Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Strong convergence analysis of iterative solvers for random operator equations

L. Herrmann

Research Report No. 2017-35

July 2017
Latest revision: August 2018

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

Funding SNF: 159940

Strong convergence analysis of iterative solvers for random operator equations *

Lukas Herrmann †

†Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, CH–8092 Zürich, Switzerland.
lukas.herrmann@sam.math.ethz.ch

August 13, 2018

Abstract

For the numerical solution of linear systems that arise from discretized partial differential equations, multigrid and domain decomposition methods are well established. Multigrid methods are known to have optimal complexity and domain decomposition methods are in particular useful for parallelization of the implemented algorithm. For random operator equations, the classical theory is not directly applicable, since condition numbers of system matrices may be close to degenerate due to non-uniform random input. It is shown that iterative methods converge in the strong, i.e., L^p, sense if the random input satisfies certain integrability conditions. As a main result, a wide range of standard multigrid and domain decomposition methods are applicable in the case of elliptic partial differential equations with lognormal diffusion coefficients and converge strongly with deterministic bounds on the computational work which are essentially optimal. This enables the application of multilevel Monte Carlo methods with rigorous, deterministic bounds on the computational work.

1 Introduction

Mathematical models of partial differential equations (PDEs) with random input receive increasing attention in recent years. In particular, diffusion equations with random coefficients to model random media are considered and respective system responses or quantities of interest are studied, cf. [17, 16, 13, 14, 37, 27]. Generally, the random input is function valued and will be referred to as random field. In the numerical analysis of these problems constants in error estimates become random variables and may have a distribution with unbounded support. In the case that quantities of interest are moments of system responses, numerical analysis has been performed if certain integrability conditions are satisfied by the random system input. There, the case of Gaussian random fields (GRFs) as random inputs is frequently considered. Multilevel techniques such as multilevel Monte Carlo (MLMC) have been established to accelerate the approximation of moments. There, sample numbers are chosen in a greedy technique to optimize the error versus the required computational cost, cf. [3, 14, 37, 26, 22, 23]. The requirement of MLMC to be efficient is a small variance between higher levels or to put it differently, the strong error between higher levels has to converge with an available rate. In the case that finite element

*This work is supported by the Swiss National Science Foundation under SNF 159940

(FE) discretizations are used and the random input is unbounded such as GRFs, the numerical analysis and numerical experiments presented in these references relied on sparse direct solvers, cf. [14, 37, 27]. Hence, the applicability of iterative methods which are known to be fast for deterministic problems is of natural interest, e.g., so called full multigrid is well known to have *optimal* complexity in the case of Poisson's equation, cf. [2].

In this paper we establish rigorously the strong convergence of a wide class of standard iterative solvers, which yields essentially optimal computational cost estimates also in the case of GRF input with low spatial regularity. As an application, MLMC is discussed with deterministic, essentially optimal estimates of the computation cost, which was previously unknown. By optimal, we mean that solutions of linear systems under consideration with dimension $\mathcal{O}(N)$ may be approximated in computational cost $\mathcal{O}(N)$ consistently with the overall discretization error. In the computational uncertainty quantification (UQ) literature, iterative solvers have been considered mostly in the context of stochastic collocation and stochastic Galerkin, cf. [39, 35, 19]. A particular variant of MLMC with multigrid for GRF inputs has been proposed in [30] and computational experiments have been performed.

The present manuscript analyzes the applicability and strong convergence of well established iterative methods for operator equations with unbouded random input in a general setting. Let $\mathcal{A}$ be a random, continuous linear operator from $\mathcal{V}$ to $\mathcal{V}^*$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that is $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.) boundedly invertible, where $\mathcal{V}$ is a Hilbert space and $\mathcal{V}^*$ its dual space. Let its expectation $\mathbb{E}(\mathcal{A})$ be well defined. In the present paper we are interested in the numerical analysis of approximations of the solution u to the linear equation that for $\mathbb{P}$-almost every ($\mathbb{P}$-a.e.) $\omega \in \Omega$

$$\mathcal{A}(\omega)u(\omega) = f, \tag{1}$$

where $f \in \mathcal{V}^*$ is deterministic, in the strong sense by iterative methods such as multigrid and domain decomposition methods. This can be rewritten in variational form to find $u : \Omega \to \mathcal{V}$ such that for $\mathbb{P}$-a.e. ω

$$a_\omega(u(\omega), v) := {}_{\mathcal{V}^*}\langle \mathcal{A}(\omega)u(\omega), v\rangle_{\mathcal{V}} = {}_{\mathcal{V}^*}\langle f, v\rangle_{\mathcal{V}} \quad \forall v \in \mathcal{V}.$$

Let us assume that there are strictly positive random variables $\hat{a}$ and $\check{a}$ such that for $\mathbb{P}$-a.e. ω

$$\check{a}(\omega)\ {}_{\mathcal{V}^*}\langle \mathbb{E}(\mathcal{A})v, v\rangle_{\mathcal{V}} \leq {}_{\mathcal{V}^*}\langle \mathcal{A}(\omega)v, v\rangle_{\mathcal{V}} \leq \hat{a}(\omega)\ {}_{\mathcal{V}^*}\langle \mathbb{E}(\mathcal{A})v, v\rangle_{\mathcal{V}} \quad \forall v \in \mathcal{V}. \tag{2}$$

We will be particularly interested in the case that $\hat{a}$ and $\check{a}^{-1}$ are unbounded random variables. This is for example the case for elliptic PDEs with lognormal coefficients. Thus, preconditioned finite dimensional discretizations suffer from random condition numbers that are unbounded and respectively interative methods contract with random contraction numbers, which may have realizations arbitrarily close to one with positive probability. At first sight one may overcome this with random interation numbers specified by a threshold of residual errors with the disadvantage of the occurrence of large iteration numbers, when samples of the random contraction number are close to one. Also, bounds on the computational work for such strategies would be probabilistic. A main new contribution of this paper is that deterministic iteration numbers exist that allow for strong convergence, i.e., convergence of well known iterative methods in the $L^q(\Omega; \mathcal{V})$-norm, $q \in [1, +\infty)$, such as multigrid, multilevel preconditioned conjugate gradient, or domain decomposition. This is possible due to tail bounds of the random contraction numbers, which for example are satisfied in the important case of elliptic PDEs with lognormal coefficients. As a consequence, deterministic, essentially optimal complexity bounds are implied for the solution of resulting random linear systems when well known multigrid or domain decomposition methods

are applied. This enables also rigorous, deterministic, essentially optimal complexity bounds for MLMC approximations of mean fields, which was previously unknown. Assumptions on the computational cost of PDE solvers that were made in previous papers [14, 37, 27] to obtain complexity bounds of MLMC and partly to calibrate the MLMC estimator are pervaded by the new theory presented in this manuscript. We will treat the case that $a_\omega(\cdot,\cdot)$ is symmetric for $\mathbb{P}$-a.e. ω. However, the presented theory can be extended to certain non-symmetric operators, see for example cf. [12, Section 11].

In Section 2, we will review iterative methods as they were formulated in [40] in order to discuss various multilevel method in a unified framework as well as the more classical approach, cf. [5, 43]. It will also be highlighted which parts in the framework and in the iterative methods are random. As a main result, we will develop integrability conditions on the random contraction numbers in Section 3 that result in sufficiently strong tail bounds in order to ensure strong convergence in the setting of multi and single-level discretizations of (1). The integrability conditions posed in Section 3 are analyzed for several multilevel methods such as multigrid, the so called BPX preconditioner, cf. [10], and domain decomposition methods in Sections 4 and 5. An important application of the presented theory are lognormal diffusion problems, which are briefly reviewed in Section 6. In particular deterministic bounds on the computational work without assumptions on the PDE solver of MLMC are implied. Numerical experiments with GRF input are presented in Section 7 and confirm the theoretical analysis.

2 Iterative methods

In this section, we review well known iterative methods to approximate solutions to linear equations on a finite dimensional inner product space $(V,(\cdot,\cdot))$, where $\|\cdot\|$ denotes the norm that is induced by $(\cdot,\cdot)$. Let us consider the random linear equation that for $\mathbb{P}$-a.e. ω

$$A(\omega)u(\omega) = f, \tag{3}$$

where $A : V \to V$ is a random linear operator that is $\mathbb{P}$-a.s. symmetric and positive definite (SPD) with respect to $(\cdot,\cdot)$. Hence, (3) is $\mathbb{P}$-a.s. uniquely solvable. Note that we will often omit dependencies of random quantities on ω for notational convenience. Let us denote the bilinear form that is induced by A by $(\cdot,\cdot)_A$ and let $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ denote the maximal and minimal eigenvalue of A. The condition number of A is denoted $\kappa(A)$ and $\rho(A)$ denotes the spectral radius. This notation will also be used for other linear operators that occur. Note that since A is random, $\lambda_{\max}(A)$, $\lambda_{\min}(A)$, $\kappa(A)$, and $\rho(A)$ are random variables. In particular, the review article [40] enables the discussion of multigrid and domain decomposition methods in an unified framework. These methods allow in some cases for optimal preconditioning or uniform contraction numbers. In this section, we will mainly follow [40] and introduce abstract algorithms, which in later sections will be used as BPX or additive Schwarz preconditioner, (un)symmetric multigrid and overlapping domain decomposition method. We will also highlight which of the occuring objects in this review section are random.

Since A is SPD $\mathbb{P}$-a.s., the well known conjugent gradient (CG) method implies after $n \in \mathbb{N}$ iterations with initial guess U^0 the error bound that for $\mathbb{P}$-a.e. ω

$$\|u(\omega) - U^n(\omega)\|_{A(\omega)} \leq 2\left(\frac{\sqrt{\kappa(A(\omega))}-1}{\sqrt{\kappa(A(\omega))}+1}\right)^n \|u(\omega) - U^0(\omega)\|_{A(\omega)}.$$

Since the random condition number $\kappa(A)$ may depend on the dimension of the linear space V, we consider the preconditioned linear system that for $\mathbb{P}$-a.e. ω

$$B(\omega)A(\omega)u(\omega) = B(\omega)f, \tag{4}$$

where the random linear operator B is chosen to be SPD with respect to $(\cdot,\cdot)$. The random operator B shall satisfy that $\mathbb{P}$-a.s. $\kappa(BA) \leq \kappa(A)$, which then accelerates the convergence of the CG method. The combination of preconditioning and CG will be referred to as preconditioned conjugent gradient (PCG) method.

Another method to be discussed is the linear iterative method of the form

$$U^{n+1}(\omega) = U^n(\omega) + B(\omega)(f - A(\omega)U^n(\omega)), \quad n \in \mathbb{N}_0, \tag{5}$$

for $\mathbb{P}$-a.e. ω, where B is a suitable random operator that is not nessarily symmetric and U^0 is given. Note that this linear iterative methods converges $\mathbb{P}$-a.s. if $\|\operatorname{Id} - BA\|_A < 1$, $\mathbb{P}$-a.s. Alternatively, one could also introduce a relaxed version of B with relaxation parameter in $(0, 2/\rho(BA))$ to guarantee convergence with random contraction number $(\kappa(BA)-1)/(\kappa(BA)+1)$, cf. [40, Proposition 2.3]. Also we remark that generally the contraction number of the PCG method is smaller, cf. [40, Proposition 2.2], which is why one may say that PCG accelerates convergence.

Lemma 2.1 *Let $\tilde{A} : V \to V$ be a SPD operator with respect to $(\cdot,\cdot)$ and let $\tilde{B} : V \to V$ be a SPD preconditioner with respect to $(\cdot,\cdot)$ for $\tilde{A}$. If there exists positive random variables c_0, c_1 such that for $\mathbb{P}$-a.e. $\omega \in \Omega$*

$$c_0(\omega)(v,v)_{\tilde{A}} \leq (v,v)_{A(\omega)} \leq c_1(\omega)(v,v)_{\tilde{A}} \quad \forall v \in V,$$

then for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$\kappa(\tilde{B}A(\omega)) \leq \kappa(\tilde{B}\tilde{A})\frac{c_1(\omega)}{c_0(\omega)}.$$

Proof. Since $\tilde{B}\tilde{A}$ is SPD with respect to $(\cdot,\cdot)_{\tilde{A}}$, it holds that

$$\kappa(\tilde{B}\tilde{A}) = |\lambda_{\max}(\tilde{B}\tilde{A})/\lambda_{\min}(\tilde{B}\tilde{A})|.$$

According to [40, Lemma 2.1] this implies

$$|\lambda_{\max}(\tilde{B}\tilde{A})|^{-1}(v,v)_{\tilde{A}} \leq (\tilde{B}^{-1}v,v) \leq |\lambda_{\max}(\tilde{B}\tilde{A})|^{-1}(v,v)_{\tilde{A}}.$$

Then, the assumption of the lemma implies that

$$c_1^{-1}|\lambda_{\max}(\tilde{B}\tilde{A})|^{-1}(Av,v) \leq (\tilde{B}^{-1}v,v) \leq c_0^{-1}|\lambda_{\min}(\tilde{B}\tilde{A})|^{-1}(Av,v), \tag{6}$$

which implies the claim with another application of [40, Lemma 2.1]. □

For $J \in \mathbb{N}$, let us assume a decomposition of V in subspaces $(V_j : j = 1,\dots,J)$, i.e., it holds that $V_j \subset V$, $j = 1,\dots,J$, and

$$V = \sum_{j=1}^{J} V_j. \tag{7}$$

We define the orthogonal projections $Q_j, P_j : V \to V_j$ for every $v \in V$ by

$$(Q_j v, w_j) := (v, w_j), \quad (AP_j v, w_j) := (Av, w_j) \quad \forall w_j \in V_j$$

and the operator $A_j : V_j \to V_j$ for every $v \in V_j$ by

$$(A_j v, w_j) := (Av, w_j) \quad \forall w_j \in V_j,$$

$j = 1, \dots, J$. Consequently it holds for every $j = 1, \dots, J$ that $A_j P_j = Q_j A$, which implies that if u is the random solution of (3), then $u_j := P_j u$ satifies for $\mathbb{P}$-a.e. ω

$$A_j(\omega) u_j(\omega) = f_j,$$

where $f_j := Q_j f$, $j = 1, \dots, J$. Note that A_j, P_j are random whereas Q_j is deterministic, $j = 1, \dots, J$. Let $R_j : V_j \to V_j$, $j = 1, \dots, J$, be random SPD operators with respect to $(\cdot, \cdot)$, which shall *approximate the inverse* of A_j respectively. Thus,

$$B^a := \sum_{j=1}^{J} R_j Q_j \tag{8}$$

is also SPD, cf. [40, Lemma 3.1], and a candidate for a random preconditioner of A, where the corresponding method is known as *parallel subspace correction*.

Algorithm 1 Apply the PCG method to (4) with the random preconditioner B^a defined in (8).

A multiplicative algorithm, cf. [40, Algorithm 3.3], can also be defined.

Algorithm 2 Let U^0 be given and U^n be already obtained, then U^{n+1} is defined by

$$U^{n+j/J} := U^{n+(j-1)/J} + R_j Q_j (f - A U^{n+(j-1)/J}), \quad j = 1, \dots, J.$$

We introduce $T_j := R_j A_j P_j = R_j Q_j A$ for every $j = 1, \dots, J$. It is well known, cf. [40, Equations (3.7) and (3.8)], that for every $n \in \mathbb{N}$ the error that results from one step of this scheme can be written for $\mathbb{P}$-a.e. ω as

$$u(\omega) - U^n(\omega) = E_J(\omega)(u(\omega) - U^{n-1}(\omega)),$$

where the random residual operator is given by

$$E_J = (\mathrm{Id} - T_J) \cdots (\mathrm{Id} - T_1).$$

For a symmetrized version, we refer to [40, Algorithm 3.4], where the respective residual operator E_J^s satisfies that $E_J^s = E_J^* E_J$, where E_J^* denotes the adjoint of E_J with respect to $(\cdot, \cdot)_A$. We observe that $\|E_J^s\|_A = \|E_J\|_A^2$, which implies that convergence for either of the algorithms implies convergence respectively for the other as well. Note that the symmetrized version that results from E_J^s can also be brought into the form of (5), where for $f \in V$

$$Bf := U^1,$$

where U^1 is defined to be the result of the symmetrized version of Algorithm 2, cf. [40, Algorithm 3.4] with $U^0 = 0$. This random symmetric operator can also be used as a preconditioner and will be denoted by B^s.

A multilevel iteration can be defined under the assumption that there are nested subspaces satisfying

$$M_1 \subset M_2 \subset \cdots \subset M_J = V, \tag{9}$$

where we also define operators $\hat{Q}_j, \hat{P}_j : M_J \to M_j$ and $\hat{A}_j : M_j \to M_j$ for $j = 1, \ldots, J$ respectively. The random multilevel iterations $\hat{B}^s_j : M_j \to M_j$, $j = 1, \ldots, J$ with parameters $m, k \in \mathbb{N}$ will be defined iteratively.

Definition 1 Let $m, k \in \mathbb{N}$ and set $\hat{B}^s_1 := \hat{A}_1^{-1}$ and assume that $\hat{B}^s_{j-1} : M_{j-1} \to M_{j-1}$ is already defined, then set $v_0 := 0$, $w_0 := 0$ for $v_0, w_0 \in M_j$ and define for every $g \in M_j$ the multigrid iteration $\hat{B}^s_j g$ by:

1. $v_\ell := v_{\ell-1} + \hat{R}_j(g - \hat{A}_j v_{\ell-1})$, for $\ell = 1, \ldots, m$
2. $v_{m+1} := v_m + w_k$, where for $i = 1, \ldots, k$

$$w_i := w_{i-1} + \hat{B}^s_{j-1}[\hat{Q}_{j-1}(g - \hat{A}_j w_{i-1}) - \hat{A}_{j-1} w_{i-1}]$$

3. $\hat{B}^s_j g := v^{2m+1}$, where v^ℓ is given by step 1 for $\ell = m+2, \ldots, 2m+1$.

Algorithm 3 Let U^0 be given, then U^n is defined by the linear iteration in (5) with $B = \hat{B}^s_J$.

Note that for $k = 1$, we have the V-cycle and for $k = 2$ the W-cycle iteration. We will consider multiple and single smoothing versions, i.e., $m \geq 2$ and $m = 1$. According to [8, Equation (2.14)] the residual operator for $m = k = 1$ is given by

$$\hat{E}^s_J = (\mathrm{Id} - \hat{B}^s_J \hat{A}_J) = (\mathrm{Id} - \hat{T}_J) \cdots (\mathrm{Id} - \hat{T}_1)(\mathrm{Id} - \hat{T}^*_1) \cdots (\mathrm{Id} - \hat{T}^*_J),$$

where $\hat{T}_j := \hat{R}_j \hat{A}_j \hat{P}_j$. There is also a non symmetric version of this algorithm, where either step 1 or step 3 are omitted. For $m = k = 1$, the residual operator for this algorithm in the case of omitting step 1 is according to [40, Equation (3.11)] given by

$$\hat{E}_J = (\mathrm{Id} - \hat{B}_J \hat{A}_J) = (\mathrm{Id} - \hat{T}_J) \cdots (\mathrm{Id} - \hat{T}_1),$$

which implies $\hat{E}^s_J = \hat{E}_J \hat{E}^*_J$. Again, it holds that $\|\hat{E}^s_J\|_A = \|\hat{E}_J\|^2_A$. The respective operators will be denoted by $\hat{B}_j$, $j = 1, \ldots, J$.

Algorithm 4 Let U^0 be given, then U^n is defined by the linear iteration in (5) with $B = \hat{B}_J$.

The case where step 3 is omitted is similar and is for example given in [9, Section 2].

For $m = k = 1$, Algorithms 2 and 4 are related, cf. [40] and will be made precise in the following proposition as a version of [40, Propositions 3.2 and 3.3].

Proposition 2.2 *In the framework of Algorithm 4 resulting from given nested subspaces* (9)*, Algorithm 2 results with $V_j = M_j$ and $R_j = \hat{R}_j$, $j = 1, \ldots, J$.*

In the framework of Algorithm 2 resulting from given nested subspaces (7)*, Algorithm 4 results with $\hat{M}_j := \sum_{i=1}^j V_i$ and $\hat{R}_j = R_j Q_j$, $j = 1, \ldots, J$.*

For the convergence of Algorithm 1 we have to prove bounds of $\kappa(BA)$, whereas for Algorithms 2 and 3 we have to show that $\|E_J\|_A \leq \delta$ for some random variable δ taking values in $(0,1)$ $\mathbb{P}$-a.s. According to Proposition 2.2, for $m = k = 1$ it is sufficient to prove the contraction property for the setting of Algorithm 2. Note that also the convergence of the symmetrized versions of the algorithms will be implied. For a brief overview how the presented algorithms are intented to be applied in the ensuing analysis, the additive preconditioner in Algorithm 1 will in later applications be the BPX or additive Schwarz preconditioner. Algorithm 2 can be used as a multiplicative domain decomposition method and Algorithms 3 and 4 refer to non-symmetric and symmetric multigrid.

In [40], assumptions are introduced involving two parameters K_0 and K_1 that allow to discuss convergence in this abstact setting. Here K_0 and K_1 are positive random variables. We recall the two conditions [40, Equations (4.2) and (4.3)].

Assumption 1 There exist positive random variables K_0 and K_1 such that

1. for every $v \in V$, there exists a decomposition $v = \sum_{j=1}^{J} v_j$ with $v_j \in V_j$, $j = 1, \ldots, J$, such that for $\mathbb{P}$-a.e. ω
$$\sum_{j=1}^{J} (R_j^{-1}(\omega) v_j, v_j) \leq K_0(\omega)(A(\omega)v, v) \tag{10}$$

2. for $\mathbb{P}$-a.e. ω and for every $S \subset \{1, \ldots, J\} \times \{1, \ldots, J\}$ and $v_j, w_j \in V$, $j = 1, \ldots, J$,
$$\begin{aligned} &\sum_{(i,j)\in S} (T_i(\omega)v_i, T_j(\omega)w_j)_{A(\omega)} \\ &\qquad \leq K_1(\omega) \left(\sum_{i=1}^{J} (T_i(\omega)v_i, v_i)_{A(\omega)} \right)^{1/2} \left(\sum_{j=1}^{J} (T_j(\omega)w_j, w_j)_{A(\omega)} \right)^{1/2}. \end{aligned} \tag{11}$$

Theorem 2.3 *Let Assumption 1 be satisfied. Let B^a be the random preconditioner given by (8). Then, for $\mathbb{P}$-a.e. ω*
$$\kappa(B^a(\omega)A(\omega)) \leq K_0(\omega)K_1(\omega). \tag{12}$$
The residual operator E_J from Algorithm 2 satisfies for $\mathbb{P}$-a.e. ω
$$\|E_J(\omega)\|_{A(\omega)}^2 \leq 1 - \frac{2-\nu}{K_0(\omega)(1+K_1(\omega))^2}, \tag{13}$$
where $\nu \geq \max_{j=1,\ldots,J}\{\rho(R_j(\omega)A_j(\omega))\}$ is deterministic.

Proof. These are explicitly [40, Theorems 4.1 and 4.4]. □

Remark 2.4 *Note that the convergence of Algorithms 3 and 4 also holds with the same random contraction number due to Proposition 2.2. The convergence of Algorithms 3 and 4 in the general cases $m, k \in \mathbb{N}$ also holds with this random contraction number respectively with its square root, cf. [9, Equations (2.12) and (2.13)]. But in the setting of Assumption 1 and Theorem 2.3 multiple smoothing or correction steps seem not to improve the contraction number.*

The random parameters K_0 and K_1 can be estimated in some cases with [40, Lemmas 4.5, 4.6 and 4.7]. Let us state a specific case of [40, Lemma 4.6] as the following lemma.

Lemma 2.5 *Let K_3 be a positive random variable that is independent of J and let $\gamma \in (0,1)$ be deterministic. If for $\mathbb{P}$-a.e. ω and for every $v, w \in V$ and every $i, j \in \{1, \dots, J\}$ it holds that*

$$(T_i(\omega)v, T_j(\omega)w)_{A(\omega)} \leq K_3(\omega)\nu\gamma^{|i-j|}(T_i(\omega)v, v)^{1/2}_{A(\omega)}(T_j(\omega)w, w)^{1/2}_{A(\omega)}, \tag{14}$$

then

$$K_1(\omega) \leq K_3(\omega)\nu\frac{2}{1-\gamma}.$$

Note that (14) is often called *strengthened Cauchy–Schwarz inequality.*

The other approach to prove convergence of multigrid methods as Algorithms 3 and 4 takes advantage of multiple smoothing steps, i.e., $m \in \mathbb{N}$, and relies on the so called *regularity and approximation* assumption, which seems to have been introduced in this form by Bramble and Pasciak in [6].

Assumption 2 Let $s \in (0,1]$. There exist positive random variables C_R and C_s such that:

1. for every $j = 1, \dots, J$ and for $\mathbb{P}$-a.e. ω

$$\frac{\|v\|^2}{\lambda_{\max}(\hat{A}_j(\omega))} \leq C_R(\omega)(\hat{R}_j(\omega)v, v) \quad \forall v \in V_j.$$

2. for every $j = 1, \dots, J$ and for $\mathbb{P}$-a.e. ω

$$((\mathrm{Id} - \hat{P}_{j-1}(\omega))v, v)_{A(\omega)} \leq C_s(\omega)\left(\frac{\|\hat{A}_j(\omega)v\|^2}{\lambda_{\max}(\hat{A}_j(\omega))}\right)^s (v, v)^{1-s}_{A(\omega)} \quad \forall v \in M_j.$$

The convergence estimate for Algorithm 3 for general $m \in \mathbb{N}$ and $s \in (0,1)$ was proven in [6], whereas the result for $s = 1$ has been known earlier, cf. [25]. The following theorem is a version of [6, Theorem 3, Theorem 1, Equation (3.28)]. There, increasing the number of smoothing steps m improves the contraction number.

Theorem 2.6 *Let Assumption 2 be satisfied. For $k = 2$, $m \in \mathbb{N}$, and the random variable $C_W := 2^{(1-s)/s}s(1-s)^{(1-s)/s}C_s^{2/s}C_R$ it holds for $\mathbb{P}$-a.e. ω*

$$\|\mathrm{Id} - \hat{B}^s_J(\omega)A(\omega)\|_{A(\omega)} \leq \frac{1}{(1 + m/C_W(\omega))^s}.$$

For $s = 1$, $k = 1$, $m \in \mathbb{N}$, and the random variable $C_V := C_s^2C_R/2$ it holds for $\mathbb{P}$-a.e. ω

$$\|\mathrm{Id} - \hat{B}^s_J(\omega)A(\omega)\|_{A(\omega)} \leq \frac{1}{1 + m/C_V(\omega)}.$$

The improvement of the contraction number with increased smoothing steps may be interesting in some application. For convergence estimates for Algorithm 4 for general $m \in \mathbb{N}$ we refer to [6, Theorem 4]. There the contraction number is the square root of the one given in the previous theorem. For the case of one correction step, i.e., $k = 1$, only non uniform convergence estimates seem to be available under Assumption 2, cf. [6, Theorems 1 and 2].

Proposition 2.7 *Let $B \in \{B^s, \hat{B}^s_J\}$ and let δ be a random variable taking values in $(0,1)$. If for $\mathbb{P}$-a.e. ω, $\|\mathrm{Id} - B(\omega)A(\omega)\|_{A(\omega)} \leq \delta(\omega)$, then $\kappa(B(\omega)A(\omega)) \leq 1/(1-\delta(\omega))$.*

Proof. Since the operator $\mathrm{Id} - BA$ can be written as E^*E for some appropriate E that was discussed above, it holds that $((\mathrm{Id} - BA)v, v)_A = \|E^*v\|_A^2 \geq 0$ for every $v \in V$. Hence,

$$(BAv, v)_A \leq (v, v)_A \quad \forall v \in V. \tag{15}$$

The assumption implies that

$$(1 - \delta)(v, v)_A \leq (BAv, v)_A \quad \forall v \in V, \tag{16}$$

which then implies the assertion. □

Remark 2.8 *The operators B^s and $\hat{B}_J^s$ can also be used as preconditioners in a PCG method to accelerate convergence. The respective condition number can be bounded by Proposition 2.7 in combination with Theorem 2.3 or Theorem 2.6.*

3 Strong convergence of iterative methods

We recall the possibly infinite dimensional Hilbert space $\mathcal{V}$ and let $(\mathcal{H}, (\cdot, \cdot))$ be another Hilbert spaces such that the embedding $\mathcal{H} \subset \mathcal{V}$ is continuous. Let $(\mathcal{V}_\ell : \ell \in \mathbb{N})$ be a nested sequence of finite dimensional subspaces of $\mathcal{V}$, i.e., $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset \mathcal{V}$. Let the finite dimensional spaces $\mathcal{V}_\ell$ have dimensions $N_\ell := \dim(\mathcal{V}_\ell)$, $\ell \in \mathbb{N}$. Similar to Section 2, we introduce random operators $\mathcal{A}_\ell$, $\mathcal{P}_\ell$, R_ℓ, T_ℓ, and Q_ℓ with respect to the inner product $(\cdot, \cdot)$ of $\mathcal{H}$, $\ell \in \mathbb{N}$. The "usual inner product" $(\cdot, \cdot)_{\mathbb{E}(\mathcal{A})}$ on $\mathcal{V}$ is given by

$$(v, w)_{\mathbb{E}(\mathcal{A})} := {}_{\mathcal{V}^*}\langle \mathbb{E}(\mathcal{A})v, w\rangle_{\mathcal{V}} \quad \forall v, w \in \mathcal{V}.$$

The inner product $(\cdot, \cdot)_{\mathcal{A}}$ with respect to the random symmetric operator $\mathcal{A}$ will also be considered.

For every $\ell \in \mathbb{N}$, we consider the variational form of (1) on the subspace $\mathcal{V}_\ell$. For every $\ell \in \mathbb{N}$, this gives rise to the random linear equation: find $u_\ell : \Omega \to \mathcal{V}_\ell$ such that for $\mathbb{P}$-a.e. ω

$${}_{\mathcal{V}^*}\langle \mathcal{A}_\ell(\omega)u_\ell(\omega), v_\ell\rangle_{\mathcal{V}} = {}_{\mathcal{V}^*}\langle f, v_\ell\rangle_{\mathcal{V}} \quad \forall v_\ell \in \mathcal{V}_\ell.$$

This is uniquely solvable by the Lax–Milgram lemma.

We assume that the random solution $u : \Omega \to \mathcal{V}$ of the problem in (1) is approximated by the Galerkin approximations u_ℓ, $\ell \in \mathbb{N}$. Specifically, we assume that there exists a positive random variable $\mathcal{C}$ that may depend on u, a scalar $s > 0$, and a decreasing sequence $(h_\ell : \ell \in \mathbb{N})$ such that for every $\ell \in \mathbb{N}$ and for $\mathbb{P}$-a.e. ω

$$\|u(\omega) - u_\ell(\omega)\|_{\mathcal{A}(\omega)} \leq \mathcal{C}(\omega)h_\ell^s. \tag{17}$$

We apply an iterative method such as Algorithm 1, 2, 3, or 4 with random contraction number δ, that is independent of ℓ, with n iterations and solve exactly on level 1, i.e., starting with $U_{\ell-1}^n$ as initial guess for level ℓ we carry out n iterations of the algorithm with a random contraction number δ that takes values in $(0, 1)$. Hence, we obtain a sequence U_ℓ^n, $\ell \in \mathbb{N}$, where we set $U_1^n := u_1$. This multilevel process was used in [2, Section 3] to derive optimal complexity bounds for the solution of certain linear equations and is also commonly referred to as "full multigrid".

Lemma 3.1 *Let us assume that* $(u_\ell : \ell \in \mathbb{N})$ *satisfies* (17) *for some* $s > 0$ *and let the sequence* $(h_\ell, \ell \in \mathbb{N})$ *satisfy that* $h_\ell = \gamma^\ell h_0$, $\ell \in \mathbb{N}$, *for some fixed* $h_0 > 0$ *and* $\gamma \in (0,1)$. *Then, for every* $\ell \in \mathbb{N}$, U_ℓ^n *and for* $\mathbb{P}$*-a.e.* ω

$$\|u(\omega) - U_\ell^n(\omega)\|_{\mathcal{A}(\omega)} \leq 2\mathcal{C}(\omega) \left(\sum_{k=0}^{\ell-1} h_{\ell-k}^s \delta(\omega)^{nk} \right) = 2\mathcal{C}(\omega) \left(\sum_{k=0}^{\ell-1} (\gamma^{-s}\delta(\omega)^n)^k \right) h_\ell^s.$$

Proof. The argument is for example given in [5, Chapter 10]. We recall that on every level $\ell \geq 2$ we will carry out n iterations of our contractive algorithm with initial guess $U_{\ell-1}^n$. Thus,

$$\|u_\ell - U_\ell^n\|_{\mathcal{A}} \leq \delta^n \|u_\ell - U_{\ell-1}^n\|_{\mathcal{A}} \quad \ell = 2, \dots, J.$$

A twofold application of the triangle inequality implies with (17) a recursion, i.e.,

$$\begin{aligned} \|u - U_\ell^n\|_{\mathcal{A}} &\leq \|u - u_\ell\|_{\mathcal{A}} + \|u_\ell - U_\ell^n\|_{\mathcal{A}} \\ &\leq \|u - u_\ell\|_{\mathcal{A}} + \delta^n \|u_\ell - U_{\ell-1}^n\|_{\mathcal{A}} \\ &\leq \|u - u_\ell\|_{\mathcal{A}}(1 + \delta^n) + \delta^n(\|u - U_{\ell-1}^n\|_{\mathcal{A}}) \\ &\leq \mathcal{C} h_\ell^s (1 + \delta^n) + \delta^n \|u - U_{\ell-1}^n\|_{\mathcal{A}}. \end{aligned}$$

Since we have set $U_1^n := u_1$, this resursion terminates and can be expanded as claimed using that $\max\{\mathcal{C}(1+\delta^n), \mathcal{C}\} \leq 2\mathcal{C}$, where the second claim follows with the assumption that $h_\ell = \gamma^\ell h_0$, $\ell \in \mathbb{N}$. □

For any Banach space $(B, \|\cdot\|_B)$ and any $p \in [1, +\infty)$, let us denote the space of strongly measurable mappings $X : \Omega \to B$ such that $\|X\|_B$ is integrable with respect to the probability measure $\mathbb{P}$ by $L^p(\Omega; B)$. For $B = \mathbb{R}$, we simply write $L^p(\Omega)$.

Theorem 3.2 *Let the assumptions of Lemma 3.1 be satisfied. Let us assume that* $(\mathcal{C}/\sqrt{\breve{a}}) \in L^p(\Omega)$ *for some* $p \in [1, +\infty)$ *and that* $1/(1-\delta) \in L^{p'}(\Omega)$ *for some* $p' \in [1, +\infty)$. *For every* $\eta \in (0,1)$, *every deterministic number of iterations* $n \in \mathbb{N}$, $q \in [1, p]$, *and* $r := p'(p-q)/(pq)$ *it holds that for every* $\ell \geq 2$

$$\|u - U_\ell^n\|_{L^q(\Omega;\mathcal{V})} \leq \mathcal{C}_r \left(h_\ell^s + n^{-r} \right),$$

where

$$\mathcal{C}_r := 2 \max \left\{ \frac{\|\mathcal{C}/\sqrt{\breve{a}}\|_{L^q(\Omega)}}{1-\eta}, \|\mathcal{C}/\sqrt{\breve{a}}\|_{L^p(\Omega)} \frac{h_1^s}{1-\gamma^s} \left\| \frac{1}{1-\delta} \right\|_{L^{p'}(\Omega)}^r \log\left(\frac{1}{\eta\gamma^s} \right)^r \right\}.$$

Proof. The idea of the proof is to decompose the probability space into $\Omega = \Omega_n \cup (\Omega_n)^{\mathsf{c}}$, where

$$\Omega_n := \{\omega \in \Omega : \delta(\omega)^n < \eta\gamma^s\}$$

and therefore $(\Omega_n)^{\mathsf{c}} = \{\omega \in \Omega : \delta(\omega)^n \geq \eta\gamma^s\}$. Note that both sets are measurable. For notational convenience, we omit ω in the following when discussing subsets of Ω. Our goal is to show, how the probability of $(\Omega_n)^{\mathsf{c}}$ tends to zero for increasing values of $n \in \mathbb{N}$, to be able to justify the applicability of a well known argument on the sets Ω_n, $n \in \mathbb{N}$. Naturally, $\Omega_{n_1} \subset \Omega_{n_2}$ for every choice of natural numbers $n_1 \leq n_2$. We recall a version of the Markov inequality, cf. [4,

Equation (2.1)], i.e., for a random variable X taking values in $(1,+\infty)$ and a non-decreasing function ϕ such that $\phi(t)>0$ it holds that

$$\mathbb{P}(X\geq t)\leq \mathbb{P}(\phi(X)\geq \phi(t))\leq \frac{\mathbb{E}(\phi(X))}{\phi(t)},\quad t\in(1,+\infty).$$

We select the non decreasing positive function $\phi(t):=t^{p'}$. Then, for $X=1/(1-\delta)$ and every $t\in(1,+\infty)$ it holds that

$$\mathbb{P}\left(\frac{1}{1-\delta}\geq t\right)\leq \mathbb{E}\left(\left(\frac{1}{1-\delta}\right)^{p'}\right)\frac{1}{t^{p'}}. \tag{18}$$

Since for every $r\in(0,1)$, $\{\delta\geq r\}=\{1/(1-\delta)\geq 1/(1-r)\}$, we conclude that

$$\mathbb{P}((\Omega_n)^{\mathsf{c}})=\mathbb{P}\left(\delta^n\geq \eta\gamma^s\right)=\mathbb{P}\left(\delta\geq(\eta\gamma^s)^{1/n}\right)=\mathbb{P}\left(\frac{1}{1-\delta}\geq\frac{1}{1-(\eta\gamma^s)^{1/n}}\right).$$

We observe that the function $x\mapsto(1-(\eta\gamma^s)^{1/x})x$ from $(1,+\infty)$ to $(0,+\infty)$ is increasing. Since the well known rule of L'Hospital implies that $\lim_{x\to+\infty}(1-(\eta\gamma^s)^{1/x})x=\log(1/(\eta\gamma^s))$, we conclude that for every $n\in\mathbb{N}$

$$1-(\eta\gamma^s)^{1/n}\leq\log\left(\frac{1}{\eta\gamma^s}\right)\frac{1}{n}.$$

For every $n\in\mathbb{N}$, we choose $t:=1/(1-(\eta\gamma^s)^{1/n})$ in (18), and conclude that for every $n\in\mathbb{N}$ it holds that

$$\mathbb{P}((\Omega_n)^{\mathsf{c}})\leq\mathbb{E}\left(\left(\frac{1}{1-\delta}\right)^{p'}\right)\log\left(\frac{1}{\eta\gamma^s}\right)^{p'}\left(\frac{1}{n}\right)^{p'}. \tag{19}$$

Hence, we have established estimates for the probability of the sets $(\Omega_n)^{\mathsf{c}}$, $n\in\mathbb{N}$. We apply Lemma 3.1, (2), and decompose the probability space into $\Omega=\Omega_n\cup(\Omega_n)^{\mathsf{c}}$ to obtain that for q as in the statement of the theorem

$$\begin{aligned}\mathbb{E}(\|u-U_\ell^n\|_{\mathbb{E}(\mathcal{A})}^q)&\leq\mathbb{E}\left(\left(2\frac{\mathcal{C}}{\sqrt{\check{a}}}\sum_{k=0}^{\ell-1}h_{\ell-k}^s\delta^{nk}\right)^q\right)\\&=\mathbb{E}\left(\left(2\frac{\mathcal{C}}{\sqrt{\check{a}}}\sum_{k=0}^{\ell-1}(\gamma^{-s}\delta^n)^k\right)^q\mathbb{1}_{\Omega_n}\right)h_\ell^{sp}\\&\quad+\mathbb{E}\left(\left(2\frac{\mathcal{C}}{\sqrt{\check{a}}}\sum_{k=0}^{\ell-1}h_{\ell-k}^s\delta^{nk}\right)^q\mathbb{1}_{(\Omega_n)^{\mathsf{c}}}\right).\end{aligned}$$

Since on Ω_n holds that $\gamma^{-s}\delta^n<\eta$, we obtain with a geometric series argument that

$$\mathbb{E}\left(\left(2\frac{\mathcal{C}}{\sqrt{\check{a}}}\sum_{k=0}^{\ell-1}(\gamma^{-s}\delta^n)^k\right)^q\mathbb{1}_{\Omega_n}\right)h_\ell^{sq}\leq 2^q\,\mathbb{E}\left(\left(\frac{\mathcal{C}}{\sqrt{\check{a}}}\right)^q\right)\left(\frac{1}{1-\eta}\right)^q h_\ell^{sq}.$$

The relation $h_\ell=\gamma^\ell h_0$, $\ell\in\mathbb{N}$, implies that for every $\ell\geq 2$ it holds that $\sum_{k=0}^{\ell-1}h_{\ell-k}^s\leq\sum_{\ell\geq1}\gamma^{\ell s}h_0=h_1^s/(1-\gamma^s)$. The Hölder inequality with $r_1=p/q$ and $r_2=p/(p-q)$ implies with

the tail bound of δ^n in (19) that

$$\begin{aligned}
&\mathbb{E}\left(\left(2\frac{\mathcal{C}}{\sqrt{\check{a}}}\sum_{k=0}^{\ell-1}h_{\ell-k}^s\delta^{nk}\right)^q \mathbb{1}_{(\Omega_n)^c}\right)\\
&\quad\le 2^q\left\|\frac{\mathcal{C}}{\sqrt{\check{a}}}\right\|_{L^p(\Omega)}^q\left(\frac{h_1^s}{1-\gamma^s}\right)^q\mathbb{E}\left(\mathbb{1}_{(\Omega_n)^c}\right)^{1/r_2}\\
&\quad\le 2^q\left\|\frac{\mathcal{C}}{\sqrt{\check{a}}}\right\|_{L^p(\Omega)}^q\left(\frac{h_1^s}{1-\gamma^s}\right)^q\left\|\frac{1}{1-\delta}\right\|_{L^{p'}(\Omega)}^{rq}\log\left(\frac{1}{\eta\gamma^s}\right)^{rq}\left(\frac{1}{n}\right)^{rq},
\end{aligned}$$

which finishes the proof of the theorem. □

Remark 3.3 *Let X be a random variable with values in $(0,1)$ such that $1/(1-X)\in L^p(\Omega)$ for some $p\in[1,+\infty)$, then (19) implies that for every $\eta<1$ and $n\in\mathbb{N}$*

$$\mathbb{P}\left(X^n\ge\eta\right)\le\left\|\frac{1}{1-X}\right\|_{L^p(\Omega)}^p\log\left(\eta^{-1}\right)^p n^{-p}.$$

The case $p'=+\infty$ in Theorem 3.2 is trivial, since then the random contraction number may be upper bounded to be uniformly strictly less than one, i.e., if $p'=+\infty$ in Theorem 3.2, then $\sup_{\omega\in\Omega}\delta(\omega)<1$. In this case, the standard theory applies with $\overline{\delta}=\sup_{\omega\in\Omega}\delta(\omega)<1$.

The convergence estimate (17) may not be available. Nevertheless, the approximated solution by an iterative scheme still converges strongly to the discrete solution on fixed levels. Let U_ℓ^n be the result of n steps of an iterative scheme with random contraction number δ_ℓ taking values in $(0,1)$ to approximate u_ℓ with initial guess $0\in\mathcal{V}_\ell$, $\ell\in\mathbb{N}$, i.e., for $\mathbb{P}$-a.e. ω

$$\|u_\ell(\omega)-U_\ell^n(\omega)\|_{\mathcal{A}(\omega)}\le\delta_\ell(\omega)^n\|u_\ell(\omega)\|_{\mathcal{A}(\omega)}.$$

Note that we allow here the random contraction number to depend on ℓ. The following result is obtained as a corollary to the proof of Theorem 3.2.

Corollary 3.4 *Let us assume that $[\omega\mapsto\|u_\ell(\omega)\|_{\mathbb{E}(\mathcal{A})}/\sqrt{\check{a}(\omega)}]\in L^p(\Omega;V)$ for some $p\in[1,+\infty)$ and that $1/(1-\delta_\ell)\in L^{p'}(\Omega)$ for some $p'\in[1,+\infty)$. For every $\eta\in(0,1)$, every deterministic number of iterations $n\in\mathbb{N}$, $q\in[1,p]$, and $r:=p'(p-q)/(pq)$ it holds that for every $\ell\ge1$*

$$\|u_\ell-U_\ell^n\|_{L^q(\Omega;\mathcal{V})}\le\mathcal{C}_r(\eta^n+n^{-r}).$$

where

$$\mathcal{C}_r:=2\max\left\{\left\|\frac{\|u_\ell\|_{\mathbb{E}(\mathcal{A})}}{\sqrt{\check{a}}}\right\|_{L^p(\Omega)},\left\|\frac{1}{1-\delta_\ell}\right\|_{L^{p'}(\Omega)}^r\log\left(\eta^{-1}\right)^r\right\}.$$

Proof. The proof is similar to the proof of Theorem 3.2 using Remark 3.3. □

4 Multigrid methods

Here, we provide sufficient conditions under which the regularity and approximation assumption and the strengthened Cauchy–Schwarz inequality can be proven with explicit dependence on the operator $\mathcal{A}$. This will allow us to show $L^q(\Omega)$ bounds of the condition numbers and tail bounds of the random contraction number in order to apply the strong error bounds from Theorem 3.2 and Corollary 3.4.

4.1 Verification of Assumption 1

We turn to a proof of the strengthened Cauchy–Schwarz inequality. We will provide a proof in the case of a random, symmetric elliptic differential operators. To be specific, let $H_0^1(D)$, $H^s(D)$, $s \in [-1,2]$, be the usual Sobolev spaces for some polytopal domain $D \subset \mathbb{R}^d$, $d \geq 1$ arbitrary, such that $H^{-s}(D)$ is the dual space of $H^s(D)$, $s \in [0,1]$, and $H^0(D) = L^2(D)$. The reader is referred to [24, Chapter 1] for details on Sobolev spaces. We consider the class of random symmetric operators

$$\mathcal{A} := -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial}{\partial x_j}\right) + a : H_0^1(D) \to (H_0^1(D))^*, \tag{20}$$

where $(a_{ij}(x))_{i,j=1,\dots,d}$ is a random symmetric matrices for a.e. $x \in D$. Let us assume that $\overline{D} = \bigcup_{k=1}^{K} \overline{D}_k$, where the subdomains D_k are pairwise disjoint with a polytopal boundary. Furthermore, we assume that the random fields a and a_{ij} are strongly measurable as mappings from Ω to $L^\infty(D)$, and $a_{ij}|_{D_k}$ is strongly measurable as a mapping from Ω to $W^{s,p}(D_k)$ such that $s > d/p$, $k = 1,\dots,K$, $i,j = 1,\dots,d$, where $W^{s,p}(D_k)$, $s \geq 0$, $p \in [1,+\infty)$, $k = 1,\dots,K$, denote the Sobolev–Slobodeckij spaces, cf. [24, Definition 1.3.2.1]. We assume that for $\mathbb{P}$-a.e. ω, $\operatorname{ess\,inf}_{x\in D}\{a(\omega,x)\} \geq 0$ and that there exists a strictly positive random variable $\check{a}$ such that for $\mathbb{P}$-a.e. ω

$$\operatorname*{ess\,inf}_{x\in D}\left\{\sum_{i,j=1}^{d} a_{ij}(\omega,x)\xi_i\xi_j\right\} \geq \check{a}(\omega)|\xi|^2 \quad \forall \xi \in \mathbb{R}^d.$$

The corresponding bilinear form $\mathfrak{a}(\cdot,\cdot)$ is given by

$$\mathfrak{a}(v,w) = \sum_{i,j=1}^{d} \int_D a_{ij}\frac{\partial v}{\partial x_j}\frac{\partial w}{\partial x_i} + avw\mathrm{d}x \quad \forall v,w \in H_0^1(D).$$

Furthermore, let $\hat{a}$ be a positive random variable such that for $\mathbb{P}$-a.e. ω

$$\check{a}(\omega)\int_D |\nabla v|^2\mathrm{d}x \leq \mathfrak{a}_\omega(v,v) \leq \hat{a}(\omega)\int_D |\nabla v|^2\mathrm{d}x \quad \forall v \in H_0^1(D). \tag{21}$$

The following example will is a class of random coefficients, which will be further discussed in Sections 6 and 7.

Example 1 The class of lognormal random coefficient fields $a_{ij} = \exp(Z)\delta_{ij}$ and $a = 0$, where $Z : \Omega \to W^{s,p}(D)$ is a strongly measurable Gaussian random field and δ_{ij} denotes the Kronecker symbol, satisfy these conditions; see ahead in Section 7 for a class of instances of such GRFs.

The proof of the strengthened Cauchy–Schwarz inequality, see (14), draws on [8, Sections 4 and 5] and [40, Section 4]. The setting in reference [8] allows for low Hölder regularity of the coefficients of elliptic operators, but does not provide a strengthened Cauchy–Schwarz inequality needed for our setting of Assumption 1. The strengthened Cauchy–Schwarz inequality proved in [40, Lemmas 6.1 and 6.3] is limited to coefficients with $W^{1,\infty}(D)$ regularity. Here, a strengthened Cauchy–Schwarz inequality will be proved with explicit dependence on the coefficients that is valid for arbitrary low Hölder regularity of the coefficients and also allows for jumps across ∂D_k (see ahead Proposition 4.2). We will identify estimates with explicit dependence on the random coefficients of $\mathcal{A}$.

Let $(\mathcal{T}_\ell, \ell \in \mathbb{N})$ be a nested sequence of shape regular simplicial, uniformly refined meshes of D, i.e., every $\tau \in \mathcal{T}_\ell$ is a finite union of elements in $\mathcal{T}_{\ell+1}$, $\ell \in \mathbb{N}$. Note that in one refinement step one simplex is refined into 2^d subsimplices. For every $k = 1, \dots, K$ and $\ell \in \mathbb{N}$, we require that $\overline{D}_k = \bigcup_{\tau \in \mathcal{T}_\ell, D_k \cap \tau \neq \emptyset} \overline{\tau}$. Let $\mathcal{V}_\ell \subset \mathcal{V}$ be the space of piecewise polynomial function with respect to the mesh $\mathcal{T}_\ell$, $\ell \in \mathbb{N}$. For simplicity, we will consider here only first order FE, i.e., polynomial degree one. Define $-\Delta_\ell : \mathcal{V}_\ell \to \mathcal{V}_\ell$ by the bilinear form $(w_\ell, v_\ell) \mapsto \int_D \nabla w_\ell \cdot \nabla v_\ell \mathrm{d}x$ over $\mathcal{V}_\ell \times \mathcal{V}_\ell$, $\ell \in \mathbb{N}$. By [8, Equation (5.1)], there exists a deterministic constant $C > 0$ such that for every $j \in \mathbb{N}$ and every $v \in \mathcal{V}_j$ with $v = \sum_{i=1}^j v_i$, $v_i \in \mathcal{V}_i$, such that

$$\sum_{i=1}^{j} \lambda_{\max}(-\Delta_i) \|v_i\|^2_{L^2(D)} \leq C \int_D |\nabla v|^2 \mathrm{d}x. \tag{22}$$

Moreover, the following inverse estimates hold: there exists a constant C such that for every $v \in \mathcal{V}_\ell$

$$\|v\|_{H^1(D)} \leq C h_\ell^{-1} \|v\|_{L^2(D)} \quad \text{and} \quad \|v\|_{H^{1+s}(D)} \leq C h_\ell^{-s} \|v\|_{H^1(D)}, \tag{23}$$

where

$$h_\ell := \max_{\tau \in \mathcal{T}_\ell} \{\mathrm{diam}(\tau)\}$$

and $s \in (0, 1/2)$, cf. [15, Theorem 3.2.6] and [11, Equation (10.1)]. These inverse estimates are sharp, which can be seen by choosing v to be a nodal basis function of the FE space $\mathcal{V}_\ell$. Since by (21) for $\mathbb{P}$-a.e. ω

$$\check{a}(\omega) \lambda_{\max}(-\Delta_\ell) \leq \lambda_{\max}(\mathcal{A}_\ell(\omega)) \leq \hat{a}(\omega) \lambda_{\max}(-\Delta_\ell), \tag{24}$$

we also observe that for $\mathbb{P}$-a.e. ω

$$\lambda_{\max}(\mathcal{A}_\ell(\omega)) \geq C \check{a}(\omega) h_\ell^{-2}, \tag{25}$$

which for $-\Delta_\ell$ is a consequence of the sharpness of (23).

We require the following assumptions on the smoothers $(R_j : j = 1, \dots, J)$: there exists a deterministic $\nu \in (0, 2)$ such that for every $j = 1, \dots, J$, and $\mathbb{P}$-a.e. ω

$$(T_j(\omega) v, T_j(\omega) v)_{\mathcal{A}(\omega)} \leq \nu (T_j(\omega) v, v)_{\mathcal{A}(\omega)} \quad \forall v \in \mathcal{V}_j. \tag{\textbf{A}}$$

There exists deterministic $c_0, c_1 > 0$ such that for every $j = 1, \dots, J$, and for $\mathbb{P}$-a.e. ω

$$c_0 \frac{\|v\|^2_{L^2(D)}}{\lambda_{\max}(\mathcal{A}_j(\omega))} \leq (R_j(\omega) v, v) \leq c_1 \frac{\|v\|^2_{L^2(D)}}{\lambda_{\max}(\mathcal{A}_j(\omega))} \quad \forall v \in \mathcal{V}_j. \tag{\textbf{B}}$$

Assume that there exists $\gamma \in (0, 1)$ such that for all $i, j \in \mathbb{N}$ satisfying $i \leq j$ it holds that

$$\frac{h_j}{h_i} \leq \gamma^{j-i}, \tag{\textbf{C}}$$

Note that (**A**) implies that $\rho(R_i A_i) \leq \nu$ for every i. There exist smoothers that satisfy these assumptions, cf. [12, Chapter 8] and [7, 10].

Lemma 4.1 *Let $s \in (0,1/2)$ and $p \in (d/s, +\infty)$, then for $\mathbb{P}$-a.e. ω, for every $\eta > 0$, $\phi \in H^1(D)$, and $\psi \in H^{1+s}(D)$ it holds that*

$$|a(\omega)(\phi,\psi)| \leq \mathcal{C}(\omega)(\eta^{-1}\|\phi\|^2_{L^2(D)} + \eta^{s/(1-s)}\|\phi\|^2_{H^1(D)})^{1/2}\|\psi\|_{H^{1+s}(D)},$$

where for a deterministic constant C independent of $(a_{ij}, a : i,j = 1,\ldots,d)$

$$\mathcal{C}(\omega) := C\left(\max_{k=1,\ldots,K}\sum_{i,j=1}^{d}\|a_{ij}(\omega)\|_{W^{s,p}(D_k)} + \|a(\omega)\|_{L^\infty(D)}\right).$$

Proof. The following argument originates from the proof of [8, Lemma 4.3]. We will track the dependence on the random elliptic coefficients $(a_{ij}, a : i,j = 1,\ldots,d)$. Let us fix $k \in \{1,\ldots,K\}$. There exists a bounded linear extension operator $I_k : H^1(D_k) \to H^1(\mathbb{R}^d)$, e.g. cf. [24, Theorem 1.4.3.1]. For every function $v : D_k \to \mathbb{R}$ the zero extension to $\mathbb{R}^d$ is denoted by $\widetilde{v}$. Let $\mathscr{F}$ denote the Fourier transform on $\mathbb{R}^d$. We obtain with Plancherel's theorem

$$\int_{D_k} a_{ij}\frac{\partial\phi}{\partial x_i}\frac{\partial\psi}{\partial x_j}\mathrm{d}x = \int_{\mathbb{R}^d}\frac{\partial(I_k\phi)}{\partial x_i}\widetilde{a_{ij}\frac{\partial\psi}{\partial x_j}}\mathrm{d}x = \left(\mathscr{F}\left(\frac{\partial(I_k\phi)}{\partial x_i}\right), \mathscr{F}\left(\widetilde{a_{ij}\frac{\partial\psi}{\partial x_j}}\right)\right)_{L^2(\mathbb{R}^d)}.$$

Recall that $\|(1+|\xi|^2)^{s/2}\mathscr{F}(v)\|_{L^2(\mathbb{R}^d)}$ is the Bessel potential norm in the Hilbert case of order s of a function v. The fact that in the Hilbert case Bessel potential and Slobodeckij spaces are equal with equivalent norms, cf. [38, Definition 2.3.1(d), Theorem 2.3.2(d), Equation 4.4.1(8)], and the boundedness of the zero extension as an operator from $H^s(D_k)$ to $H^s(\mathbb{R}^d)$, cf. [24, Corollary 1.4.4.5], imply with the Cauchy–Schwarz inequality and differentiation rules for $\mathscr{F}$ that exists a constant C such that

$$\begin{aligned}\int_{D_k} a_{ij}\frac{\partial\phi}{\partial x_i}\frac{\partial\psi}{\partial x_j}\mathrm{d}x &\leq C\left(\int_{\mathbb{R}^d}\frac{|\xi|^2}{(1+|\xi|^2)^s}\mathscr{F}(I_k\phi)\mathrm{d}\xi\right)^{1/2}\left\|a_{ij}\frac{\partial\psi}{\partial x_j}\right\|_{H^s(D_k)}\\ &\leq C(\eta^{-1}\|\phi\|^2_{L^2(D_k)} + \eta^{s/(1-s)}\|\phi\|^2_{H^1(D_k)})^{1/2}\left\|a_{ij}\frac{\partial\psi}{\partial x_j}\right\|_{H^s(D_k)},\end{aligned}$$

where the inequality $|\xi|^2/(1-|\xi|^2)^s \leq \eta^{-1} + \eta^{s/(1-s)}(1+|\xi|^2)$ is derived with elementary manipulations for every $\eta > 0$ and every $\xi \in \mathbb{R}^d$. By [24, Theorem 1.4.4.2], the multiplication of elements of $W^{s,p}(D_k)$ is a bounded linear operator on $H^s(D_k)$. Thus, by summing over k and the Cauchy–Schwarz inequality there exists a constant C such that

$$\begin{aligned}&\int_D a_{ij}\frac{\partial\phi}{\partial x_i}\frac{\partial\psi}{\partial x_j}\mathrm{d}x\\ &\quad\leq C\max_{k=1,\ldots,K}\|a_{ij}\|_{W^{s,p}(D_k)}(\eta^{-1}\|\phi\|^2_{L^2(D)} + \eta^{s/(1-s)}\|\phi\|^2_{H^1(D)})^{1/2}\|\psi\|_{H^{1+s}(D)}.\end{aligned}$$

Since it also holds that

$$\int_D a\phi\psi\mathrm{d}x \leq \|a\|_{L^\infty(D)}(\eta^{-1}\|\phi\|^2_{L^2(D)} + \eta^{s/(1-s)}\|\phi\|^2_{H^1(D)})^{1/2}\|\psi\|_{H^{1+s}(D)},$$

the claim of the lemma follows. □

Proposition 4.2 *Let Assumptions* (**A**) *and* (**B**) *be satisfied by the smoothers* $(R_j : j = 1 \dots, J)$*, let Assumption* (**C**) *hold, and let* $s \in (0, 1/2)$ *and* $p \in (d/s, +\infty)$*. Then for some deterministic constant* C *independent of* J *the inequality* (11) *from Assumption 1 holds with the random variable*

$$K_1 := C \left(\max_{k=1,\dots,K} \sum_{i,j=1}^{d} \|a_{ij}\|_{W^{s,p}(D_k)} + \|a\|_{L^\infty(D)} \right)^2 \left(\frac{1}{\check{a}} \right)^2.$$

Proof. The proof of this proposition merges ideas of the proofs of [8, Lemma 4.2] and [40, Lemma 6.3] to obtain a strengthened Cauchy–Schwarz inequality with explicit dependence on the coefficients $(a_{ij}, a : i, j = 1, \dots, d)$ in the setting considered here that allows for low spatial regularity of (a_{ij}), $i, j = 1, \dots, d$. We may assume that $j \geq i$ due to the symmetry of $(\cdot, \cdot)_{\mathcal{A}}$ and let $w \in \mathcal{V}_i$ and $\phi \in \mathcal{V}_j$ be arbitrary, which are both elements of $H^{1+s}(D)$ due to $s < 1/2$. The first inverse estimate in (23) and Lemma 4.1 imply that

$$|\mathfrak{a}(w, \phi)| \leq \mathcal{C}(\eta^{-1} + \eta^{s/(1-s)} h_j^{-2})^{1/2} \|\phi\|_{L^2(D)} \|w\|_{H^{1+s}(D)},$$

where we tacitly absorbed the deterministic constant in (23) into $\mathcal{C}$. Then, the second inverse estimate in (23), (25), and the choice $\eta := h_j^{2(1-s)}$ results in

$$\mathfrak{a}(w, \phi) \leq \mathcal{C} \sqrt{\frac{\lambda_{\max}(\mathcal{A}_j)}{\check{a}}} 2 \left(\frac{h_j}{h_i} \right)^s \|\phi\|_{L^2(D)} \|w\|_{H^1(D)},$$

where we again tacitly absorbed the deterministic constant in (25) into $\mathcal{C}$ as well as the constant in (23). Since

$$\frac{\|\mathcal{A}_j w\|_{L^2(D)}^2}{\lambda_{\max}(\mathcal{A}_j)} = \lambda_{\max}(\mathcal{A}_j)^{-1} \left(\sup_{\phi \in \mathcal{V}_j} \frac{\mathfrak{a}(w, \phi)}{\|\phi\|_{L^2(D)}} \right)^2,$$

we conclude with (21) and the assumption of the lemma that for every $w \in \mathcal{V}_i$

$$(T_j w, w)_{\mathcal{A}} \leq c_1 \frac{\|\mathcal{A}_j w\|_{L^2(D)}^2}{\lambda_{\max}(\mathcal{A}_j)} \leq c_1 4 \mathcal{C}^2 \left(\frac{1}{\check{a}} \right)^2 \left(\frac{h_j}{h_i} \right)^{2s} (w, w)_{\mathcal{A}}. \tag{26}$$

We argue in a similar fashion as in the second part of the proof of [40, Lemma 6.3], i.e., we conclude with the Cauchy–Schwarz inequality, (26), and with the scaling property of the smoothers in (**A**) that

$$\begin{aligned}
(T_j v, T_i w)_{\mathcal{A}} &\leq (T_j v, v)_{\mathcal{A}}^{1/2} (T_j T_i w, T_i w)_{\mathcal{A}}^{1/2} \\
&\leq \mathcal{C}^2 \left(\frac{1}{\check{a}} \right)^2 (T_j v, v)_{\mathcal{A}}^{1/2} \left(\frac{h_j}{h_i} \right)^{2s} (T_i w, T_i w)_{\mathcal{A}}^{1/2} \\
&\leq \mathcal{C}^2 \eta \left(\frac{1}{\check{a}} \right)^2 \left(\frac{h_j}{h_i} \right)^{2s} (T_j v, v)_{\mathcal{A}}^{1/2} (T_i w, w)_{\mathcal{A}}^{1/2},
\end{aligned}$$

where we again absorbed deterministic constants into $\mathcal{C}$. Since $h_j/h_i \leq \gamma^{j-i}$ by assumption, the statement of the proposition follows with Lemma 2.5. □

Proposition 4.3 *Let the smoothers* $(R_j : j = 1, \dots, J)$ *satisfy Assumption* **B** *with a determistic constant* c_0*. There exists a deterministic constant* C *independent of* j *such that inequality* (10) *from Assumption 1 holds with the random variable*

$$K_0 := C \frac{\hat{a}}{\check{a}}.$$

Proof. Since the assumption implies that $(R_i^{-1}v_i, v_i) \leq \lambda_{\max}(\mathcal{A}_i)/c_0\ (v_i, v_i)$, the claim of the proposition follows with (21), (22), and (24). □

4.2 Verification of Assumption 2

The proof of the regularity and approximation assumption here with explicit dependence of the constants on the random operator stems from [12, Section 9.1]. Let $(H^s, \|\cdot\|_s)$, $s \in [0,2]$, be nested Hilbert spaces, i.e., $H^s \subset H^{s'}$ for every $s > s'$, respresenting smoothness with $H^1 = \mathcal{V}$ such that the norms $\|\cdot\|_1$ and $\|\cdot\|_{\mathbb{E}(\mathcal{A})}$ are equivalent. Let $(H^{-s}, \|\cdot\|_{-s})$, $s \in (0,2]$ denote the respective dual spaces and identify H^0 with its dual $(H^0)^*$.

Proposition 4.4 *If there exists a random variable C_1 such that for $\mathbb{P}$-a.e. ω, for every $f \in H^{-1+s}$ the solution u of (1) takes values in H^{1+s} and satisfies the regularity estimate*

$$\|u(\omega)\|_{1+s} \leq C_1(\omega)\|f\|_{-1+s}$$

and if there exists a random variable C_2 such that for $\mathbb{P}$-a.e. ω, every $v \in H^{1+s}$ is approximated by its Galerkin projection, i.e.,

$$\|(\mathrm{Id} - \mathcal{P}_{j-1}(\omega))v\|_{\mathcal{A}(\omega)} \leq C_2(\omega)(\lambda_{\max}(\mathcal{A}_j(\omega)))^{-s/2}\|v\|_{1+s}$$

then, Assumption 2 holds with the random variable

$$C_s := C\hat{a}^{3-s}C_1^2C_2^2,$$

where $C > 0$ is some deterministic constant.

Proof. We follow the proof of [12, Lemma 9.1] with the specific aim to make the dependence of the constant C_s on the operator $\mathcal{A}$ explicit. Let $v \in M_j$ be arbitrary, define $g := \mathcal{A}_j^{1-s}(\mathrm{Id} - \mathcal{P}_{j-1})v$, and let w solve $\mathcal{A}w = g$ uniquely. Note that by definition of g

$$(\mathcal{A}_j^{1-s}(\mathrm{Id} - \mathcal{P}_{j-1})v, (\mathrm{Id} - \mathcal{P}_{j-1})v) = ((\mathrm{Id} - \mathcal{P}_{j-1})v, g). \tag{27}$$

We follow the argument that proves [12, Equation (9.9)] to conclude with both assumptions of the proposition and (2) that

$$\begin{aligned}((\mathrm{Id} - \mathcal{P}_{j-1})v, g) &= (\mathcal{A}w, (\mathrm{Id} - \mathcal{P}_{j-1})v)\\ &\leq \|v\|_{\mathcal{A}}\|(\mathrm{Id} - \mathcal{P}_{j-1})w\|_{\mathcal{A}}\\ &\leq C_1C_2\hat{a}^{(1+s)/2}(\lambda_{\max}(\mathcal{A}_j))^{-s/2}\|v\|_{\mathcal{A}}\|g\|_{-1+s}.\end{aligned} \tag{28}$$

By the definition of $\mathcal{A}_j$ and Q_j and (2), there exists a deterministic C such that for every $v \in H^1$

$$(\mathcal{A}_jQ_jv, Q_jv)^{1/2} \leq C\sqrt{\hat{a}}\|v\|_1.$$

Then, by interpolation, e.g. [12, Theorem A.4], it follows that

$$(\mathcal{A}_j^{1-s}Q_j\phi, Q_j\phi)^{1/2} \leq (\hat{a}C)^{1-s}\|\phi\|_{1-s} \quad \forall \phi \in H^1,$$

which thereafter implies with the Cauchy–Schwarz inequality that

$$\begin{aligned}|(g,\phi)| = |(g, Q_j\phi)| &\leq (\mathcal{A}_j^{-1+s}g, g)^{1/2}(\mathcal{A}_j^{1-s}Q_jv', Q_jv)^{1/2}\\ &\leq (\sqrt{\hat{a}}C)^{1-s}(\mathcal{A}_j^{-1+s}g, g)^{1/2}\|\phi\|_{1-s}.\end{aligned} \tag{29}$$

By duality it holds that $\|g\|_{-1+s} = \sup_{\phi\in H^1}(g,\phi)/\|\phi\|_{1-s}$, which implies with (29) and the definition of g that

$$\|g\|_{-1+s} \leq (\hat{a}C)^{1-s}(\mathcal{A}_j^{1-s}(\mathrm{Id}-\mathcal{P}_{j-1})v, (\mathrm{Id}-\mathcal{P}_{j-1})v)^{1/2}. \tag{30}$$

Combining (27), (28), and (30) we have shown that

$$(\mathcal{A}_j^{1-s}(\mathrm{Id}-\mathcal{P}_{j-1})v, (\mathrm{Id}-\mathcal{P}_{j-1})v) \leq \hat{a}^{3-s}C^{2(1-s)}C_1^2C_2^2(\lambda_{\max}(\mathcal{A}_j))^{-s}(\mathcal{A}v, v),$$

which corresponds to the assumption [12, Equation (A.2)]. This condition is sufficient to conclude the regularity and approximation assumption with the given expression for C_s by [12, Lemma 3.2]. □

Remark 4.5 *For uniformly refined FE discretizations, the assumption on the random variable C_2 can be proven by Galerkin orthogonality and an equivalence between the mesh width and the maximal eigenvalue of the stiffness matrix. In the case of the elliptic operators the regularity estimate has been derived explicitly in the coefficients of elliptic operators in Euclidean domains in [14, 37]. Respective statements for bounded, closed, smooth submanifolds in $\mathbb{R}^3$ are given in [27, Appendix C].*

5 Domain decomposition methods

We will mainly consider overlapping domain decomposition methods in the setting of Assumption 1 anf for the random elliptic operator defined in (20). We denote by $\mathcal{V}_0 \subset \mathcal{V}$ a *coarse* first order FE space with meshwidth h_0 of the type introduced in Section 4.1. The first order FE space with a *fine* grid is denoted by $\widetilde{\mathcal{V}} \subset \mathcal{V}$. For a given set of overlapping subdomains $(D_j : j = 1,\dots,J)$ of D such that $\overline{D} = \bigcup_{j=1}^J \overline{D}_j$. These subdomains result for example by extending a given disjoint set of subdomains by a multiple of h_0 in each spatial direction such that the union of its closures contains D. Also we assume that the boundary ∂D_j aligns with the considered mesh, $j = 1,\dots,J$. The FE spaces $\mathcal{V}_j$, $j = 1,\dots,J$, are subspaces of $\widetilde{\mathcal{V}}$ and are defined by

$$\mathcal{V}_j := \{v \in \widetilde{\mathcal{V}} : v(x) = 0 \ \forall x \in D\backslash D_j\}, \quad j = 1,\dots,J$$

So, we consider the redundant space decomposition

$$\widetilde{\mathcal{V}} = \sum_{j=0}^{J} \mathcal{V}_j.$$

We consider the case that symmetric multigrid solvers from Section 4.1 (Assumptions **A**, **B**, and **C** are satisfied) are used as so called *subpaces solvers* $(R_j : j = 0,\dots,J)$, which are random here. Therefore, suppose that the spaces $\mathcal{V}_j$ have nested subspaces $\mathcal{M}_{j,1} \subset \ldots \subset \mathcal{M}_{j,J'(j)} = \mathcal{V}_j$, $j = 0,\dots,J$. Naturally, only few levels are used on the subspace $\mathcal{V}_0$, i.e. $J'(0) = \mathcal{O}(1)$. As in Section 4.1, we seek for random variables K_0 and K_1 with explicit dependence on the random operator $\mathcal{A}$ in (20), in order to obtain $L^q(\Omega)$-estimates for the condition numbers using additive Schwarz preconditioners and tail bounds of the random contraction number of multiplicative domain decomposition methods.

Proposition 5.1 *There exists a deterministic constant $C > 0$ that is independent of J and $J'(j)$, $j = 0, \dots, J$, such that inequality* (10) *in Assumption 1 holds with the random variable*

$$K_0 := C \left(\frac{\hat{a}}{\check{a}}\right)^4 \frac{(\hat{a})^2}{(\check{a})^6} \left(1 + \left(\max_{k=1,\dots,K} \sum_{i,j=1}^{d} \|a_{ij}\|_{W^{s,p}(D_k)} + \|a\|_{L^\infty(D)}\right)^4\right).$$

Proof. By (21) and [40, Lemmas 4.5 and 7.1],

$$K_0 \leq \frac{\hat{a}}{\check{a}} \frac{1}{\min_{j=0,\dots,J} \lambda_{\min}(R_j \mathcal{A}_j)}.$$

Since the R_j's are chosen to be symmetric multigrid solvers, Propositions 4.3 and 4.2, Theorem 2.3, and (16) imply there esists a deterministic constant $c > 0$ such that for every $j = 0, \dots, J$,

$$\lambda_{\min}(R_j \mathcal{A}_j) \geq c \frac{(\check{a})^5}{\hat{a}} \left(1 + \left(\max_{k=1,\dots,K} \sum_{i,j=1}^{d} \|a_{ij}\|_{W^{s,p}(D_k)} + \|a\|_{L^\infty(D)}\right)^4\right)^{-1}.$$

This implies the assertion of the proposition. □

Proposition 5.2 *Inequality* (11) *holds with the deterministic number*

$$K_1 := (1 + |\{(i,j) \in \{1, \dots, J\}^2 : D_i \cap D_j\}|)$$

Proof. The assertion will follow by [40, Lemma 4.7] after we show an estimate of the form of Assumption (**A**). By (15), it holds that $\lambda_{\max}(R_j \mathcal{A}_j) \leq 1$, $j = 0, \dots, J$. Let $j = 0, \dots, J$ be arbitrary. Since R_j is a symmetric multigrid solver, $R_j \mathcal{A}_j$ is symmetric and positive definite with respect to $(\cdot, \cdot)_{\mathcal{A}}$. There exists an orthonormal basis of eigenvectors of $R_j \mathcal{A}_j$ with respect to $(\cdot, \cdot)_{\mathcal{A}}$ such that $R_j \mathcal{A}_j v_i = \lambda_i v_i$. Hence, for every $v \in \mathcal{V}_j$, $v = \sum_i (v, v_i)_{\mathcal{A}} v_i$ and

$$(R_j \mathcal{A}_j v, R_j \mathcal{A}_j v)_{\mathcal{A}} = \sum_i \lambda_i^2 ((v, v_i)_{\mathcal{A}})^2 \leq \lambda_{\max}(R_j \mathcal{A}_j) \sum_i \lambda_i ((v, v_i)_{\mathcal{A}})^2 = (R_j \mathcal{A}_j v, v)_{\mathcal{A}},$$

which implies the estimate of the proposition with [40, Lemma 4.7]. □

Remark 5.3 *Nonoverlapping domain decomposition methods can also be formulated in the general framework introduced in [40, Section 3], that is recalled in Section 2, relying on Assumption 1, cf. [42, Section 2.2]. Strong convergence of PCG with deterministic iteration numbers using preconditioners for random* interface operators *of nonoverlapping domain decompositions resulting in uniform condition numbers, cf. [42, Sections 7.1 and 7.2], can also be established with Lemma 2.1 and either Theorem 3.2 or Corollary 3.4.*

6 Application to lognormal diffusion problems

The presented theory in Sections 3, 4, and 5 is in particular applicable to lognormal diffusion problems. Let Z be a GRF on D that takes values in Hölder spaces $C^t(\overline{D})$ such that for some $t \in (0, 1]$

$$Z \in L^q(\Omega; C^t(\overline{D})) \quad \text{and} \quad \exp(Z) \in L^q(\Omega; C^0(\overline{D})) \quad \forall q \in [1, +\infty), \tag{**D**}$$

see ahead Section 7 for a class of instances of such GFRs. Since for every $v \in C^t(\overline{D})$, $t \in (0,1]$, $\|\exp(v)\|_{C^t(\overline{D})} \leq \|\exp(v)\|_{C^0(\overline{D})}(1 + \|v\|_{C^t(\overline{D})})$, the assumption in (**D**) implies by the Cauchy–Schwarz inequality

$$\exp(Z) \in L^q(\Omega; C^t(\overline{D})) \quad \forall q \in [1, +\infty)$$

For the lognomal coefficient $a := \exp(Z)$, we consider the elliptic diffusion probem with Dirichlet boundary conditions in variational form: find $u : \Omega \to \mathcal{V}$ such that for $\mathbb{P}$-a.e. ω

$$a_\omega(u(\omega), v) = \int_D a(\omega)\nabla u(\omega) \cdot \nabla v = {}_{\mathcal{V}^*}\langle f, v\rangle_{\mathcal{V}} \quad \forall v \in \mathcal{V}. \tag{31}$$

where $\mathcal{V} = H^1_0(D)$. Well-posedness and approximation by Finite Elements is well known, cf. [13, 14, 37]. We use the FE spaces $\mathcal{V}_\ell$ from Section 4.1 with maximal meshwidth h_ℓ of $\mathcal{T}_\ell$, $\ell \in \mathbb{N}$ and remark that for each ℓ, the space $\mathcal{V}_\ell$ may have the additional structure for overlapping domain decomposition methods with multigrid subspace solvers as introduced in Section 5.

Elements of $H^{1+s}(D)$, $s \in [0,1]$, can be approximated by functions in $\mathcal{V}_\ell$, cf. [15, Theorem 3.2.1], i.e., there exists a deterministic constant $C > 0$ such that for every $v \in H^{1+s}(D)$ there is $w_\ell \in \mathcal{V}_\ell$ such that

$$\|v - w_\ell\|_{\mathcal{V}} \leq C h_\ell^s \|v\|_{H^{1+s}(D)}. \tag{32}$$

Note that the approximation property stated in [15, Theorem 3.2.1] can be interpolated to also hold for non-integer order Sobolev spaces. The following regularity estimate makes the dependence on the coefficient a explicit. For every $s \in [0, \min\{t, t_{-\Delta}\})\backslash\{1/2\}$ there exists a deterministic constant $C > 0$ such that for $\mathbb{P}$-a.e. ω

$$\|u(\omega)\|_{H^{1+s}(D)} \leq C \frac{\|a(\omega)\|_{C^0(\overline{D})}\|a(\omega)\|^2_{C^t(\overline{D})}}{(\min_{x\in\overline{D}} a(\omega, x))^4} \|f\|_{H^{-1+s}(D)}, \tag{33}$$

where $t_{-\Delta}$ is the maximal value such that the inverse of the Dirichlet Laplacian satisfies $(-\Delta)^{-1} : H^{-1+t_{-\Delta}}(D) \to \mathcal{V} \cap H^{1+t_{-\Delta}}(D)$ is bounded. For $d = 2$, the estimate (33) is due to [37, Lemma 5.2] (for $d = 3$, the reader is referred to [37, Remark 5.2(c)]). The solution u can be approximated in $\mathcal{V}_\ell$ by the FE approximation denoted by u_ℓ, in the $L^q(\Omega, \mathcal{V})$-norm, $\ell \in \mathbb{N}$. Specifically, by Céa's lemma, (32), and (33), there exists a deterministic constant $C > 0$ that is independent of $\ell \in \mathbb{N}$ and a such that for $\mathbb{P}$-a.e. ω

$$\|u(\omega) - u_\ell(\omega)\|_{L^q(\Omega;\mathcal{V})} \leq C \left\| \frac{\|a(\omega)\|^2_{C^0(\overline{D})}\|a(\omega)\|^2_{C^t(\overline{D})}}{(\min_{x\in\overline{D}} a(\omega, x))^5} \right\|_{L^q(\Omega)} \|f\|_{H^{-1+s}(D)}\, h_\ell^s, \quad \ell \in \mathbb{N}.$$

Finiteness of the right hand side follows by the Cauchy–Schwarz inequality using the Assumption (**D**). Since the embedding $C^t(\overline{D}) \subset W^{s,p}(D)$ is continuous for every $0 < s < t$ and every $p \in [1, +\infty)$, the conditions from Sections 4 and 5 are satisfied. Let

$$U_\ell^n : \Omega \to \mathcal{V}_\ell \tag{34}$$

be the result of $n \in \mathbb{N}$ iterations of an iterative algorithm introduced in the previous sections, with initial guess $U_{\ell-1}^n$, $2 \leq \ell \in \mathbb{N}$, and $U_1^n = u_1$. The iterative methods are (symmetric) multigrid (see Algorithms 4 and 3) or PCG using the additive (BPX) preconditioner (see Algorithm 1) in Section 4. In the setting of Section 5, $U_\ell^n : \Omega \to \mathcal{V}_\ell$ in (34) may result from $n \in \mathbb{N}$ iterations of PCG using the additive Schwarz preconditioner (see Algorithm 1) or a multiplicative domain decomposition method (see Algorithm 2), where symmetric multigrid (see Algorithm 3) is used as subspace solvers.

Theorem 6.1 *For $0 < s < t \leq 1$ and every $q, r \in [1, +\infty)$, there exists a constant $C > 0$ such that for every number of iterations $n \in \mathbb{N}$ and $\ell \geq 2$,*

$$\|u - U_\ell^n\|_{L^q(\Omega;\mathcal{V})} \leq C(h_\ell^s + n^{-r}).$$

Proof. By the Cauchy–Schwarz inequality, $\|a\|^2_{C^0(\overline{D})}\|a\|^2_{C^t(\overline{D})}(\min_{x\in\overline{D}} a(x))^{-5} \in L^p(\Omega)$ for every $p \in [1, +\infty)$, which is one of the conditions of Theorem 3.2.

It remains to verify the needed properties of the random contraction number in the conditions of Theorem 3.2. In the framework of Assumption 1, by Theorem 2.3 the random contraction number δ satisfies for the multiplicative iteration $1/(1-\delta) \leq K_0(1+K_1)^2/(2-\nu)$. In the case of multigrid, Propositions 4.2 and 4.3 and the Cauchy–Schwarz inequality imply that $1/(1-\delta) \in L^{p'}(\Omega)$ for every $p' \in [1, +\infty)$. For overlapping domain decomposition methods, this statement is due to Propositions 5.1 and 5.2. If the additive preconditioner is applied with PCG, the random contraction number δ satisfies by Theorem 2.3, $1/(1-\delta) \leq \sqrt{K_0 K_1} + 1$. By the same argument, $1/(1-\delta) \in L^{p'}(\Omega)$ for every $p' \in [1, +\infty)$. In the setting of Assumption 2, the assumptions of Proposition 4.4 are satisfied by (32) and Galerkin orthogonality and (33). By a similar argument using Theorem 2.6, $1/(1-\delta) \in L^{p'}(\Omega)$ for every $p' \in [1, +\infty)$. Hence, the parameter r in Theorem 3.2 may be arbitrarily large, which implies the assertion. □

Corollary 6.2 *In the setting of Theorem 6.1, let U_ℓ^n result from $n \in \mathbb{N}$ iterations of PCG with a deterministic preconditioner $\widetilde{B}_j$ such that $\kappa(\widetilde{B}_j\,\mathbb{E}(\mathcal{A}_j))$ is bounded uniformly in j. Then, the strong convergence estimate of Theorem 6.1 also holds.*

Proof. By Lemma 2.1, for $\mathbb{P}$-a.e. ω, $\kappa(\widetilde{B}_j\mathcal{A}_j(\omega)) \leq \hat{a}(\omega)/\check{a}(\omega)\kappa(\widetilde{B}_j\,\mathbb{E}(\mathcal{A}_j))$. Since $\hat{a}/\check{a} \in L^{q'}(\Omega)$ for every $q' \in [1, +\infty)$, the claim follows as in the proof of Theorem 6.1. □

Corollary 6.3 *In the setting of Theorem 6.1, for every $\varepsilon > 0$ there exists a constant $C_{q,\varepsilon,s} > 0$ that is independent of h_ℓ such that for every $\ell \geq 2$*

$$\|u - U_\ell^n\|_{L^q(\Omega;\mathcal{V})} \leq C_{q,\varepsilon,s}\, h_\ell^s.$$

with a deterministic number of iterations given by $n = \lceil n_0 h_\ell^{-\varepsilon} \rceil$ for a deterministic constant $n_0 > 0$. The cost for one sample of U_ℓ^n is

$$\mathcal{O}(h_\ell^{-d-\varepsilon})$$

with deterministic *constants that are independent of h_ℓ, $\ell \geq 0$.*

Proof. The cost of one iteration is $\mathcal{O}(h_\ell^{-d})$, since the matrix vector product has cost $\mathcal{O}(h_{\ell'}^{-d})$ for the sparse stiffness matrices that result taking the nodal basis of $\mathcal{V}_{\ell'}$, $\ell' \leq \ell$. The error contributions in the estimate of Theorem 6.1 are equilibrated for this choice of iteration number, since r in Theorem 6.1 can be chosen arbitrarily large, i.e., $r = s/\varepsilon$ is admissible. □

6.1 Mean field approximation with MLMC

Let $G : \Omega \times \mathcal{V} \to \mathcal{W}$ be a random operator, where $\mathcal{W}$ is a Banach space, that is Lipschitz continuous, i.e., there exists a positive random variable $\mathcal{C}$ such that for $\mathbb{P}$-a.e. ω

$$\|G(\omega, v) - G(\omega, w)\|_{\mathcal{W}} \leq \mathcal{C}(\omega)\|v - w\|_{\mathcal{V}} \quad \forall v, w \in \mathcal{V}.$$

Let us assume that $\mathcal{C} \in L^{q'}(\Omega)$ for some $q' \in (2, +\infty)$. For the computation of the mean field $\mathbb{E}(G(u))$, MLMC methods are well established, cf. [14, 23]. For a maximum number of levels $L \in \mathbb{N}$, we define the MLMC estimator by

$$E_L^{\mathrm{ML}}(G(U_L^{n_L}) := \sum_{\ell=1}^{L} E_{M_\ell}(G(U_\ell^{n_\ell}) - G(U_{\ell-1}^{n_{\ell-1}})),$$

where $(E_{M_\ell} : \ell = 1, \dots, L)$ are Monte Carlo estimators that are mutually independent. We used the convention that $U_0^{n_0} := 0$. The MLMC estimator satisfies the general error estimate

$$\begin{aligned} &\| \mathbb{E}(G(u)) - E_L^{\mathrm{ML}}(G(U_L^{n_L}))\|_{L^2(\Omega;\mathcal{W})} \\ &\qquad \le \|G(u) - G(U_L^{n_L})\|_{L^1(\Omega;\mathcal{W})} + \left(\sum_{\ell=1}^{L} \frac{1}{M_\ell} \|G(U_\ell^{n_\ell}) - G(U_{\ell-1}^{n_{\ell-1}})\|_{L^2(\Omega;\mathcal{W})}^2 \right)^{1/2}, \end{aligned} \tag{35}$$

see for example [27, Lemma 5.1]. For the effectiveness of MLMC, bounds on the $L^2(\Omega;\mathcal{V})$-norm of the difference between two consecutive discretization levels have to be known. Specifically, by Theorem 6.1 and Corollary 6.3, the Hölder inequality, and the triangle inequality

$$\|G(U_\ell^{n_\ell}) - G(U_{\ell-1}^{n_{\ell-1}})\|_{L^2(\Omega;\mathcal{W})} \le \|\mathcal{C}\|_{L^{q'}(\Omega)} \|U_\ell^{n_\ell} - U_{\ell-1}^{n_{\ell-1}}\|_{L^{2q'/(q'-2)}(\Omega;\mathcal{V})} = \mathcal{O}(h_{\ell-1}^s),$$

where $n_\ell = \lceil n_0 h_\ell^{-\varepsilon} \rceil$ for a deterministic constant $n_0 > 0$ and a small deterministic number $\varepsilon > 0$. Inserting this estimate into (35) yields

$$\mathrm{error}_L := \| \mathbb{E}(G(u)) - E_L^{\mathrm{ML}}(G(U_L^{n_L}))\|_{L^2(\Omega;\mathcal{W})} = \mathcal{O}\left(h_L^{2s} + \sum_{\ell=1}^{L} \frac{h_{\ell-1}^{2s}}{M_\ell} \right)^{1/2}. \tag{36}$$

Let us assume that the stiffness matrix on every level $\ell \in \mathbb{N}$ can be assembled in computational work $\mathcal{O}(h_\ell^{-d} \log(h_\ell^{-1}))$. Possible techniques to achieve this are in certain cases the circulant embedding or FFT, cf. [34, Chapter 7]. See also [36] for approximation techniques of the so called *Karhunen–Loève* expansion. Also, suppose that the computational cost of the evaluation of $G(v)$ is $\mathcal{O}(h_\ell^{-d})$ for every $v \in \mathcal{V}_\ell$. For example circulant embedding or FFT are applicable if $\log(a)$, which is a Gaussian random field, has a stationary covariance kernel and is the restriction of such a Gaussian random field defined on a product domain containing D (see ahead Section 7 for a class of GRFs, where FFT is applicable). Thus, the computational work to compute $E_L^{\mathrm{ML}}(G(U_L^{n_L}))$ is

$$\mathrm{work}_L = \mathcal{O}\left(\sum_{\ell=1}^{L} M_\ell h_\ell^{-d-\varepsilon} \right). \tag{37}$$

The sample numbers $(M_\ell)_{\ell=1,\dots,L}$ are still to be chosen such that the error in (36) as a function of the work in (37) is optimized. This is a generic task, which has for this type of work model been analyzed in [28, Section 6] (see also [32, 23]). Specifically, we take the sample numbers according to [28, Equations (44) and (47)], i.e., for $M^* \in \mathbb{N}$

$$M_\ell = \lceil M_1 h_\ell^{(2s+d+\varepsilon)/2)} \rceil, \quad \ell = 2, \dots, L, \tag{38}$$

and

$$M_1 = M^* \begin{cases} \lceil 2^{sL/2} \rceil & \text{if } d + \varepsilon < 2s, \\ \lceil 2^{(2s+d+\varepsilon)L/2} \rceil & \text{if } d + \varepsilon > 2s. \end{cases} \tag{39}$$

Theorem 6.4 *Let the sample numbers be given by* (38) *and* (39)*, then an error threshold* $0 <$ TOL*, i.e.,*

$$\| \mathbb{E}(G(u) - E_L^{\mathrm{ML}}(G(U_L^{n_L}))\|_{L^2(\Omega;\mathcal{V})} = \mathcal{O}(\mathrm{TOL}),$$

can be achieved with computational cost

$$\mathrm{work}_L = \begin{cases} \mathcal{O}(\mathrm{TOL}^{-2}) & \text{if } 2s > d + \varepsilon, \\ \mathcal{O}(\mathrm{TOL}^{-(d+\varepsilon)/s}) & \text{if } 2s < d + \varepsilon, \end{cases} \tag{40}$$

where $d = 1, 2, 3$ *is the dimension of the domain* D*.*

Proof. The estimates for the error and the work in (36) and (37) are of the type considered in [28, Section 6]. The used sample numbers are those obtained in [28, Section 6], where the error as a function of the work is minimized. Hence, the complexity theorem [28, Theorem 6.2] (applied with $\tau = s$, $\bar{\chi} = 1/2$) implies the assertion. □

Note that the case $2s = d + \varepsilon$ is not included, since here s satisfies a strict inequality, which renders this case unimportant. The estimate on the computational work of MLMC for lognormal diffusion coefficients in (40) are deterministic and do not have assumptions on the complexity of the PDE solver. The computational work of MLMC FE for lognormal diffusions is for $d = 2, 3$ essentially optimal, since $\varepsilon > 0$ may be chosen arbitrarily small.

Remark 6.5 *The established theory in this paper is also applicable for deterministic preconditioners that do not imply uniform condition numbers. A possible class of examples are so called "algebraic multigrid" (AMG) preconditioners for which a multilevel convergence theory does not seem to be available, cf. [41, Section 5]. However, if such a deterministic AMG preconditioner could be tuned to* $\mathbb{E}(\mathcal{A}_j)$ *or any operator which is spectrally equivalent in the sense of* (2)*, the respective iterative method also converges strongly as in Theorem 6.1 by a similar argument applying Corollaries 3.4 and 6.2. Note that in the example* (31) *(see also Example 1), if* Z *is stationary, then* $\mathbb{E}(\mathcal{A})$ *is the Dirichlet Laplacian multiplied by* $\mathbb{E}(\exp(Z))$*, which is constant with respect to* $x \in D$*.*

7 Numerical experiments

We consider a class of centered, stationary GRFs that are solutions to the white noise stochastic PDE

$$(-\Delta + \kappa^2)^{\alpha/2} Z = W \quad \text{on } \widetilde{D}, \tag{41}$$

where $\widetilde{D}$ is a superset of the domain D, W is spatial white noise on $\widetilde{D}$ (cf. [1]), $\nu = \alpha - d/2 > 0$, and $\kappa > 0$. It is well known that for $\widetilde{D} = \mathbb{R}^d$, the GRF Z has so called Matérn covariance with smoothness parameter ν and length scale parameter $\lambda = \sqrt{\nu}/\kappa$, cf. [33].

In our numerical experiments, we will choose $d = 2$ and $D = [0, 1]^2$. A stationary GRF as a solution to (41) results by restricting Z to the domain D. It is convenient to choose $\widetilde{D} = [0, 2]^2$ with periodic boundary conditions. The solution Z to (41) can be obtained by a

spectral Galerkin method. The set

$$\begin{aligned}\Bigg\{ &\frac{1}{2}, \frac{1}{\sqrt{2}}\cos(\pi k_1 x_1), \frac{1}{\sqrt{2}}\sin(\pi k_1 x_1), \frac{1}{\sqrt{2}}\cos(\pi k_2 x_2), \frac{1}{\sqrt{2}}\sin(\pi k_2 x_2),\\ &\cos(\pi k_1 x_1)\cos(\pi k_2 x_2), \cos(\pi k_1 x_1)\sin(\pi k_2 x_2),\\ &\sin(\pi k_1 x_1)\cos(\pi k_2 x_2), \sin(\pi k_1 x_1)\sin(\pi k_2 x_2) : \qquad k_1, k_2 \in \mathbb{N}\Bigg\}\end{aligned} \tag{42}$$

is an orthonormal basis (ONB) of the square integrable, periodic functions on $\widetilde{D}$ denoted by $L^2_{\text{periodic}}(\widetilde{D})$. We observe that white noise W applied to any ONB yields a sequence of independent standard normally distributed random variables. Since the ONB (42) diagonalizes the operator $(-\Delta + \kappa^2)^{\alpha/2}$, the stationary GRF Z can be explicitly expanded with respect to this basis. The random coefficients with respect to the ONB (42) are given by

$$\mu_{k_1,k_2}(\omega) := \frac{1}{(\pi^2(k_1^2 + k_2^2) + \kappa^2)^{\alpha/2}} y_{k_1,k_2}(\omega), \quad k_1, k_2 \in \mathbb{N}_0,$$

where $(y_{k_1,k_2})_{k_1,k_2\in\mathbb{N}_0}$ is a sequence of independent, standard normally distributed random variables. For $\sigma > 0$, to be determined later, the GRF Z will be rescaled by

$$\sigma \left(\frac{1}{4\kappa^{2\alpha}} + \sum_{k\geq 1} \frac{1}{(\pi^2 k^2 + \kappa^2)^{\alpha}} + \sum_{k_1,k_2\geq 1} \frac{1}{(\pi^2(k_1^2 + k_2^2) + \kappa^2)^{\alpha}} \right)^{-1/2}$$

such that the pointwise variance satisfies

$$\mathbb{E}(Z(x)^2) = \sigma^2, \quad \forall x \in D.$$

The expansion of Z can be truncated for numerical purposes and efficiently implemented with FFT. For any $\widetilde{k} \in \mathbb{N}$, let us denote the truncation of the expansion of Z to the terms such that $k_1 \leq \widetilde{k}$ and $k_2 \leq \widetilde{k}$ by $Z^{\widetilde{k}}$. The following proposition quantifies the strong convergence of $Z^{\widetilde{k}}$ to Z in L^q-Hölder spaces, which also reveals the regularity of the GRF.

Proposition 7.1 *For every $t \in (0, \nu)$, $\varepsilon \in (0, \nu - t)$, and every $q \in [1, +\infty)$, there exists a constant $C > 0$ such that for every truncation $\widetilde{k} \in \mathbb{N}$*

$$\|Z - Z^{\widetilde{k}}\|_{L^q(\Omega; C^t(\widetilde{D}))} \leq C\widetilde{k}^{-(\nu - t - \varepsilon)}.$$

This proposition can be proven similarly to [27, Theorem 2.2], we omit the proof for convenience of the reader and a concise presentation. The additional error introduced by truncating the expansion of Z is consistent with the FE discretization error if $\widetilde{k}$ is chosen level-dependently such that $\widetilde{k}_\ell = \lceil \widetilde{k}_1 h_\ell^{-\nu} \rceil$ for some $\widetilde{k}_1 > 0$ at our disposal and FE meshwidth h_ℓ. The reader is referred to the discussion in [27, Sections 3 and 5]. As a consequence of Fernique's theorem, we may conclude similarly as in [13, Proposition 3.10]. that $\exp(Z) \in L^q(\Omega; C^0(\overline{D}))$ for every $q \in [1, +\infty)$; see also [27, Proposition B.1]. Thus, the assumptions in (**D**) are verified by the GRF Z and the developed theoretical error estimates of this paper, in particular Corollary 6.3 and Theorem 6.4, hold.

The GRF Z is taken as a random input and we seek to approximate the expectation of the solution to (31) with a MLMC estimator, where $a = \exp(Z)$ and right hand side $f(x_1, x_2) = \sin(\pi x_1)\sin(\pi x_2)$. We use the sample numbers in (38) and (39) with $M^* = 5$ and $L = 1, \dots, 8$ levels. For the level dependent truncation of the expansion of the GRF, we use $\widetilde{k}_\ell = \lceil \widetilde{k}_1 h_\ell^{-\nu} \rceil$ with $\widetilde{k}_1 = 2$. We consider a full multigrid cycle with PCG and a direct solver on the lowest level. More precisely, CG will be preconditioned by the BPX preconditioner using Jacobi smoothers. We recall that the iteration numbers depend mildly on the level of the MLMC estimator and will be chosen by $n_\ell = \lceil n_0 h_\ell^{-\varepsilon} \rceil$ for $0 < \varepsilon \ll 1$. We use a triangulation of D, which is uniformly refined, resulting in FE spaces $\mathcal{V}_\ell \subset \mathcal{V}$, $\ell \geq 1$, with maximal meshwidth h_ℓ, which incorporate Dirichlet boundary conditions. The FE spaces are spanned by a nodal basis. The implementation uses the FE C++ library `BETL`, cf. [29], the GRF is implemented using `FFTW`, cf. [18], and the execution of MLMC is parallelized using the MPI-based wrapper `gMLQMC`, cf. [20]. The $L^2(\Omega; \mathcal{V})$-norm will be estimated by

$$\| \mathbb{E}(u) - E_L^{ML}(U_L^{n_L}) \|_{L^2(\Omega;\mathcal{V})} \approx \sqrt{\frac{1}{R-1} \sum_{j=1}^{R} \left\| \mathbb{E}(u) - \widehat{E_L^{\mathrm{ML}}(U_L^{n_L})}^{(j)} \right\|_{\mathcal{V}}^2},$$

where R i.i.d. realizations of $E_L^{\mathrm{ML}}(U_L^{n_L})$ are used. The reference solution is approximated by the average of $R = 20$ realizations of $E_L^{\mathrm{ML}}(U_L^{n_L})$ with $L = 9$, sample numbers (38) and (39) with $M^* = 40$, and a sparse direct solver.

In Figures 1, 2(a), and 2(b), the error is plotted versus the degrees of freedom in the FE space of the highest level that is active in the MLMC estimator. The empirical rate is computed by least squares taking into account the five data pairs corresponding to finer resolution. The fitted lines are shifted down for better visability. We remark that the relation to the total required work is asymptotically $\mathrm{work}_L = \mathcal{O}(h_L^{-d-\varepsilon})$. Thus, the rate implied by Theorem 6.4 is for $\nu = 1.0$ approximately 0.5, see Figure 1. For $\nu = 0.5$, Theorem 6.4 implies a rate of approximately 0.25. We observe that the performance of the iterative solver is in the required range of accuracy as good as a sparse direct solver. In Figures 2(a) and 2(b), the rate seems to suffer from the smaller value of the correlation length. But the performance of the iterative solver is still as good as the sparse direct solver.

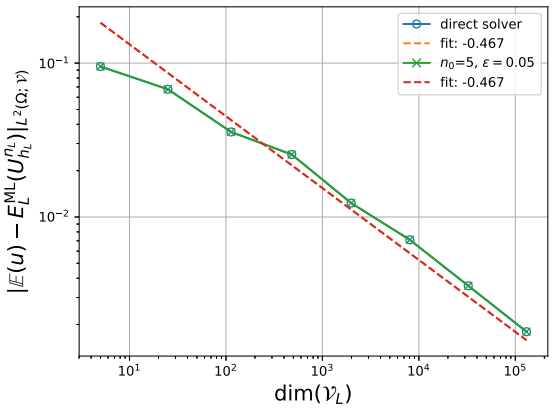
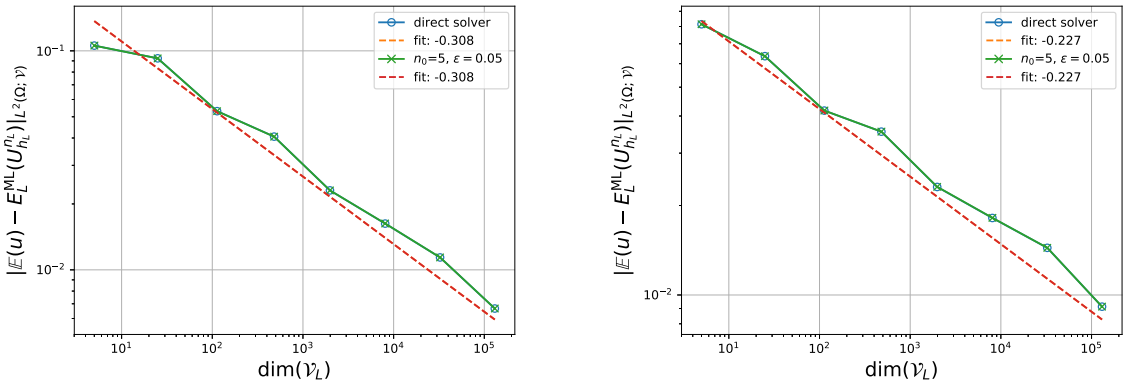


Figure 1: Comparison of Full Multigrid Cycle using PCG with BPX with $n_0 = 5$, $\varepsilon = 0.05$ to a sparse direct solver for GRF input with $\nu = 1.0, \sigma = 1.0, \lambda = 1.0$, and $R = 20$



(a) $\nu = 0.5, \sigma = 1.0, \lambda = 0.5$, and $R = 20$

(b) $\nu = 0.5, \sigma = 1.0, \lambda = 0.1$, and $R = 20$

Figure 2: Comparison of Full Multigrid Cycle using PCG with BPX with $n_0 = 5$, $\varepsilon = 0.05$ to a sparse direct solver

8 Conclusions and extensions

In the study of random operator equations with non-uniform input a rigorous framework has been established to verify strong convergence of a wide range of iterative solvers. This includes standard solvers such as multigrid, domain decomposition, and preconditioned conjugate gradient. In the case of lognormal random input, essentially optimal, deterministic complexity bounds are implied. This offers an alternative to direct solvers for this type of problems. Numerical experiments with CG preconditioned by BPX confirm the theoretical results and we conclude that standard iterative solvers are applicable in the case of lognormal Hölder continuous input. This was tested here in connection with MLMC for the approximation of the expectation of the random solution. Assumptions on the computational cost of PDE solvers are also common in the context of multilevel quasi-Monte Carlo methods (MLQMC) for PDE problems, cf. [32, 31, 21, 28]. Applicability of iterative solvers for MLQMC will be analyzed in a forthcoming work.

References

[1] Robert J. Adler. *The Geometry of Random Fields*. John Wiley & Sons, Ltd., Chichester, 1981. Wiley Series in Probability and Mathematical Statistics.

[2] Randolph E. Bank and Todd Dupont. An optimal order process for solving finite element equations. *Math. Comp.*, 36(153):35–51, 1981.

[3] Andrea Barth, Christoph Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numer. Math.*, 119(1):123–161, 2011.

[4] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[5] James H. Bramble. *Multigrid Methods*. Pitman Research Notes in Mathematics Series, Vol. 294. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1993.

[6] James H. Bramble and Joseph E. Pasciak. New convergence estimates for multigrid algorithms. *Math. Comp.*, 49(180):311–329, 1987.

[7] James H. Bramble and Joseph E. Pasciak. The analysis of smoothers for multigrid algorithms. *Math. Comp.*, 58(198):467–488, 1992.

[8] James H. Bramble and Joseph E. Pasciak. New estimates for multilevel algorithms including the V-cycle. *Math. Comp.*, 60(202):447–471, 1993.

[9] James H. Bramble, Joseph E. Pasciak, Jun Ping Wang, and Jinchao Xu. Convergence estimates for multigrid algorithms without regularity assumptions. *Math. Comp.*, 57(195):23–45, 1991.

[10] James H. Bramble, Joseph E. Pasciak, and Jinchao Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55(191):1–22, 1990.

[11] James H. Bramble, Joseph E. Pasciak, and Jinchao Xu. The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms. *Math. Comp.*, 56(193):1–34, 1991.

[12] James H. Bramble and Xuejun Zhang. The analysis of multigrid methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 173–415. North-Holland, Amsterdam, 2000.

[13] Julia Charrier. Strong and weak error estimates for elliptic partial differential equations with random coefficients. *SIAM J. Numer. Anal.*, 50(1):216–246, 2012.

[14] Julia Charrier, Robert Scheichl, and Aretha Teckentrup. Finite element error analysis for elliptic PDEs with random coefficients and applications. *SIAM J. Numer. Anal.*, 50(1):322–352, 2013.

[15] Philippe G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[16] Gh. de Marsily, F. Delay, J. Gonçalvès, Ph. Renard, V. Teles, and S. Violette. Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1):161–183, Mar 2005.

[17] J. P. Delhomme. Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research*, 15(2):269–280, 1979.

[18] Matteo Frigo and Steven G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[19] D. Galindo, P. Jantsch, C. G. Webster, and G. Zhang. Accelerating stochastic collocation methods for partial differential equations with random input data. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):1111–1137, 2016.

[20] Robert N. Gantner. A Generic C++ Library for Multilevel Quasi-Monte Carlo. In *Proceedings of the Platform for Advanced Scientific Computing Conference*, PASC '16, pages 11:1–11:12, New York, NY, USA, 2016. ACM.

[21] Robert N. Gantner, Lukas Herrmann, and Christoph Schwab. Multilevel QMC with product weights for affine-parametric, elliptic PDEs. In J. Dick, F. Y. Kuo, and H. Woźniakowski, editors, *Comtemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan*, pages 373–405. Springer International Publishing, Cham, 2018.

[22] Michael B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[23] Michael B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.

[24] Pierre Grisvard. *Elliptic problems in nonsmooth domains*, volume 69 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Reprint of the 1985 original [MR0775683], With a foreword by Susanne C. Brenner.

[25] Wolfgang Hackbusch. *Multi-Grid Methods and Applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.

[26] Stefan Heinrich. Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing*, Lecture Notes in Computer Science, Vol. 2179, pages 58–67. Springer-Verlag, Berlin, 2001.

[27] Lukas Herrmann, Annika Lang, and Christoph Schwab. Numerical analysis of lognormal diffusions on the sphere. *Stoch. Partial Differ. Equ. Anal. Comput.*, 6(1):1–44, 2018.

[28] Lukas Herrmann and Christoph Schwab. Multilevel quasi-Monte Carlo integration with product weights for elliptic PDEs with lognormal coefficients. Technical Report 2017-19, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2017.

[29] Ralf Hiptmair and Lars Kielhorn. BETL – a generic boundary element template library. Technical Report 2012-36, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2012.

[30] Prashant Kumar, Cornelis W. Oosterlee, and Richard P. Dwight. A multigrid multilevel Monte Carlo method using high-order finite-volume scheme for lognormal diffusion problems. *Int. J. Uncertain. Quantif.*, 7(1):57–81, 2017.

[31] Frances Y. Kuo, Robert Scheichl, Christoph Schwab, Ian H. Sloan, and Elisabeth Ullmann. Multilevel Quasi-Monte Carlo methods for lognormal diffusion problems. *Math. Comp.*, 86(308):2827–2860, 2017.

[32] Frances Y. Kuo, Christoph Schwab, and Ian H. Sloan. Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients. *Found. Comput. Math.*, 15(2):411–449, 2015.

[33] Finn Lindgren, Hå vard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(4):423–498, 2011. With discussion and a reply by the authors.

[34] Gabriel J. Lord, Catherine E. Powell, and Tony Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[35] Catherine E. Powell and Elisabeth Ullmann. Preconditioning stochastic Galerkin saddle point systems. *SIAM J. Matrix Anal. Appl.*, 31(5):2813–2840, 2010.

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

[37] Aretha L. Teckentrup, Robert Scheichl, Michael B. Giles, and Elisabeth Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numer. Math.*, 125(3):569–600, 2013.

[38] Hans Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. Johann Ambrosius Barth, Heidelberg, second edition, 1995.

[39] Elisabeth Ullmann, Howard C. Elman, and Oliver G. Ernst. Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems. *SIAM J. Sci. Comput.*, 34(2):A659–A682, 2012.

[40] Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34(4):581–613, 1992.

[41] Jinchao Xu and Ludmil Zikatanov. Algebraic multigrid methods. *Acta Numer.*, 26:591–721, 2017.

[42] Jinchao Xu and Jun Zou. Some nonoverlapping domain decomposition methods. *SIAM Rev.*, 40(4):857–914, 1998.

[43] Harry Yserentant. Old and new convergence proofs for multigrid methods. *Acta Numer.*, 2:285–326, 1993.