theorem 4.5. *Under Assumptions 3.1, 3.5 and 4.2, the MLMC–FE approximation, given by Equation* (4.9)*, of the expectation of the solution of the stochastic elliptic boundary value problem (Equation* (3.1) – (3.3)*) in the polyhedral domain $D \subset \mathbb{R}^d$ with M_l samples on mesh level l given by*

$$M_l = l^{2+2\epsilon}2^{2(L-l)}h_0, \qquad l = 1, 2, ..., L,$$

where $\epsilon > 0$ is arbitrarily small, admits the error bound

$$\|\mathbb{E}[u] - E^L[u]\|_{L^2(\Omega;V)} \leq C h_L \left(\|f\|_{L^2(\Omega;L^2(D))} + \|g\|_{L^2(\Omega;H^{3/2}(\Gamma_N))}\right).$$

If, at each level l the Finite Element equations for each sample $\hat{u}_l^i$ in the estimator $E_{M_l}[u_l]$ are solved approximately with a full Multigrid method to accuracy $\mathrm{O}(h_l)$ in the energy norm, the total work Work(L) and memory for computing $E^L[u]$ approximately to accuracy $\mathrm{O}(h_L)$ is bounded by

$$\mathit{Work}(L) \leq C_\epsilon \begin{cases} N_L^2 & \text{for} \quad d = 1, \\ N_L(\log N_L)^{3+\epsilon} & \text{for} \quad d = 2, \\ N_L(\log N_L)^{2+\epsilon} & \text{for} \quad d = 3, \end{cases}$$

where the constant C depends on ϵ but is independent of L.

Proof. The convergence result in Lemma 4.4 suggests that we choose M_l such that the overall rate of convergence is $\mathrm{O}(h_L)$. With the choice

$$M_l = l^{2+2\epsilon}(h_l/h_L)^2 = \mathrm{O}(l^{2+2\epsilon}2^{2(L-l)}), \qquad l = 1, \ldots, L \tag{4.11}$$

for some $\epsilon > 0$, we obtain from Equation (4.10) the asserted error bound, since for $\epsilon > 0$ this implies

$$\begin{aligned}\sum_{l=1}^{L} h_l M_l^{-1/2} &\leq C \sum_{l=1}^{L} 2^{-l} h_0 (l)^{-(1+\epsilon)} 2^{(l-L)} h_0 \\ &\leq C 2^{-L} h_0 \sum_{l=1}^{L} (l)^{-(1+\epsilon)} \\ &\leq C h_L \sum_{l=1}^{L} (l)^{-(1+\epsilon)} \\ &= C(\epsilon) h_L.\end{aligned}$$

To estimate the work, we observe that the approximate solution given by the Finite Element equation solved by a full Multigrid method at mesh level l to accuracy h_l is of linear complexity in the number N_l of unknowns at mesh level l (see, e.g., [8, 9]). For M_l samples (possibly in parallel) this requires a total of $\mathrm{O}(M_l N_l)$ computational work and memory. This amounts to the following bound for the overall work for the MLMC–FE method at level L

$$\begin{aligned}\mathrm{Work}(L) &\lesssim \sum_{l=1}^{L} M_l N_l \\ &\leq \sum_{l=1}^{L} (l)^{2+2\epsilon} 2^{2(L-l)} 2^{dl} \\ &= 2^{dL} \sum_{l=1}^{L} (l)^{2+2\epsilon} 2^{2(L-l)} 2^{d(l-L)} \\ &= 2^{dL} \sum_{l=1}^{L} (l)^{2+2\epsilon} 2^{(d-2)(l-L)} \\ &\lesssim N_L \begin{cases} \sum_{l'=0}^{L-1} (L-l')^{2+2\epsilon} 2^{l'} & \text{for} \quad d=1, \\ \sum_{l'=0}^{L-1} (L-l')^{2+2\epsilon} & \text{for} \quad d=2, \\ \sum_{l'=0}^{L-1} (L-l')^{2+2\epsilon} 2^{-l'} & \text{for} \quad d=3. \end{cases}\end{aligned}$$

This implies the asserted work estimates, if we use in the case $d = 1$ for $0 < \epsilon < 1$ summation by parts three times. □

Remark 4.6. We remark that in the particular case $d = 1$, i.e. when the domain D coincides with an interval, with the standard "hat function" basis for $\mathcal{S}^{1,1}(D, \mathcal{T}_l)$ the stiffness matrix is tridiagonal and symmetric positive definite provided Assumption 3.1 is satisfied. Therefore, direct solvers are applicable with complexity $\mathrm{O}(N_l)$.

Remark 4.7. In the same particular case $d = 1$, the approximation with standard "hat functions" is already too accurate for the model problem. Since we equilibrate the errors of the MC method and the FE method, the MC error is dominating the overall error, leading to an increase of the samples which causes the quadratic complexity. In the cases of higher space dimensions the accuracy of the FE approximation, expressed in terms of the degrees of freedom, is lower. In this case the overall error is not dominated by the MC error.

In the subsequent chapter we further detail these results for the mean field to the approximation of higher moments of the solution of the elliptic model problem.

5. Multi–level Monte Carlo Finite Element approximation of higher moments

We now address two generalizations of the MLMC–FE approximation: the efficient computation of k-th moments of the stochastic solution $u \in L^2(\Omega; V)$, and the use of Finite Elements which are based on continuous, piecewise polynomials of degree $p \geq 1$. In particular the case $k = 2$, i.e. second moments, is of substantial interest in practice. Since, however, k-th moments (which are sometimes referred to as k-point correlation functions) are functions on the k-fold product domain $D^k = D \times \ldots \times D$, a naive MC estimation (with M samples) of the product of the solution vectors will entail complexity MN_L^k. The main result of the present section states that in order to recover log-linear complexity of k-th moments with $k \geq 2$, the MLMC–FE approximation must be combined with a wavelet compression of the Finite Element solutions for each sample. For $k = 1$, the results constitute a generalization of the preceding analysis to higher order elements. For $k = 2$ in two spatial dimensions (i.e. when $d = 2$) we obtain in particular a log–linear complexity scheme for the computation of a Galerkin approximation to the so–called "$4d$-VAR" of the stochastic solution.

Therefore, we first establish the regularity of the k-th moment of the solution of the elliptic problem given certain smoothness and regularity conditions on the data and the coefficient and we introduce wavelet bases for the hierarchical meshes. Under these assumptions we derive full and sparse tensor error bounds for the Finite Element approximation. Those bounds are essential for the error of the sparse tensor MLMC–FE approximation.

5.1. Existence and Regularity of k-th Moments.

We are interested in statistical moments of the stochastic solution u: for any $k \in \mathbb{N}$ we denote the k-fold tensor products of a separable Hilbert space X as

$$X^{(k)} = \underbrace{X \otimes \cdots \otimes X}_{k\text{-times}},$$

equipped with the natural norm $\| \cdot \|_{X^{(k)}}$. This norm has the property that for every $u_1, \ldots, u_k \in X$ there holds the isometry

$$\|u_1 \otimes \cdots \otimes u_k\|_{X^{(k)}} = \|u_1\|_X \cdots \|u_k\|_X \ .$$

For $u \in L^k(\Omega; X)$ we now consider the random field $(u)^{(k)}$ defined by $u(\omega) \otimes \cdots \otimes u(\omega)$. Then $(u)^{(k)} = u \otimes \cdots \otimes u \in L^1(\Omega, X^{(k)})$ and we have the isometry

$$\|(u)^{(k)}\|_{L^1(\Omega;X^{(k)})} = \int_\Omega \|u(\omega) \otimes \cdots \otimes u(\omega)\|_{X^{(k)}} dP(\omega) = \int_\Omega \|u(\omega)\|_X^k dP(\omega) = \|u\|_{L^k(\Omega;X)}^k. \tag{5.1}$$

Therefore, we define the moment $\mathcal{M}^k u$ as the expectation of $(u)^{(k)} = \underbrace{u \otimes \cdots \otimes u}_{k-times}$:

Definition 5.1. For $u \in L^k(\Omega; V)$, for some integer $k \geq 1$, the k-th moment (or k-point correlation function) of $u(\omega)$ is defined by

$$\mathcal{M}^k u = \mathbb{E}[(u)^{(k)}] = \mathbb{E}[\underbrace{u \otimes \cdots \otimes u}_{k-times}] = \int_{\omega \in \Omega} \underbrace{u(\omega) \otimes \cdots \otimes u(\omega)}_{k-times} \, dP(\omega) \in V^{(k)}. \tag{5.2}$$

As above, the numerical analysis of the higher order MLMC–FE method requires a regularity theory for solutions of Equation (3.1) - (3.3). To this end we introduce a smoothness scale $(Y_s)_{s\geq 0}$ for the data f, g with $Y_0 = H^{-1}(D) \times H^{-1/2}(\Gamma_N)$ and with $Y_s \subset Y_t$ for $s > t$.

We assume that we have a corresponding scale $(X_s)_{s\geq 0}$ of "smoothness spaces" for the solutions with $X_0 = V = H^1_{\Gamma_D}(D)$ and with $X_s \subset X_t$ for $s > t$, such that $(A(\omega))^{-1}\colon Y_s \to X_s$, defined in Equation (3.7), is continuous for all coefficient realization $\mathbb{P}$-a.s..

For example, for our model problem, Equation (3.1) - (3.3) with smooth random coefficients $a(\omega, x)$ in a domain D with smooth boundary ∂D and with $\Gamma_N = \emptyset$, we may choose $Y_s = V^* \cap H^{-1+s}(D) \times H^{-1/2+s}(\Gamma_N)$ and $X_s = V \cap H^{1+s}(D)$ for any $s > 0$. We remark that in non–smooth domains such as polyhedra in $\mathbb{R}^3$ the spaces X_s are weighted spaces which contain functions which are singular at corners and edges (see, e.g. [20]). We can now state our assumptions on the data of the model problem, given by Equation (3.1) - (3.3):

Assumption 5.2. For some $r^* \geq 2$ and some $s^* > 0$, the data (f, g) in Problem (3.1) - (3.3) belong to $L^{r^*}(\Omega; Y_{s^*})$ and the mapping $\Omega \ni \omega \to a(\omega, \cdot)$ is such that the operator $A(\omega)$ is boundedly invertible from Y_s to $X_s \subset V$ $\mathbb{P}$-a.s. for all $0 < s \leq s^*$ for some $s^* > 0$. Moreover, the random inputs a, f and g are independent.

We remark that Assumption 5.2 is satisfied if $a(\cdot, \omega) \in W^{s,\infty}(D)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$ and every $0 \leq s \leq s^*$.

Theorem 5.3. *If Assumptions 3.1 and 5.2 hold, then for every $2 \leq k \leq r^*$, for all $1 \leq r \leq r^*/k$, and every $0 \leq s < s^*$ holds the apriori estimate*

$$\|(u)^{(k)}\|_{L^r(\Omega; X_s^{(k)})} \leq C\|(f,g)^{(k)}\|_{L^r(\Omega; Y_s^{(k)})} \leq C\|(f,g)\|^k_{L^{rk}(\Omega; Y_s)}. \tag{5.3}$$

Proof. Under Assumption 5.2, the operator $A(\omega)^{(k)}$ is boundedly invertible from Y_s to X_s for each coefficient realization $\mathbb{P}$-a.s.. The stochastic solution satisfies, for $\omega \in \Omega$ $\mathbb{P}$-a.s., the apriori estimate

$$\|u(\omega, \cdot)\|_{X_s} \leq C(s, \omega)\|(f, g)(\omega, \cdot)\|_{Y_s}, \qquad 0 \leq s \leq s^*,$$

with a constant $C(\omega)$ bounded independently of ω. Raising both sides of the bounds to the r-th power and integrating the resulting inequality over $\omega \in \Omega$ with respect to the probability measure $\mathbb{P}(d\omega)$, we obtain the first inequality. The second inequality follows from the isometry given in Equation (5.1). □

Note in particular that in the case $s = 1$, we have $W = X_1$, $Y_1 = H^{1/2}(\Gamma_N) \times L^2(D)$ and for $k = 2$ Assumption 3.5 and Equation (3.12) imply the a-priori estimates

$$\begin{aligned}\|\mathcal{M}^2 u\|_{W^{(2)}} = \|\mathbb{E}[(u)^{(2)}]\|_{W^{(2)}} &\leq \|u\|^2_{L^4(\Omega;W)} \\ &\leq C(a)\left(\|f\|^2_{L^4(\Omega;L^2(D))} + \|g\|^2_{L^4(\Omega;H^{1/2}(\Gamma_N))}\right),\end{aligned}$$

and

$$\|(u)^{(2)}\|_{L^2(\Omega;W^{(2)})} \leq C(a)\left(\|f\|^2_{L^4(\Omega;L^2(D))} + \|g\|^2_{L^4(\Omega;H^{1/2}(\Gamma_N))}\right).$$

5.2. **Finite elements with uniform mesh refinement.** We will now generalize the subspaces V_l to simplicial Finite Elements of order $p \geq 1$.

Let us first consider the case of a bounded polyhedron $D \subset \mathbb{R}^d$. Let $\{\mathcal{T}_l\}_{l=0}^\infty$ be the sequence of partitions obtained by uniform mesh refinement: we can bisect the edges of $\mathcal{T}_l$ and obtain

a new partition into simplices which belong to finitely many congruency classes. Then we set $V_l = S^{p,1}(D, \mathcal{T}_l)$ and $h_l = \max\{\mathrm{diam}(K) : K \in \mathcal{T}_l\}$. We obtain $N_l = \dim V_l = O(h_l^{-d})$.

With V as before, and $X_s = V \cap H^{1+s}(D)$ the standard Finite Element approximation results give that the following bound holds for $s \in [0, p]$.

$$\inf_{v \in V_l} \|u - v\|_V \leq C\, N_l^{-s/d} \|u\|_{X_s}\,. \tag{5.4}$$

For a d–dimensional domain $D \subset \mathbb{R}^d$ with a smooth boundary we first divide D into patches D_J which can be mapped to a simplex $\mathcal{S}$ by smooth, bijective mappings $\Phi_J \colon D_J \to \mathcal{S}$ (which must be C^0 compatible where two pieces D_J and $D_{J'}$ touch). Then we define on D Finite Elements functions which on D_J are of the form $\gamma \circ \Phi_J$ where γ is a polynomial.

For a d–dimensional smooth surface $D \subset \mathbb{R}^{d+1}$ we similarly divide D into patches which can be mapped to simplices in $\mathbb{R}^d$, and again define Finite Elements using these mappings.

5.3. **Wavelet basis for V_l.** We introduce a hierarchical basis for the nested spaces $V_0 \subset \cdots \subset V_L$: we start with a basis $\{(\psi_l)_j\}_{j=1,\ldots,N_0}$ for the space V_0. We write the finer spaces V_l with $l > 0$ as a direct sum $V_l = V_{l-1} \oplus \mathcal{W}_l$ with a suitable space $\mathcal{W}_l$. *We assume available explicitly basis functions* $\{(\psi_l)_j\}_{j=1,\ldots,\bar{N}_l}$. Therefore we have that $V_L = V_0 \oplus \mathcal{W}_1 \oplus \cdots \oplus \mathcal{W}_L$, and $\{\,(\psi_l)_j \mid l = 0, \ldots, L;\ j = 1, \ldots, \bar{N}_l\,\}$ is a hierarchical basis for V_L where $\bar{N}_0 := N_0$:

(W1) $V_l = \mathrm{span}\{(\psi_l)_j \,|1 \leq j \leq \bar{N}_k,\, 0 \leq k \leq l\}$,

with $N_l := \dim V_l$ and, $\bar{N}_l := N_l - N_{l-1}$ for $l \geq 0$.

Property (W1) is in principle sufficient for the formulation and implementation of the sparse MC–FE method and the deterministic sparse Finite Element method. In order to obtain an algorithm with log-linear complexity we will need that the hierarchical basis satisfies the additional properties (W2)–(W6) of a *wavelet basis*. This will allow us to perform matrix compression, and to obtain optimal preconditioning for the iterative linear system solver.

(W2) **Small support:** $\mathrm{diam}\ \mathrm{supp}((\psi_l)_j) = O(2^{-l})$.

(W3) **Biorthogonal Basis:** there exists a biorthogonal basis $\widetilde{\Psi} = \{(\tilde{\psi}_l)_j \ :\ 1 \leq j \leq \bar{N}_k,\, 0 \leq k \leq l = 1, 2, ...\}$ such that

$$\langle (\psi_l)_j, (\tilde{\psi}_{l'})_{j'} \rangle = \delta_{ll'}\delta_{jj'}.$$

(W4) **Energy norm stability:** there is a constant $C_B > 0$ independent of level L, such that for all $v_L = \sum_{l=0}^{L} \sum_{j=1}^{\bar{N}_l} (v_l)_j\, (\psi_l)_j(x) \in V_L$ holds $(v_l)_j = \langle v, (\psi_l)_j \rangle$ and

$$\frac{1}{C_B} \sum_{l=0}^{L} \sum_{j=1}^{\bar{N}_l} |(v_l)_j|^2 \leq \|v_L\|_V^2 \leq C_B \sum_{l=0}^{L} \sum_{j=1}^{\bar{N}_l} |(v_l)_j|^2.$$

(W5) Wavelets $(\psi_l)_j$ with $l \geq l_0$ have **vanishing moments** up to order $p_0 \geq p - 2$

$$\int (\psi_l)_j(x)\, x^\alpha\, dx = 0, \qquad 0 \leq |\alpha| \leq p_0.$$

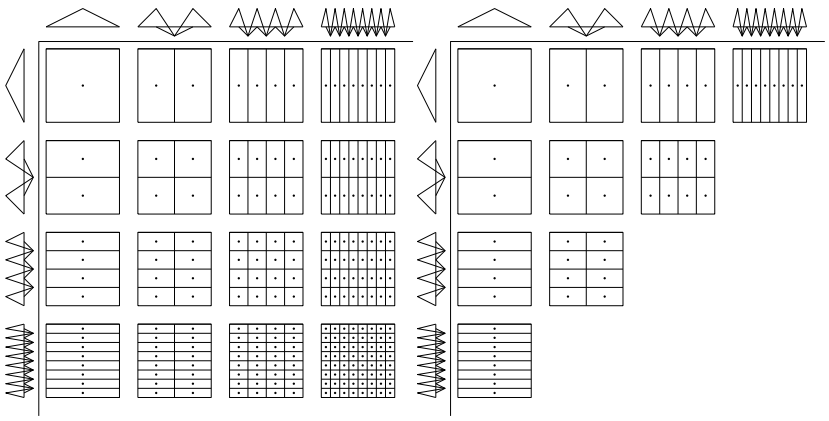
Except possibly for wavelets where the closure of the support intersects the boundary ∂D or the boundaries of the coarsest mesh.

(W6) **Decay of coefficients for "smooth" functions in X_s:**
there exists $C > 0$ independent of L such that for every $v \in X_s$ and every L holds

$$\sum_{l=0}^{L} \sum_{j=1}^{\bar{N}_l} |(v_l)_j|^2 \; 2^{2ls} \leq C L^{\nu} \|v\|_{X_s}^2,$$

$$\nu = \begin{cases} 0 & \text{for } 0 \leq s < p, \\ 1 & \text{for } s = p \, . \end{cases}$$

Many concrete, piecewise polynomial wavelet systems satisfying (W1)-(W6) are available, also in polygonal and polyhydral domains D. Any function $u \in V$ admits a wavelet expansion



(a) Full grid (b) Sparse grid

FIGURE 1. Biorthogonal, piecewise linear spline wavelets

$\sum_{l=0}^{\infty} \sum_{j=1}^{\bar{N}_l} (u_l)_j \, (\psi_l)_j$. We define the projection $P_L : V \to V_L$ by the truncating this wavelet expansion of u, i.e.,

$$P_L u := \sum_{\ell=0}^{L} \sum_{j=1}^{\bar{N}_l} (u_l)_j \, (\psi_l)_j, \qquad (u_l)_j = \langle u, (\tilde{\psi}_l)_j \rangle.$$

With the stability (W3) and the approximation property in Equation (5.4) we obtain that the wavelet projection P_L is quasioptimal: with $N_L = \dim V_L$, we have for $0 \leq s \leq s^*$ and $u \in X_s$ the asymptotic error bound

$$\|u - P_L u\|_V \leq C \, N_L^{-s/d} \|u\|_{X_s}.$$

5.4. **Full and sparse tensor product spaces.** To compute MLMC–FE approximations for $\mathcal{M}^k u \in V \otimes \cdots \otimes V = V^{(k)}$ (cf. Equation (5.2)), we project $\mathcal{M}^k u$ onto a finite dimensional subspace of $V^{(k)}$. The choice of the k-fold tensor product space $V_L^{(k)} = V_L \otimes \cdots \otimes V_L$ leads to the *full tensor MC–FE estimates* for $\mathcal{M}^k u$ in Equation (5.2):

$$E_M[(u_l)^{(k)}] = \frac{1}{M} \sum_{i=1}^{M} (\hat{u}_l^i)^{(k)} \, . \tag{5.5}$$

Here, the $\hat{u}_l^i \in V_l$ are the previously discussed Galerkin approximations for $i = 1, ..., M_l$ i.i.d. samples of the stochastic coefficients.

The space $V_L^{(k)}$ has dimension N_L^k and even forming one tensor product $(\hat{u}_L^j)^{(k)}$ of a Finite Element sample in the Monte Carlo estimate of Equation (5.5) would destroy the linear complexity of the MC–FE estimator for moments of order $k > 1$.

A reduction in cost at, as we shall show, essentially no loss in accuracy, is possible by using so-called sparse tensor products of the Finite Element spaces V_l which we define next.

We now define the k-fold *sparse tensor product space* $\hat{V}_L^{(k)}$ by

$$\hat{V}_L^{(k)} = \sum_{\substack{\vec{\ell}\in\mathbb{N}_0^k \\ |\vec{\ell}|\le L}} V_{\ell_1} \otimes \cdots \otimes V_{\ell_k},$$

where we denote by $\vec{\ell}$ the vector $(\ell_1, \ldots, \ell_k) \in \mathbb{N}_0^k$ and its length by $|\vec{\ell}| = \ell_1 + \cdots + \ell_k$. We can write V as a direct sum by using the complement spaces $\mathcal{W}_l$:

$$\hat{V}_L^{(k)} = \sum_{\substack{\vec{\ell}\in\mathbb{N}_0^k \\ |\vec{\ell}|\le L}} \mathcal{W}_{\ell_1} \otimes \cdots \otimes \mathcal{W}_{\ell_k}.$$

We define a projection operator $\widehat{P}_L^{(k)} : V^{(k)} \to \hat{V}_L^{(k)}$, for $x = (x_1, ..., x_k) \in D^{(k)}$ by truncating the wavelet expansion:

$$(\widehat{P}_L^{(k)} v)(x) := \sum_{\substack{0\le \ell_1+\cdots+\ell_k\le L \\ 1\le j_\nu \le \tilde{N}_{\ell_\nu}, \nu=1,\ldots,k}} (v_{\ell_1\ldots\ell_k})_{j_1\ldots j_k}\, (\psi_{\ell_1})_{j_1}(x_1)\ldots(\psi_{\ell_k})_{j_k}(x_k)\,. \tag{5.6}$$

Here, the coefficients are given by

$$(v_{\ell_1\ldots\ell_k})_{j_1\ldots j_k} = {}_{V^{(k)}}\langle v, (\tilde{\psi}_{\ell_1})_{j_1} \otimes ... \otimes (\tilde{\psi}_{\ell_k})_{j_k}\rangle_{(V^{(k)})'}.$$

With the projections $\Pi_l := P_l - P_{l-1}$, $l = 0, 1, \ldots$ and $P_{-1} := 0$ we can express $\widehat{P}_L^{(k)}$ as

$$\widehat{P}_L^{(k)} = \sum_{0\le\ell_1+\cdots+\ell_k\le L} \Pi_{\ell_1} \otimes \cdots \otimes \Pi_{\ell_k}.$$

The approximation property of sparse tensor products of the finite element spaces, i.e. of $\hat{V}_L^{(k)}$, was established for example in [29, 30, Proposition 4.2], [19], [33].

Proposition 5.4. *For $u \in X_s^{(k)}$ with $0 \le s \le s^*$ we have*

$$\inf_{v\in\widehat{V}_L^{(k)}} \|u - v\|_{V^{(k)}} \le C(k) \begin{cases} N_L^{-s/d} \|u\|_{X_s^{(k)}} & \text{if } 0 \le s < p, \\ N_L^{-s/d} L^{(k-1)/2} \|u\|_{X_s^{(k)}} & \text{if } s = p\,. \end{cases}$$

The stability property (W3) implies the following result (see, e.g., [33]):

Lemma 5.5. *(Properties of $\widehat{P}_L^{(k)}$)*
Assume (W1)–(W6) and that the component spaces V_ℓ of $\widehat{V}_L^{(k)}$ have the approximation property given in Equation (5.4). Then for $u \in V^{(k)}$ the truncated tensorized wavelet expansion is stable, i.e. for every $k \in \mathbb{N}$ exists $C(k) > 0$ such that for every $u \in V^{(k)}$ and every L holds

$$\|\widehat{P}_L^{(k)} u\|_{V^{(k)}} \le C(k)\, \|u\|_{V^{(k)}}\,. \tag{5.7}$$

For $u \in X_s^{(k)}$ and $0 \le s \le s^*$ we have quasioptimal convergence of $\widehat{P}_L^{(k)} u$:

$$\|u - \widehat{P}_L^{(k)} u\|_{V^{(k)}} \le C(k) N_L^{-s/d} (\log N_L)^{(k-1)/2} \|u\|_{X_s^{(k)}}. \tag{5.8}$$

This results provide us with the necessary tools to estimate the rate of convergence for the sparse tensor MLMC–FE method for $\mathcal{M}^k u$.

5.5. **Sparse tensor Multi–level Monte Carlo approximation of higher moments.** We aim at estimating

$$\mathcal{M}^k u = \mathbb{E}[(u)^{(k)}] = \mathbb{E}[u \otimes ... \otimes u].$$

To do so, we have at our disposal coefficient samples $a(\omega_i, x)$ and the Galerkin Finite Element approximations $\hat{u}_l^i(x)$ defined in Equation (4.2). We therefore define MLMC–FE estimates as statistical averages of the compressed tensor products of the Galerkin Finite Element approximations as follows:

$$\widehat{E}^L[(u_L)^{(k)}] := \sum_{l=1}^{L} E_{M_l}\left[\widehat{P}_l^{(k)}(u_l)^{(k)} - \widehat{P}_{l-1}^{(k)}(u_{l-1})^{(k)}\right], \tag{5.9}$$

where we once again used the convention that $u_0 := 0$ and that $\widehat{P}_0^{(k)} := 0$. We remark that due to $\widehat{P}_l^{(1)} = P_l$, the estimator in Equation (5.9) will coincide with our standard MLMC–FE estimator in the case $k = 1$, i.e. for estimating the expectation of u. We can now state our MLMC–FE error bound for moments $\mathcal{M}^k u$ of order $k \geq 2$.

Theorem 5.6. *Assume that $(f, g) \in L^{2k}(\Omega, Y_s)$ and that the operators $A(\omega) \in \mathcal{L}(X_s, Y_s)$, as defined in (3.7), are boundedly invertible for $0 \leq s \leq s^*$ $\mathbb{P}$-a.s., and that the Finite Element spaces $\mathcal{S}^{p,n}(D, \tau_l)$, for $l = 1, \ldots, L$, defined in Equation (4.1), satisfy the approximation property in Equation (5.4).*

Then there holds for $0 \leq s \leq \min(s^*, p)$ and for any numbers M_l of coefficient samples in the Galerkin Finite Element method on mesh $\mathcal{T}_l$ the bound

$$\left\|\mathcal{M}^k u - \widehat{E}^L[(u_L)^{(k)}]\right\|_{L^2(\Omega;V^{(k)})} \lesssim \left(\sum_{l=1}^{L} M_l^{-1/2} h_l^s |\log h_l|^{(k-1)/2}\right) \|(f,g)\|_{L^{2k}(\Omega;Y_s)}^k .$$

Here, $M_L = 1$ and the constant in $\lesssim$ depends on s, p, k but is independent of the number L of mesh refinements and of the distribution of the numbers M_l of samples at mesh levels l.

Proof. We write

$$\begin{aligned}
\|\mathcal{M}^k u - \widehat{E}^L[(u_L)^{(k)}]\|_{L^2(\Omega;V^{(k)})} &= \|\mathbb{E}[(u)^{(k)}] - \widehat{E}^L[(u_L)^{(k)}]\|_{L^2(\Omega;V^{(k)})} \\
&\leq \|\mathbb{E}[(u)^{(k)}] - \mathbb{E}[\widehat{P}_L^{(k)}(u_L)^{(k)}]\|_{L^2(\Omega;V^{(k)})} \\
&+ \|\mathbb{E}[\widehat{P}_L^{(k)}(u_L)^{(k)}] - \sum_{l=1}^{L} E_{M_l}\left[\widehat{P}_l^{(k)}(u_l)^{(k)} - \widehat{P}_{l-1}^{(k)}(u_{l-1})^{(k)}\right] \|_{L^2(\Omega;V^{(k)})} \\
&=: I + II.
\end{aligned}$$

We estimate the terms separately. For term I, we obtain with Jensen's inequality and Equation (5.7), for any $0 \le s \le \min(p, s^*)$, the error bound

$$\begin{aligned}
I &= \left\| \mathbb{E}[(u)^{(k)}] - \mathbb{E}[\widehat{P}_L^{(k)}(u_L)^{(k)}] \right\|_{L^2(\Omega;V^{(k)})} \\
&\le \left\| \mathbb{E}[(u)^{(k)} - \widehat{P}_L^{(k)}(u_L)^{(k)}] \right\|_{V^{(k)}} \\
&\le \left\| (u)^{(k)} - \widehat{P}_L^{(k)}(u)^{(k)} \right\|_{L^1(\Omega;V^{(k)})} + \left\| \widehat{P}_L^{(k)} \left((u)^{(k)} - (u_L)^{(k)} \right) \right\|_{L^1(\Omega;V^{(k)})} \\
&\lesssim \left\| \left(I - \widehat{P}_L^{(k)} \right) (u)^{(k)} \right\|_{L^1(\Omega;V^{(k)})} + \|(u)^{(k)} - (u_L)^{(k)}\|_{L^1(\Omega;V^{(k)})} \\
&=: I_a + I_b.
\end{aligned}$$

Term I_a is a consistency error which is bounded with Equation (5.8). To estimate term I_b, we denote the k dependence of this term by $I_b(k)$, then we write

$$\begin{aligned}
I_b(k) &= \|(u)^{(k)} - (u_L)^{(k)}\|_{L^1(\Omega;V^{(k)})} \\
&\le \|(u - u_L) \otimes (u)^{(k)}\|_{L^1(\Omega;V^{(k)})} + \|u_L \otimes ((u)^{(k-1)} - (u_L)^{(k-1)})\|_{L^1(\Omega;V^{(k)})} \\
&\le \|u - u_L\|_{L^2(\Omega;V)} \|(u)^{(k-1)}\|_{L^2(\Omega;V^{(k-1)})} + \|u_L\|_{L^\infty(\Omega;V)} \|(u)^{(k-1)} - (u_L)^{(k-1)}\|_{L^1(\Omega;V^{(k-1)})} \\
&= \|u - u_L\|_{L^2(\Omega;V)} \|u\|^{k-1}_{L^{2k-2}(\Omega;V)} + \|u_L\|_{L^\infty(\Omega;V)} I_b(k-1) \\
&\le C(s) N_L^{-s/d} \|f\|_{L^2(\Omega;Y_s)} \|f\|^{k-1}_{L^{2k-2}(\Omega;V^*)} + C(a) I_b(k-1) \, .
\end{aligned}$$

Induction with respect to k leads to the overall bound for I

$$I \le I_a + I_b(k) \le C(a, f, k) N_L^{-s/d} (\log N_L)^{(k-1)/2} \, .$$

We estimate term II as follows.

$$\begin{aligned}
II &= \|\mathbb{E}\left[\widehat{P}_L^{(k)}(u_L)^{(k)} \right] - \sum_{l=1}^{L} E_{M_l} \left[\widehat{P}_l^{(k)}(u_l)^{(k)} - \widehat{P}_{l-1}^{(k)}(u_{l-1})^{(k)} \right] \|_{L^2(\Omega;V^{(k)})} \\
&= \| \sum_{l=0}^{L} \left\{ (\mathbb{E} - E_{M_l}) \left[\widehat{P}_l^{(k)}(u_l)^{(k)} - \widehat{P}_{l-1}^{(k)}(u_{l-1})^{(k)} \right] \right\} \|_{L^2(\Omega;V^{(k)})} \\
&\le \sum_{l=1}^{L} M_l^{-1/2} \left\| \widehat{P}_l^{(k)}(u_l)^{(k)} - \widehat{P}_{l-1}^{(k)}(u_{l-1})^{(k)} \right\|_{L^2(\Omega;V^{(k)})} \\
&\le \sum_{l=1}^{L} M_l^{-1/2} \left\{ \left\| (u)^{(k)} - \widehat{P}_l^{(k)}(u_l)^{(k)} \right\|_{L^2(\Omega;V^{(k)})} + \left\| (u)^{(k)} - \widehat{P}_{l-1}^{(k)}(u_{l-1})^{(k)} \right\|_{L^2(\Omega;V^{(k)})} \right\} \\
&=: \sum_{l=1}^{L} M_l^{-1/2} (III(l) + III(l-1)).
\end{aligned}$$

Each of the terms in the sum is bounded as

$$\begin{aligned} III(l) :=& \left\|(u)^{(k)} - \widehat{P}_l^{(k)}(u)_l^{(k)}\right\|_{L^2(\Omega;V^{(k)})} \\ \leq& \left\|(u)^{(k)} - \widehat{P}_l^{(k)}(u)^{(k)}\right\|_{L^2(\Omega;V^{(k)})} + \left\|\widehat{P}_l^{(k)}((u)^{(k)} - (u_l)^{(k)})\right\|_{L^2(\Omega;V^{(k)})} \\ \lesssim& \left\|(u)^{(k)} - \widehat{P}_l^{(k)}(u)^{(k)}\right\|_{L^2(\Omega;V^{(k)})} + \left\|(u)^{(k)} - (u_l)^{(k)}\right\|_{L^2(\Omega;V^{(k)})} \\ =:& \; III_a + III_b. \end{aligned}$$

We estimate III_a with Equation (5.8). Term III_b is bounded as term $I_b(k)$, with l in place of L. Combining the bounds for III_a and III_b, we obtain with $h_l \simeq N_l^{-1/d}$ for every $k \geq 1$ and every $l \geq 0$ the error estimate

$$\begin{aligned} III &\leq C(k) N_l^{-\min(s,p)/d} (\log N_l)^{(k-1)/2} \|(u)^{(k)}\|_{L^2(\Omega;X_s^{(k)})} + h_l^{k\min(s,p)} \|u\|^k_{L^{2k}(\Omega;X_s)} \\ &\leq C(k,s) N_l^{-\min(s,p)/d} (\log N_l)^{(k-1)/2} \|u\|^k_{L^{2k}(\Omega;X_s)} \\ &= C(k,s) h_l^{\min(s,p)} |\log h_l|^{(k-1)/2} \|u\|^k_{L^{2k}(\Omega;X_s)}. \end{aligned}$$

Using this estimate for each $l = 0, ..., L$ to bound II, and referring to Equation (5.3) with $p = 2$, we obtain with the estimate for I the asserted error bound. □

We observe that in the case $k = 1$ and $p = 1$, with the choices $X_0 = V$ and $X_1 = W$, we recover the previous results. We now optimize the selection of MC samples $\{M_l\}_{l=0}^L$ and state the resulting overall convergence rate of the MLMC–FE method for moments, $\mathcal{M}^k u$, for any order $k \geq 1$.

Theorem 5.7. *Assume that $(f, g) \in L^{2k}(\Omega, Y_s)$ and that the operators $A(\omega) \in \mathcal{L}(X_s, Y_s)$ are boundedly invertible, for $0 \leq s \leq \min(s^*, p)$ for $\omega \in \Omega$, $\mathbb{P}$-a.s..*

Given any $k \in \mathbb{N}$, we choose the number of MC samples in the MC–FE method at level l used in the computation of the MLMC–FE estimators in Equation (5.9) *as*

$$M_l = \mathcal{O}(2^{2s(L-l)}(l/L)^{k-1}), \qquad l = 1, ..., L. \tag{5.10}$$

Then there holds for $0 \leq s \leq \min(s^*, p)$ the error bound

$$\left\|\mathcal{M}^k u - \widehat{E}^L[(u_L)^{(k)}]\right\|_{L^2(\Omega;V^{(k)})} \lesssim h_L^s |\log h_L|^{(k+1)/2} \|(f,g)\|^k_{L^{2k}(\Omega;Y_s)} .$$

and the total work $\widehat{W}(L)$ for computing the MLMC–FE estimator in Equation (5.9) *is bounded by*

$$\widehat{W}(L) \leq C(k) \begin{cases} N_L (\log N_L)^{k-1} & 2s \leq d, \\ N_L^{2s/d} (\log N_L)^{k-1} & 2s > d. \end{cases} \tag{5.11}$$

Proof. In Theorem 5.6, we choose the numbers M_l of samples at mesh level l such that the error contributions from the levels to the error bound are equilibrated. This gives, for $l = 1, 2, ..., L$,

$$M_l^{-1/2} = 2^{-s(L-l)} (L/l)^{(k-1)/2},$$

which implies Equation (5.10). Inserting this into the error bound of Theorem 5.6, we obtain Equation (5.7).

To estimate the complexity, we observe that the work to solve the Galerkin Finite Element equations to the required accuracy $\mathcal{O}(h_l^s)$ in the $\|\cdot\|_V$-norm can be achieved in linear complexity, i.e. in $\mathcal{O}(N_l) = \mathcal{O}(2^{ld})$ work and memory; this complexity estimate can be attained

in two ways: either by using the standard (one-scale) Finite Element basis and full Multigrid (see e.g. [8, 9]), or by using a diagonally preconditioned Richardson iteration in wavelet bases. In the latter case, the wavelet Galerkin Finite Element solution vector is directly obtained in the wavelet representation, so that the formation of the sparse tensor approximation in Equation (5.6) of the k-th moment of the Galerkin Finite Element approximation is obtained at cost $\mathcal{O}(N_l(\log N_l)^{k-1})$ work and memory.

In case the Finite Element solution is computed in the standard (one-scale) basis, the solution vector for each sample must first be transformed into the wavelet basis. This is achieved as usual in $\mathcal{O}(N_l)$ work and memory by the pyramid scheme (see e.g. [10]). Then the formation of the k-fold sparse tensor product of $\widehat{P}_l u_l^{(k)}$ proceeds again according to Equation (5.6).

For the work estimate, we therefore obtain

$$\begin{aligned}
\widehat{W}(L) &= \sum_{l=1}^{L} M_l N_l (\log N_l)^{k-1} \\
&\lesssim \sum_{l=1}^{L} 2^{2s(L-l)} (l/L)^{k-1} 2^{dl} l^{k-1} \\
&= 2^{dL} \sum_{l=1}^{L} L^{-(k-1)} l^{2(k-1)} 2^{2sL+l(d-2s)-dL} \\
&= N_L L^{-(k-1)} \sum_{l=1}^{L} l^{2(k-1)} 2^{(l-L)(d-2s)} \\
&= N_L L^{-(k-1)} \sum_{l'=0}^{L-1} (L-l')^{2(k-1)} 2^{l'(2s-d)} \\
&\lesssim N_L L^{-(k-1)} \begin{cases} L^{2(k-1)} & 2s \le d, \\ \sum_{l'=0}^{L-1} (L-l')^{2(k-1)} 2^{l'(2s-d)} & 2s > d. \end{cases}
\end{aligned}$$

□

Remark 5.8. We remark that the case discussed in Theorem 4.5 corresponds to the case $k = 1$, $s = 1$ and $p = 1$ in Theorems 5.6 and 5.7. Upon comparing both error bounds and the corresponding work estimates, we observe slight differences in the logarithmic terms; this is due to the slightly more conservative choice of the numbers M_l of samples in Equation (4.11) which we made in order to avoid the appearance of $\log h_L$ terms in the error bound of Equation (4.10). For moments of order $k \ge 2$, however, such terms appear in any case due to the sparse tensor approximation error bound in Proposition 5.4 which is sharp, so that the slightly more straightforward selection in Equation (5.10) is sufficient to achieve the expected convergence rates.

Remark 5.9. The complexity bound in Equation (5.11) in Theorem 5.7 indicates loss of log–linear complexity as soon as $2s > d$. In this case, the smoothness s of the solution mapping $A(\omega)^{-1}$ allows for higher convergence rates of the Galerkin Finite Element approximation in D which, when combined with a linear complexity solver such as Multigrid or a diagonally preconditioned wavelet solver, will imply that the efficiency of the MLMC-FE method (i.e. accuracy versus work) is dominated by the "weaker" of the two methods. In the case $s > d/2$, this is the MC method. We conclude from Theorem 5.7 that, therefore, the use of a MLMC-FE method is only advisable in connection with low order Finite Element methods: in spatial

dimension $d = 2$, log-linear complexity will be retained with linear simplicial Finite Element methods where $p = 1$. In spatial dimension $d = 3$, linear complexity can be retained up to $s = 3/2$; to access this range of convergence orders, it will suffice to use simplicial Finite Element methods of polynomial degree $p = 2$. With these methods, convergence for the expectation and k–th moments can be achieved in overall complexity of $\mathcal{O}(N_L^{4/3}(\log N_L)^{k-1})$ for work and $\mathcal{O}(N_L(\log N_L)^{k-1})$ for memory.

6. Implementation and Examples

In this section we discuss the approximation of the stochastic coefficient a for $\omega \in \Omega$. Followed by numerical examples in $\mathbb{R}$ and $\mathbb{R}^2$.

6.1. **Coefficient Representations.** For numerical simulations, the random field $a(\omega, x)$ in Equation (3.1) must be represented parametrically. Here, we discuss the implementation and the complexity of two choices: a Karhúnen-Loève expansion and a (multi)wavelet expansion of $a(\omega, x)$.

6.1.1. *Karhúnen-Loève -expansion.* Random diffusion coefficients $a(\omega, x) \in L^2(\Omega; L^2(D))$, admit a Karhúnen-Loève expansion in terms of the eigenpairs $(\lambda_k, \varphi_k)_{k=1}^{\infty}$ of the covariance operator which is the compact and self-adjoint integral operator with kernel q_a given by

$$q_a := \mathbb{E}[(a - \mathbb{E}[a]) \otimes (a - \mathbb{E}[a])] \tag{6.1}$$

or defined pointwise formally by

$$q_a(x, x') := \mathbb{E}[(a(\cdot, x) - \mathbb{E}[a](x))(a(\cdot, x') - \mathbb{E}[a](x'))] \,, \quad x, x' \in D. \tag{6.2}$$

We assume that the eigenfunctions φ_k are normalized in $L^2(D)$ and the λ_k are enumerated in decreasing magnitude, then the random diffusion coefficient admits the Karhúnen-Loève expansion

$$a(\omega, x) = \mathbb{E}[a](x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} Y_i(\omega) \varphi_i(x), \tag{6.3}$$

where the random coefficients $Y_i(\omega)$, for $i = 1, 2, \ldots$, are defined by

$$Y_i(\omega) = \begin{cases} \dfrac{1}{\sqrt{\lambda_i}} \displaystyle\int_D (a(\omega, x) - \mathbb{E}[a](x)) \varphi_i(x) dx \,, & \text{if} \quad \lambda_i > 0, \\ 0 \qquad \text{otherwise.} \end{cases} \tag{6.4}$$

The Karhúnen-Loève series in Equation (6.3) converges in $L^2(\Omega; L^2(D))$ (see [31]).

Estimation of the probability density function for the stochastic coefficients Y_k in the Karhúnen-Loève expansion, Equation (6.3), from an ensemble $\{\hat{a}^i \in L^\infty(D) : i = 1, ..., I\}$, can be performed via Equation (6.4) *provided* the Covariance $Q_a(x, x')$ in Equation (6.2) is known. In this case, the smoothness of the covariance kernel $q_a(x, x')$ is well known to determine the rate of decay of the eigenvalues to zero in the Karhúnen-Loève expansion (e.g. [32]). Moreover, approximate eigenpairs can be computed via variational methods using Finite Element subspaces, and rates of pointwise convergence in D can be established almost surely. We emphasize that to determine the Karhúnen-Loève expansion of the stochastic coefficient a explicit knowledge of the covariance kernel q_a in Equation (6.1) and (6.2) is required.

6.1.2. *Wavelet–expansion.* The Finite Element spaces $V_l = \mathcal{S}^{1,1}_{\Gamma_D}(D, \mathcal{T}_l)$, as defined in Equation (4.1), in the domain D are built on the nested sequence $\{\mathcal{T}_l\}_{l=0}^{\infty}$ of regular, simplicial triangulations τ_l, obtained by l uniform refinements of some initial, regular partition $\mathcal{T}_0$ of D into simplices $(K_0)_j$, $j = 1, ..., \#\mathcal{T}_0$. Therefore, for each $l \in \mathbb{N}_0$, every simplex $(K_l)_j \in \mathcal{T}_l$ is affinely equivalent to the reference simplex $\hat{K} = \{\hat{x} \in \mathbb{R}^d_+ : \|\hat{x}\|_1 < 1\}$: there are affine mappings

$$(F_l)_j : \hat{K} \ni \hat{x} \to x \in (K_l)_j \in \mathcal{T}_l,$$

such that, for all $j = 1, ..., \#(T_l)$,

$$\det|D(F_l)_j| = |(K_l)_j|/|\hat{K}| = \mathcal{O}(2^{-ld}) .$$

We observe that for every $p \geq 1$ and any regular, simplicial partition $\mathcal{T}$ of D holds

$$\nabla(\mathcal{S}^{p,1}_{\Gamma_D}(D, \mathcal{T})) \subseteq \mathcal{S}^{p-1,0}(D, \mathcal{T})^d \subset L^2(D)^d . \tag{6.5}$$

For any $L, q \in \mathbb{N}_0$, we have the orthonormal decomposition

$$\mathcal{S}^{q,0}(D, \mathcal{T}_L) = \bigoplus_{l=0}^{L} R_l,$$

where

$$R_l := \mathcal{S}^{q,0}(D, \mathcal{T}_l) \cap \mathcal{S}^{q,0}(D, \mathcal{T}_{l-1})^{\perp} \quad \text{if} \quad l \geq 1, \qquad \text{and} \quad R_0 := \mathcal{S}^{q,0}(D, \mathcal{T}_0) .$$

An $L^2(D)$-orthonormal basis of $\mathcal{S}^{q,0}(D, \mathcal{T})$ can be explicitly constructed as follows: let $\hat{\mathcal{T}}_0 = \{\hat{K}\}$ and define $\hat{\mathcal{T}}_1 = \{(\hat{K}_1)_j : j = 1, ..., 2^d\}$, the set of 2^d many simplices $(\hat{K}_1)_j$ that are obtained by regular subdivision of the reference simplex $\hat{K}$. We define for any $q \in \mathbb{N}_0$,

$$N_q := \dim(\mathcal{S}^{q,0}(\hat{K}, \hat{\mathcal{T}}_0)) = \begin{pmatrix} q+d \\ d \end{pmatrix},$$

and, for $d = 1, 2, ...$ and $q = 0, 1, ...,$

$$\tilde{N}_q := \dim(\mathcal{S}^{q,0}(\hat{K}, \hat{\mathcal{T}}_1) \cap \mathcal{S}^{q,0}(\hat{K}, \hat{\mathcal{T}}_0)^{\perp}) = (2^d - 1) \begin{pmatrix} q+d \\ d \end{pmatrix}.$$

Denote by $\{\hat{\varphi}_n\}_{n=1}^{N_q}$ an $L^2(\hat{K})$ orthonormal basis of

$$\hat{W}_0 := \mathcal{S}^{q,0}(\hat{K}, \hat{\mathcal{T}}_0) = \mathcal{P}_q(\hat{K})$$

and by $\{\hat{\psi}_n\}_{n=1}^{\tilde{N}_q}$ an $L^2(\hat{K})$ orthonormal basis of

$$\hat{W}_1 := \mathcal{S}^{q,0}(\hat{K}, \hat{\mathcal{T}}_1) \cap \mathcal{S}^{q,0}(\hat{K}, \hat{\mathcal{T}}_0)^{\perp} . \tag{6.6}$$

For $l = 0$ we define the basis Ψ_0 by

$$\Psi_0 := \left\{(\psi_0)_{j,n} | \, \forall (K_0)_j \in \mathcal{T}_0 : \ (\psi_0)_{j,n}|_{(K_0)_j} \circ (F_0)_j = \hat{\varphi}_n, ; \hat{\varphi}_k \in \hat{W}_0\right\} \tag{6.7}$$

and, for every $l \geq 1$, we define Ψ_l by

$$\Psi_l := \{(\psi_l)_{j,n} : j = 1, ..., \#(\mathcal{T}_{l-1}), \ n = 1, ..., \tilde{N}_q\}, \tag{6.8}$$

i.e. by the set of affine images of the (mother-wavelets) $\hat{\psi}_n$ under $(F_{l-1})_j$:

$$(\psi_l)_{j,n} \circ (F_{l-1})_j = \hat{\psi}_n, \quad l \geq 1, \ j = 1, ..., \#(\mathcal{T}_{l-1}), \ n = 1, ..., \tilde{N}_q.$$

By construction, $(\psi_l)_{j,n} = \hat{\psi}_k \circ ((F_{l-1})_j)^{-1}$ forms an $L^2(D)$ orthogonal system.

Proposition 6.1. *Assume that the elements $(\psi_l)_{j,n}$ of the sets Ψ_l defined in Equation (6.7) and (6.8) are $L^2(D)$ normalized, i.e. that*

$$((\psi_l)_{j,n}, (\psi_{l'})_{j',n'})_{L^2(D)} = \delta_{l,l'}\delta_{j,j'}\delta_{n,n'}, \quad \forall l, l' \in \mathbb{N}_0,\ j = 1, ..., \#(\mathcal{T}_l), j' = 1, ..., \#(\mathcal{T}_{l'}).$$

Then

$$L^2(D) = \bigoplus_{l=0}^{\infty} R_l, \quad \text{where} \quad R_l := \text{span}\{\Psi_l\}, \quad l \geq 0. \tag{6.9}$$

Proof. Since the $(\psi_l)_{j,n}$ are $L^2(D)$ orthonormal by construction, the algebraic sums $\Psi_0 + \Psi_1 + ...$ of subspaces are direct. Since, for every $L \in \mathbb{N}_0$ and every $q \in \mathbb{N}_0$

$$\mathcal{S}^{q,0}(D, \mathcal{T}_L) = \bigoplus_{l=0}^{L} \Psi_l \supseteq \mathcal{S}^{0,0}(D, \mathcal{T}_L) \tag{6.10}$$

and since the space of simple functions on the partition $\mathcal{T}_L$ coincides with $\mathcal{S}^{0,0}(D, \mathcal{T}_L)$, the sequence of subspaces defined in Equation (6.10) is dense in $L^2(D)$ as $L \to \infty$, which proves Equation (6.9). □

Every stochastic diffusion coefficient $a(\omega, x) \in L^2(\Omega; L^2(D)) = L^2(\Omega, \mathcal{A}, \mathbb{P}; L^2(D))$ admits, by Equation (6.9), a multi-wavelet expansion

$$a(\omega, x) = \sum_{l=0}^{\infty} \sum_{j=1}^{\tilde{N}_l} \sum_{n=1}^{\tilde{N}_q} (a_l)_{j,n}(\omega)(\psi_l)_{j,n}(x), \tag{6.11}$$

where the "coefficients" $(a_l)_{j,n}(\omega) \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R})$ are random variables defined by

$$(a_l)_{j,n}(\omega) = \int_D a(\omega, x)(\psi_l)_{j,n}(x)dx = (a(\omega, \cdot), (\psi_l)_{j,n})_{L^2(D)}. \tag{6.12}$$

The convergence in Equation (6.11) is, as in the case of the Karhúnen-Loève expansion, in $L^2(\Omega, L^2(D))$. However, unlike in the case of a Karhúnen-Loève expansion, in certain cases the Finite Element discretization of Equation (4.2) on mesh $\mathcal{T}_L$ coincides *exactly* with the discretization of a diffusion problem where the wavelet coefficient expansion, Equation (6.11), is truncated at level L.

Proposition 6.2. *Assume that the stochastic coefficient $a(\omega, x)$ in Equation (3.1) is given in the form of Equation (6.11). Denote for $1 \leq L < \infty$ by $a_L(\omega, x)$ the partial sum*

$$a_L(\omega, \cdot) = \sum_{l=0}^{L} \sum_{j=1}^{\tilde{N}_l} \sum_{n=1}^{\tilde{N}_q} (a_l)_{j,n}(\omega)(\psi_l)_{j,n} \tag{6.13}$$

and define the corresponding bilinear form $B_L(\cdot, \cdot)$ by

$$B_L(v, w) = \mathbb{E}\left[\int_D a_L(\omega, x)\nabla_x v \cdot \nabla_x w dx\right], \quad v, w \in L^2(\Omega; V). \tag{6.14}$$

Then, under the assumption

$$q \geq 2p - 2, \tag{6.15}$$

the bilinear forms $B(\cdot, \cdot)$ in Equation (3.6) and $B_L(\cdot, \cdot)$ in Equation (6.14) coincide on the FE spaces $\mathcal{S}^{p,1}(D, \mathcal{T}_L)$:

$$\forall v_L, w_L \in L^2(\Omega; \mathcal{S}^{p,1}(D, \mathcal{T}_L)) : \qquad B(v_L, w_L) = B_L(v_L, w_L). \tag{6.16}$$

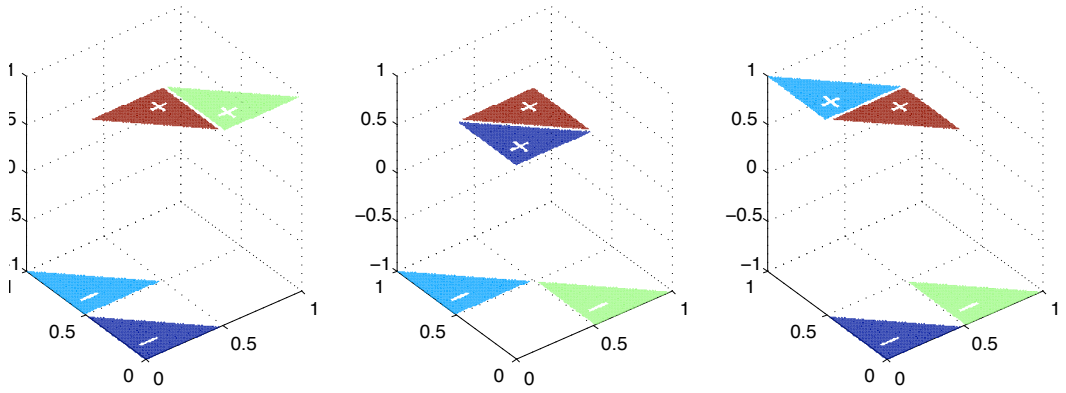


FIGURE 2. $\tilde{N}_q = 3$ Mother–Multi–wavelets in Equation (6.6) for $q = 0$ and $d = 2$.

Proof. The proof follows from the definition of $B(\cdot,\cdot)$ and of $B_L(\cdot,\cdot)$, upon noting that by Equation (6.5) for every $v_L, w_L \in \mathcal{S}^{p,1}(D,\mathcal{T}_L)$ it holds that $\nabla v_L \cdot \nabla w_L \in \mathcal{S}^{2p-2,0}(D,\mathcal{T}_L)$. The orthogonal sum property of the decomposition in Equation (6.9) then implies with Equation (6.15) the assertion. □

Remark 6.3. The identity, Equation (6.16), has the important implication that in the MLMC–FE method, in *one Finite Element simulation at mesh level* l the bilinear form $B(\cdot,\cdot)$ can be evaluated on the *exact* stochastic diffusion coefficient $a(\omega,x)$ in Equation (6.11) with $\mathcal{O}(N_l)$ work. This is easily verified from Equation (6.13) together with the identity in Equation (6.16).

Remark 6.4. From Equation (6.15) and Remark 5.9 we see that for linear scaling MLMC–FE methods for the most important spatial dimensions $d = 2, 3$, piecewise constant (i.e. $q = 0$ for $p = 1$) and piecewise quadratic (i.e. $q = 2$ for $p = 2$) discontinuous multiwavelets will have to be used in the wavelet representation, Equation (6.11), of the stochastic diffusion coefficient. For $q = 0$ in spatial dimension $d = 2$, the $\tilde{N}_q = 3$ generating mother–wavelets $\hat{\psi}_n$ are shown in Figure 2.

6.2. **Numerical example on** $D = [0,1]$**.** In our implementation the mesh $\mathcal{T}_l$ at level l is the family of intervals of the form $[(i-1)2^{-l}, i2^{-l}]$ for $i = 1, ..., 2^l$, the mesh width is then given by $h_l = 2^{-l}h_0 = 2^{-l}$, with 2^l elements per level. This results in a 2-shape regular mesh and the family $\{\mathcal{T}_l\}_{l=1}^{\infty}$ is nested. Here we employ Dirichlet boundary conditions, i.e. $S^{1,0}_{\Gamma_D}(D,\mathcal{T}_0) = S^{1,0}(D,\mathcal{T}_0) \cap H^1_0(D) = \{0\}$, which implies no degrees of freedom on the boundary. The hat basis $(b_l)_i$ at each level l is defined, for $i = 1, \ldots, 2^l - 1$, as:

$$(b_l)_i(x) = 2^l \begin{cases} x - (i-1)h_l & \text{for } x \in [(i-1)h_l, ih_l], \\ (i+1)h_l - x & \text{for } x \in [ih_l, (i+1)h_l], \\ 0 & \text{otherwise.} \end{cases} \tag{6.17}$$

We consider the following example adapted from [6]

Example 6.1. *Let* $D = [0,1]$, $\mathbb{E}_a(x) = 5 + x$, $q_a(x,x') = \frac{min\{x,x'\}+1}{2} \in H^1(D) \otimes H^1(D)$. *The corresponding eigenpairs in the Karhúnen-Loève expansion are given by* $\tilde{\lambda}_m = \frac{8}{\pi^2(2m-1)^2}$, $\tilde{\phi}_m(x) = \sin((x+1)/\sqrt{2\tilde{\lambda}_m})$, *for* $m \geq 1$. *The eigenvalues feature algebraic decay with rate* 2. *The data* f, *on the right hand side of Equation* (3.1), *is set equal to 1.*

The diffusion coefficient a expressed in the Karhúnen-Loève expansion (see Equation (6.3)) was truncated after the first term. With independent and $[-1,1]$-uniformly distributed random variables Y_k we may write

$$a_1(\omega, x) := 5 + x + \frac{2\sqrt{2}}{\pi} \cdot Y(\omega) \cdot \sin\left(\frac{\pi(x+1)}{4}\right).$$

This leads to an increased consistency error between the exact moment of the solution of Equation (3.1) and the MC-FE method, resp. MLMC-FE approximation.

To establish the error bounds, proven in the previous chapters, in simulations we calculate the exact solution of Equation (3.1). Therefore, we integrate Equation (3.1), given the stochastic diffusion coefficient a_1 and the right hand side $f = 1$, to obtain for the first moment:

$$\mathbb{E}[u(\omega,x)] = \sum_{i=0}^{\infty} \frac{2\sqrt{2}}{\pi(2i+1)} \int_0^x \frac{c-y}{5+y} \left(\frac{\sin\left(\frac{\pi(y+1)}{4}\right)}{5+y}\right)^{2i} dy,$$

The constant c is the solution of the above expression set to zero when integrating over the whole domain $D = [0,1]$. The integral was calculated with Mathematica [2], terminating the sum after $i = 5$, this leads to $c \approx 0.4850$. The integration in each term in the sum is tedious. For the simulations we terminate the series after $i = 5$. The L^2–norm between the expansion up to $i = 4$ and to $i \geq 5$ is of order $\mathrm{O}(10^{-10})$. Up to level $L = 10$, where $h_L \approx 10^{-3}$ we can neglect the remainder of the series. This error is insignificant given the accuracy of the approximation for simulations up to level $L = 10$ for point estimates and for the L^2-norm on the domain $D = [0,1]$.

The error estimates are calculated in the first order Sobolev semi-norm accordingly, given in Equation (3.5). The m-th order Sobolev semi-norm, for $m \in \mathbb{N}$, for sufficiently smooth $u : D \to \mathbb{R}$ is defined as

$$|u|^2_{H^m(D)} := \sum_{\alpha \in \mathbb{N}^d, |\alpha| \leq m} \int_D |D^\alpha u|^2 dx.$$

For $u_l \in V_l = S^{1,1}_{\Gamma_D}(D, \mathcal{T}_l)$ we have

$$|u_l|^2_{H^1(D)} = \int_D \Big(\sum_{i=1}^{2^l-1} (u_l)_i \nabla (b_l)_i(x)\Big)\Big(\sum_{i=1}^{2^l-1} (u_l)_i \nabla (b_l)_i(x)\Big)\, dx.$$

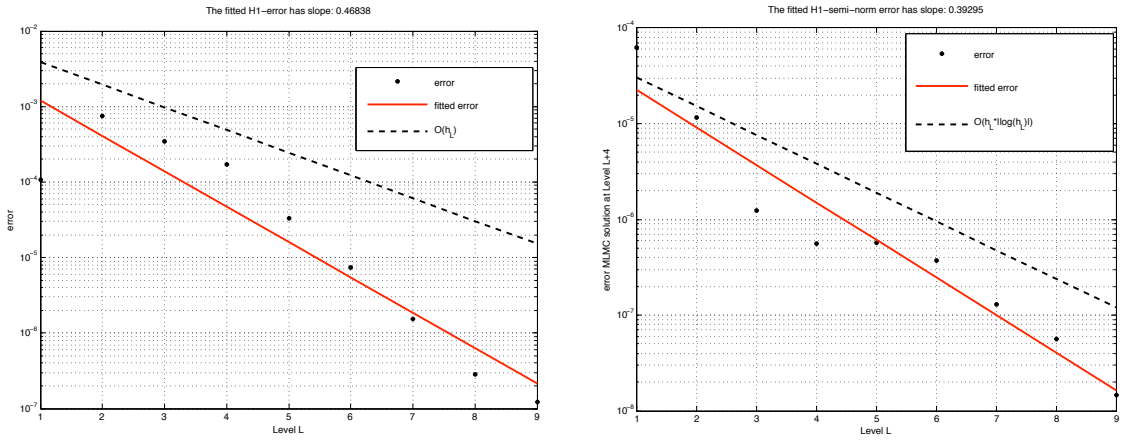
If we denote by S_l the stiffness matrix of the Laplace operator with respect to the nodal hat basis b_l at level l we may write

$$|u_l|_{H^1(D)} = (u_l S_l u_l)^{1/2} \tag{6.18}$$

The simulation was carried out on a Computer with a two GHz processor with one GB RAM using Matlab [3]. As a solver we used the backslash operator in Matlab.

[2] Wolfram Research, Inc., Mathematica, Version 7.0; Champaign, IL (2008).

[3] MATLAB, version 7.9.0.529 (R2009b); Natick, Massachusetts: The MathWorks Inc., 2009.

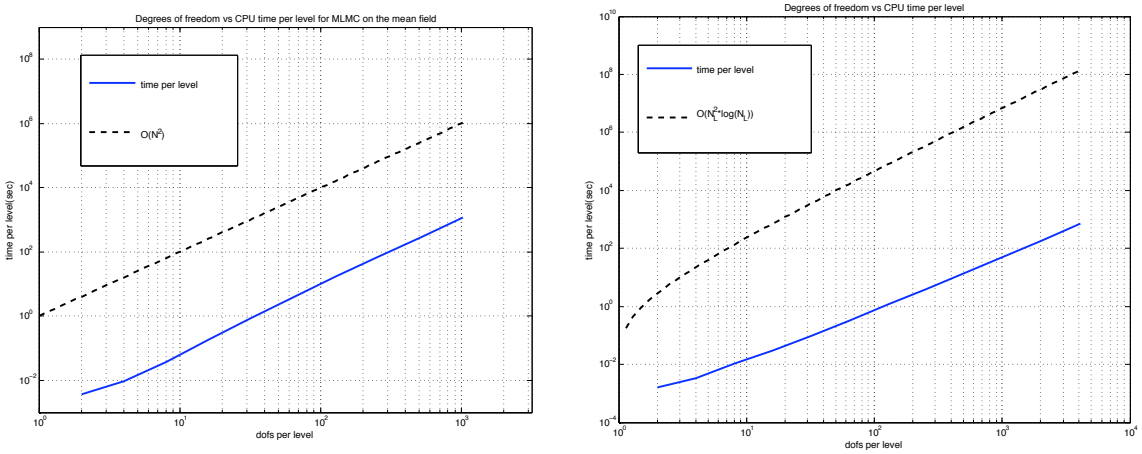


(a) (b)

FIGURE 3. Rate of convergence of the MLMC–FE method with respect to the H^1-semi-norm for the approximation of $\mathbb{E}(u)$ in (a) and $\mathcal{M}^2(u)$ in (b) in dimension $d = 1$ against the level.

6.2.1. *MLMC-FE method for the approximation of* $\mathbb{E}[u]$. Our aim is to verify the theoretical approximation error for the MLMC–FE method $\|\mathbb{E}[u] - E^L[u]\|_V$, given in Theorem 4.5. Therefore, we consider the nested family $\{\mathcal{T}_l\}_{l=1}^L$ and the spaces $V_l = S^{1,1}(D, T_l)$ with basis functions $\{(b_l)_i\}_{i=1}^{2^l-1}$, defined in Equation (6.17), on each level $l = 1, \ldots, L$. Each level is constructed by adding the mid points between two vertices to the mesh of the previous level. Thus, we get $2^{(l-1)}$ additional linear independent basis functions passing from V_{l-1} to V_l. To construct the single scale basis $\{(b_l)_i\}_{i=1}^{2^l-1}$, we transform each basis function of level $l-1$ into the basis function of level l plus $2^{(l-1)}$ additional basis functions. This allows us to calculate $E^L[u]$, since we need to subtract the solution in V_{l-1} from the solution in V_l (see Equation (4.9)). For each sample of the stochastic coefficient on each level we assemble the stiffness matrix and solve the deterministic system of equations, given in Equation (3.6), to obtain u_l^i. With this and the exact solution we get with Equation (6.18) the desired error.

The rate of convergence of the MLMC–FE approximation, depending on the level L, is displayed in Figure 3(a). The theoretical convergence rate of Theorem 4.5, $\mathrm{O}(h_L)$, is resembled in the simulation (as indicated by the reference slope). Figure 4(a) shows the total CPU–time needed to calculate $E^L[u]$ for different levels L. It reflects the calculated expected behavior of the total work $\mathrm{Work}(L) \leq C_\epsilon N_L^2$ in Theorem 4.5 for $d = 1$. In Figure 5(a) is the CPU–time



(a) (b)

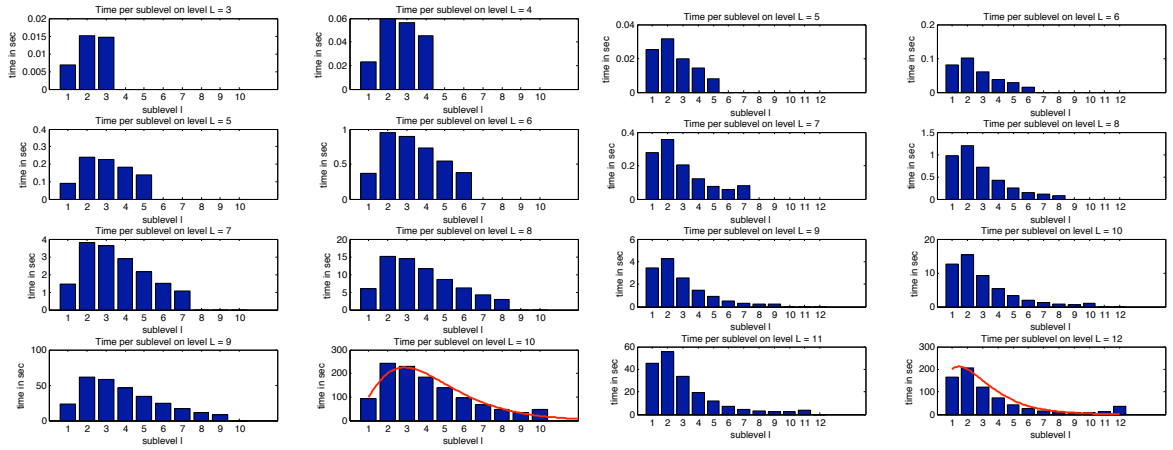
FIGURE 4. Total CPU time for the MLMC–FE approximation of $E^L(u)$ in (a) and $\mathcal{M}^2(u)$ in (b) in $1d$ against the level.

per sublevel l, for $l = 1, \ldots, L$, depicted. For M_l as in Theorem 4.5 the CPU–time at each sublevel l is of rate $\mathrm{O}(l^2\, 2^{-l})$.

6.2.2. *MLMC-FE method for the approximation of* $\mathcal{M}^2[u]$. The calculation of the second moment $\mathcal{M}^2 u$ is performed in three steps. First we set up the linear Finite Element equation using a standard nodal hat basis for a given level l, as described above, to compute u_l^i. In a second step, we transform the result into a hierarchic B–spline linear wavelet basis. Finally, in a third step, we generate the sparse tensor product by implementing Equation (5.6), the sparse tensor projection. This algorithm is repeated for each level and, according to Equation (5.8), this leads to the MLMC–FE approximation $\hat{E}^L((u_L)^2)$ of $\mathcal{M}^2 u$.

In Figure 3(b) we compare the sparse tensor product solution to the sparse tensor product of the solution $\hat{E}^L((u_{L+4})^2)$. The error resembles the theoretical results of Theorem 5.7, as the reference slope indicates. The total CPU–time in dependence of the degrees of freedom has quadratic growth as stated in Equation (5.11), is displayed in Figure 4(b). Figure 5(b) shows the CPU–time on all sublevels l, for $l = 1, \ldots, L$ for a fixed level L. Theoretically this is, for fixed level L, $M_l\, N_l = \mathrm{O}(l\, 2^{-l})$, with M_l as in Theorem 4.5.

6.3. **Numerical example on** $D = [0,1]^2$**.** We consider the unit square and define level $l = 0$ to be the space of the boundary basis functions with four vertices $P_1 = (0,0)$, $P_2 = (1,0)$,



(a) (b)

FIGURE 5. CPU–time per sublevel for different levels for the MLMC–FE approximation of $\mathbb{E}(u)$ in (a) and $\mathcal{M}^2(u)$ in (b) in dimension $d = 1$.

$P_3 = (1, 1)$ and $P_4 = (0, 1)$, the triangulation of the unit square is given by the triangles $P_1P_2P_4$ and $P_2P_3P_4$. Given the Dirichlet boundary condition the simulation on this level is superfluous. The nested family $\mathcal{T}_{l+1}$ is constructed by dividing each triangle of level l into four congruent triangles of the same size. The resulting mesh is then $(2 - \sqrt{2})$-shape regular. Figure 6 shows the $\mathcal{T}_2$ mesh. Similar to the example in one space dimension, we apply Dirichlet boundary conditions. We adapt Example 6.1 to $\mathbb{R}^2$ as follows:

Example 6.2. *Let $D = [0, 1]^2$ and choose the sequence $\{f_m\}_{m\geq 1}$ as the tensor product of the sequence from Example 6.1, ordered by the magnitude of the resulting eigenvalues $\{\tilde{\lambda}\}_{m\geq 1}$. The eigenvalues λ_m are chosen as $\lambda_m = (\tilde{\lambda}_m)^\theta$ with $\theta = 2.5$, such that the algebraic decay of $\{\lambda\}_{m\geq 1}$ is of rate $5/2$.*

The simulation in dimension $d = 2$ was carried out on a cluster compute server with AMD Opteron Processors, between 2.4 and 2.8 GHz per core. We used here, as in the case $d = 1$, the backslash operator to solve each linear system.

6.3.1. *MLMC–FE method for the approximation of* $\mathbb{E}[u]$. For $d = 2$ we did not calculate the exact solution $\mathbb{E}[u_L]$ as before. In this case the reference solution is either the solution of a Monte Carlo simulation with high sample count (10000 samples) or the solution of the MLMC–FE simulation on level $L+1$. Further, we did not integrate the entries of the stiffness matrix,

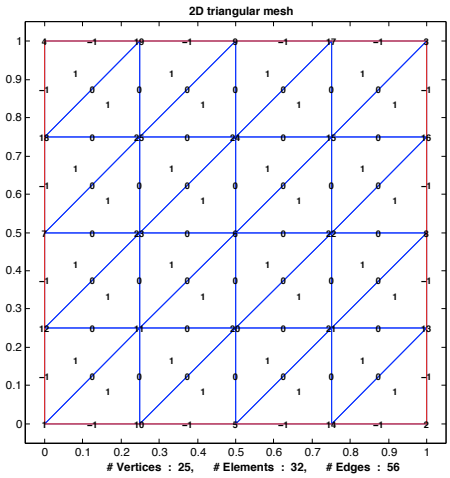
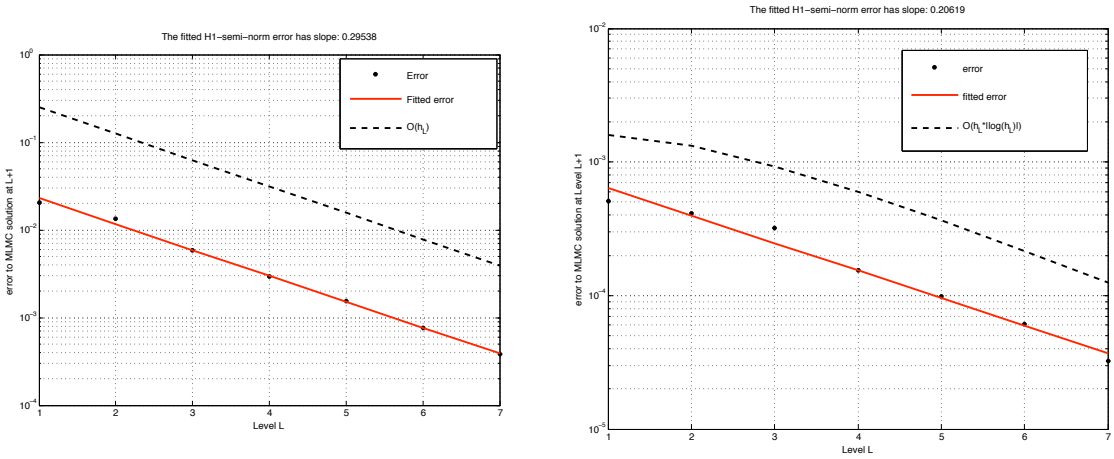


FIGURE 6. Grid for level 2 ($d = 2$)

B given in Equation (3.6), exactly. Here we use a seven point Gaussian quadrature rule of order six. Figure 7(a) shows the error of the MLMC–FE approximation for the mean field in dependence of the level. The theoretical results from Theorem 4.5 are resembled. For the total computational costs we calculated in Theorem 4.5, for $d = 2$, $\mathrm{Work}(L) = \mathrm{O}(N_L(\log N_l)^{3+\epsilon})$. This is also apparent in Figure 8(a). We deduce that the MLMC–FE method has log–linear computational time, whereas the convergence is the same as in the Monte Carlo method. The results can be compared to those in Figure 4(a) for $d = 1$, where the CPU–time was quadratic. The work load on each sublevel l, for $l = 1, \ldots, L$ can easily be computed as $\mathrm{O}(l^2)$, matching the results of the simulation in Figure 9(a).

6.3.2. *MLMC–FE method for the second moment.* For the calculation of the error of the MLMC–FE method for $\mathcal{M}^2[u]$ we proceed as in the one–dimensional case. Results in Figure 7(b) reflect the theoretical error for the MLMC–FE approximation of rate $\mathrm{O}(h_L(\log h_L)^{3/2})$ as stated in Theorem 5.7 for $k = 2$ and $s = 1$. Equally the simulation results on the total CPU-time pictured in Figure 8(b) for the theoretical result in Equation 5.11 ($\widehat{W}(L) = \mathrm{O}(N_L(\log N_L))$), slight differences in the rate of convergence are due to the nonoptimality of the backslash operator for large systems. This indicates that the use of a Multigrid method

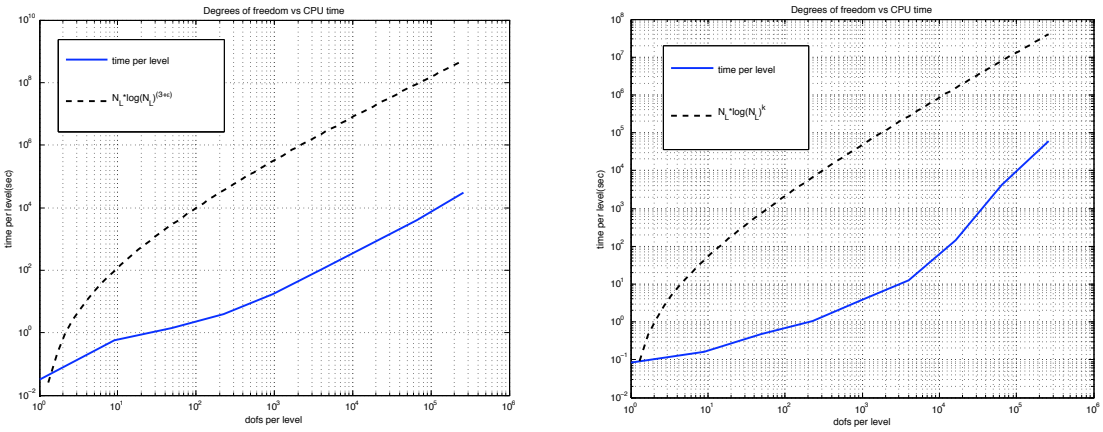


(a) (b)

FIGURE 7. Rate of convergence of the MLMC–FE method with respect to the H^1-semi-norm for the approximation of $\mathbb{E}(u)$ in (a) and $\mathcal{M}^2(u)$ in (b) in dimension $d = 2$ against the level.

should give optimal results for large L. For the CPU–time on the sublevels l, for $l = 1, \ldots, L$, we have a linear growth with increasing sublevel. This can be easily calculated with M_l as in Equation (5.10) and $N_l = 2^{2l}$. The results in Figure 9(b) are influenced by the use of the backslash operator, such that the linear behaviour is not fully reflected.

Remark 6.5. The serial Monte Carlo and therefore the MLMC–FE simulations can easily be implemented for parallel computing. No communication between processes is needed during execution if available memory allows for the handling of one sample on one processor. This means load balancing can easily be achieved without the need for communication between the processors during execution. However, the random number streams for each processor should be low correlated. The results are gathered at the end of the computation. In the case of a single–level MC-FE method each result can be added independently to the result of any other processor and data loss only leads to a reduction in the convergence speed. In the case of a MLMC–FE method the correct order of summation of the final result on each level has to be respected. This means that the MLMC–FE simulation is much more sensitive to data loss then is the MC–FE simulation. In our example the sample size of the finest level was normalized to one. In dimension $d = 1$ for the calculation of the mean field and the second moment we have decreasing load with increasing mesh refinements (see Figure 5). If



(a) (b)

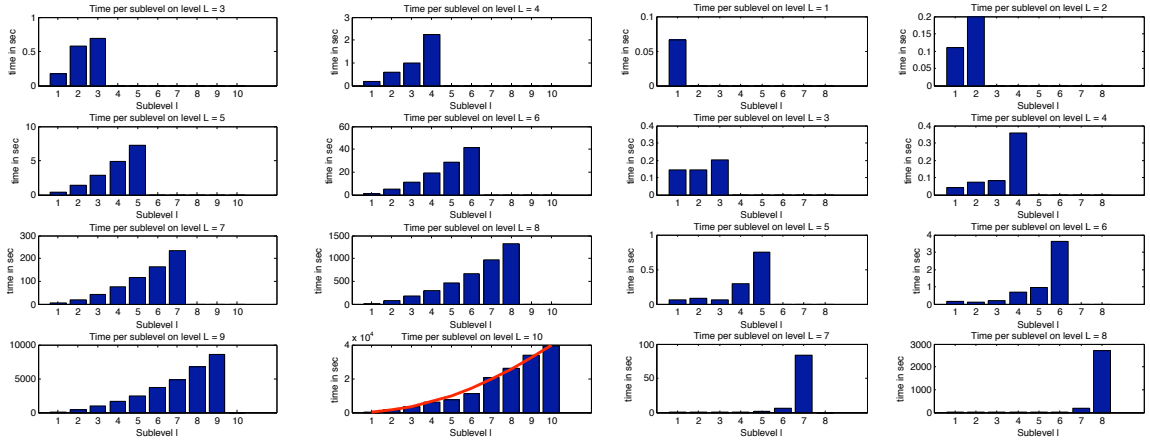
FIGURE 8. Total CPU time for the MLMC–FE approximation of $E^L(u)$ in (a) and $\mathcal{M}^2(u)$ in (b) in $2d$ against the level.

the calculation on the finest sublevel (normalized to one sample) can be performed on one processor, we can balance the work load for the other sublevels, with higher work load due to higher sample sizes, by splitting the random streams to more than one processor. In the case of $d = 2$ we have an increasing work load with finer meshes. This means that load balancing can only be achieved by domain decomposition, or, depending on the number of processors available, both, domain decomposition and sample splitting.

For all the simulations we truncated the sum in the Karhúnen-Loève expansion after the first term. This truncation could be coupled to the degrees of freedom of the spatial approximation as well, or it could be fixed to some higher term. This leads to a more complex calculation of the exact solution and the stiffness matrix, but also to an error reduction in the approximation of the exact moments of the solution. An algorithm to generate correlated Gaussian random fields is given in [23].

7. Conclusions

Our error and complexity analysis reveals that for low order Finite Element discretizations in the physical domain D, the proposed MLMC–FE method achieves an approximation of the mean field of the random solution and its k-th moments with *efficiency (i.e. error versus*



(a) (b)

FIGURE 9. CPU–time per sublevel for different levels for the MLMC–FE approximation of $\mathbb{E}(u)$ in (a) and $\mathcal{M}^2(u)$ in (b) in dimension $d = 2$.

computational work) which is comparable to one solve of a linear complexity Finite Element method for a deterministic elliptic problem of the same type.

The analysis also shows that *preservation of increased convergence rates of higher order Finite Element methods for the approximation of stochastic solutions with higher spatial regularity entails corresponding increase of the MC samples at each mesh level.* This implies a loss of the overall log-linear complexity of the MLMC–FE scheme. For problems with random solutions that exhibit high spacial smoothness as well as high summability, higher convergence rates of the overall discretization scheme will require apart from high order Finite Element methods in the physical domain also improved discretization strategies in the stochastic domain such as spectral, polynomial chaos based discretizations (see, e.g. [11, 12, 2, 1, 28, 27, 7, 6]).

Therefore, the MLMC–FE method proposed here is competitive for stochastic PDE problems whose solutions have low smoothness in physical space, and moderate summability in $\omega \in \Omega$ as, e.g., finite second moments. This is typically the case in Gaussian models of porous media where realizations of a are, roughly speaking, Hölder continuous with exponent at most $1/2$.

In closing, we emphasize that the presently proposed MLMC—FE method does not require stationarity or Gaussianity of the stochastic diffusion coefficient a in any way. For stationary, Gaussian random inputs, linear scaling simulation methods can be built on tensorized

Fast Fourier Transform algorithms (see, e.g. [14] and the references therein). Both methods proposed here, the Karhúnen-Loève based and the wavelet based parametrization of $a(\omega, x)$, allow for non–stationary and irregular random inputs.

The representations in Equation (6.11) and Equation (6.13) can, due to the $L^2(D)$ orthogonality in Equation (6.12) of the (multi) wavelets be utilized *directly* in scenario generation based on a stream of "coefficient realizations" of a, possibly in digital form with uniform pixel resolution $\geq L$. In this case, Equation (6.16) implies that a "forward" MLMC—FE simulation on mesh $\mathcal{T}_L$ can account for all available data on a *exactly*.

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 (2007), pp. 1005–1034 (electronic).

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

[3] A. Barth, *A finite element method for martingale-driven stochastic partial differential equations*, Comm. Stoch. Anal., 4 (2010).

[4] A. Barth and A. Lang, *Almost sure convergence of a galerkin–milstein approximation for stochastic partial differential equations*. in review, February 2010.

[5] A. Barth and A. Lang, *Simulation of stochastic partial differential equations using finite element methods*. in review, March 2010.

[6] M. Bieri, R. Andreev, and C. Schwab, *Sparse tensor discretization of elliptic SPDEs*, SIAM J. Sci. Comput., 31 (2009/10), pp. 4281–4304.

[7] M. Bieri and C. Schwab, *Sparse high order FEM for elliptic sPDEs*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1149–1170.

[8] D. Braess, *Finite Elemente*, Springer, Berlin, 3rd ed., 2002.

[9] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer, Berlin, 2nd ed., 2002.

[10] A. Cohen, *Numerical analysis of wavelet methods*, Studies in Mathematics and its Applications, North-Holland, Amsterdam, 2003.

[11] A. Cohen, R. DeVore, and C. Schwab, *Convergence rates of best n-term approximations for a class of elliptic spdes*, to appear in Found. of Comp. Math., (2010).

[12] ———, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic pdes*, to appear J. Analysis and Applications, (2011).

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

[14] C. R. Dietrich and G. N. Newsam, *Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix*, SIAM J. Sci. Comput., 18 (1997), pp. 1088–1107.

[15] M. Geissert, M. Kovács, and S. Larsson, *Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise*, BIT Numerical Mathematics, 49 (2009), pp. 343–356.

[16] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[17] M. Giles, *Improved multilevel monte carlo convergence using the milstein scheme*, Preprints NA-06/22, Oxford University Computing Laboratory, Parks Road, Oxford, U.K., 2006.

[18] M. B. Giles, *Multilevel monte carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[19] M. Griebel and P. Oswald, *On additive Schwarz preconditioners for sparse grid discretizations*, Numer. Math., 66 (1994), pp. 449–463.

[20] P. Grisvard, *Elliptic problems in nonsmooth domains*, vol. 24 of Monographs and Studies in Mathematics, Pitman (Advanced Publishing Program), Boston, MA, 1985.

[21] S. Heinrich, *Multilevel Monte Carlo Methods*, in Large-Scale Scientific Computing, Third International Conference, LSSC 2001, Sozopol, Bulgaria, June 6-10, 2001, Revised Papers, S. Margenov, J. Wasniewski, and P. Y. Yalamov, eds., vol. 2179 of Lecture Notes in Computer Science, Springer, 2001, pp. 58–67.

[22] M. Kovács, S. Larsson, and F. Saedpanah, *Finite element approximation of the linear stochastic wave equation with additive noise*, SIAM J. Numer. Anal., 48 (2010), pp. 408–427.

[23] A. Lang and J. Potthoff, *Fast simulation of Gaussian random fields*. in review, April 2009.
[24] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer-Verlag, New York, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
[25] R. L. Naff, D. F. Haley, and E. Sudicky, *High-resolution monte carlo simulation of flow and conservative transport in heterogeneous porous media 1. methodology and flow results*, Water Resour. Res., 34 (1998), pp. 663–677.
[26] ———, *High-resolution monte carlo simulation of flow and conservative transport in heterogeneous porous media 2. transport results*, Water Resour. Res., 34 (1998), pp. 679–697.
[27] F. Nobile, R. Tempone, and C. G. Webster, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442.
[28] ———, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345.
[29] C. Schwab and R.-A. Todor, *Sparse finite elements for elliptic problems with stochastic loading*, Numer. Math., 95 (2003), pp. 707–734.
[30] C. Schwab and R. A. Todor, *Sparse finite elements for stochastic elliptic problems—higher order moments*, Computing, 71 (2003), pp. 43–63.
[31] C. Schwab and R. A. Todor, *Karhunen-Loève approximation of random fields by generalized fast multipole methods*, J. Comput. Phys., 217 (2006), pp. 100–122.
[32] ———, *Karhunen-Loève approximation of random fields by generalized fast multipole methods*, J. Comput. Phys., 217 (2006), pp. 100–122.
[33] T. von Petersdorff and C. Schwab, *Numerical solution of parabolic equations in high dimensions*, M2AN Math. Model. Numer. Anal., 38 (2004), pp. 93–127.

(Andrea Barth)
ETH, Seminar für Angewandte Mathematik
Rämistrasse 101
8092 Zürich
E-mail address: andrea.barth@math.ethz.ch

(Christoph Schwab)
ETH, Seminar für Angewandte Mathematik
Rämistrasse 101
8092 Zürich
E-mail address: schwab@math.ethz.ch

(Nathaniel Zollinger)
ETH, Seminar für Angewandte Mathematik
Rämistrasse 101
8092 Zürich
E-mail address: zollingn@student.ethz.ch

Research Reports

No.	Authors/Title
10-18	*A. Barth, C. Schwab and N. Zollinger* Multi-Level Monte Carlo Finite Element method for elliptic PDE's with stochastic coefficients
10-17	*B. Kågström, L. Karlsson and D. Kressner* Computing codimensions and generic canonical forms for generalized matrix products
10-16	*D. Kressner and C. Tobler* Low-Rank tensor Krylov subspace methods for parametrized linear systems
10-15	*C.J. Gittelson* Representation of Gaussian fields in series with independent coefficients
10-14	*R. Hiptmair, J. Li and J. Zou* Convergence analysis of Finite Element Methods for $H(\mathrm{div};\Omega)$-elliptic interface problems
10-13	*M.H. Gutknecht and J.-P.M. Zemke* Eigenvalue computations based on IDR
10-12	*H. Brandsmeier, K. Schmidt and Ch. Schwab* A multiscale hp-FEM for 2D photonic crystal band
10-11	*V.H. Hoang and C. Schwab* Sparse tensor Galerkin discretizations for parametric and random parabolic PDEs. I: Analytic regularity and gpc-approximation
10-10	*V. Gradinaru, G.A. Hagedorn, A. Joye* Exponentially accurate semiclassical tunneling wave functions in one dimension
10-09	*B. Pentenrieder and C. Schwab* *hp*-FEM for second moments of elliptic PDEs with stochastic data. Part 2: Exponential convergence
10-08	*B. Pentenrieder and C. Schwab* *hp*-FEM for second moments of elliptic PDEs with stochastic data. Part 1: Analytic regularity
10-07	*C. Jerez-Hanckes and J.-C. Nédélec* Asymptotics for Helmholtz and Maxwell solutions in 3-D open waveguides