Block and multilevel preconditioning for stochastic Galerkin problems with lognormally distributed parameters and tensor product polynomials

Ivana Pultarová

Mathematical Institute, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Prague 8, Czech Republic;
Department of Mathematics, Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29 Prague 6, Czech Republic; email: ivana.pultarova@cvut.cz

Abstract.
The stochastic Galerkin method is a popular numerical method for the solution of differential equations with randomly distributed data. We focus on isotropic elliptic problems with lognormally distributed coefficients. We study the block-diagonal preconditioning and the algebraic multilevel preconditioning based on the block splitting according to some hierarchy of approximation spaces for the stochastic part of the solution. We introduce upper bounds for the resulting condition numbers, and we derive a tool for obtaining sharp guaranteed upper bounds for the strengthened Cauchy-Bunyakowsky-Schwarz constant, which can serve as an indicator of the efficiency of some of these preconditioning methods. The presented multilevel approach yields a tool for efficient guaranteed two-sided a posteriori estimates of algebraic errors and for adaptive algorithms as well.

Keywords. Stochastic Galerkin method, differential equations with uncertainties, algebraic multilevel preconditioning, block-diagonal preconditioning, uncertainty quantification.

Grants. This work was supported by the ERC-CZ project LL1202.

1 Introduction

We focus on the numerical solution of elliptic differential equations with randomly distributed data. In practice, several approaches are usually used. We can apply the Monte Carlo method, which may suffer from slow convergence due to a large number of test problems which must be solved. The multilevel Monte Carlo method [22] is more efficient, but it still needs the solution of many deterministic problems and does not offer any systematic error estimate. Collocation methods need only a small number of samples and solutions of associated problems, but the whole set of sample parameters must be changed and associated problems must be recomputed if we need some refinement of the current solution. Introducing a weak formulation with respect to physical and stochastic components of the problem gives rise to the stochastic Galerkin method (SGM), which transforms the parametrized differential equation into a high dimensional deterministic problem [6, 8, 11, 39]. The convergence theory of SGM and a priori error estimates are well understood [6, 7, 11, 12] even for problems with lognormally distributed coefficients [5, 23]. Systems of linear equations arising from SGM and their spectral properties represent relatively new issues within numerical linear algebra [17, 28]. In this paper, we are interested in preconditioning and in a posteriori estimates of algebraic errors of approximate solutions of SGM.

Preconditioning methods usually use some kind of approximate inverse of a given matrix derived, for example, from some reduction or simplification of a given problem. For example, we may use diagonal blocks of the SGM matrix or blocks which correspond to some underlying deterministic problem. These two methods coincide and are called mean based preconditioning, if the coefficients of the equation are expanded into a series which is linear in stochastic variables [16, 28, 41]. They

are efficient especially if the coefficient of variation of the data is small, or for log-transformed problems [37]. The Kronecker product of two sparse matrices of stochastic and physical variables, respectively, was used and studied in [36]. It can be expected that any type of preconditioning employing some appropriate partitioning of physical variables can be successfully applied to SGM. For example, good efficiency of multigrid algorithms and of Schwarz and domain decomposition methods was reported in [9, 15, 25, 26] and in [34, 35], respectively. Some recent papers revealed that hierarchical or Schur complement preconditioning can be successfully applied to stochastic discretization spaces as well. Numerical results for normally or lognormally distributed coefficients for such methods were presented in [32, 33] and some upper bounds for the resulting condition numbers were theoretically proved for uniform or normal distributions in [30].

The goal of this paper is to study block-diagonal preconditioning and algebraic multilevel (AML) splittings of stochastic approximation spaces and associated preconditioning of problems with lognormally distributed coefficients. These coefficients yield more complicated structures of the matrices of discretized problems than the coefficients which depend linearly on random variables. The efficiency of AML preconditioners [2, 13, 24], can be quantified by means of the strengthened Cauchy-Bunyakowsky-Schwarz (CBS) constant. Sharp uniform upper bounds for the CBS constants can be obtained for many types of finite element (FE) discretization spaces, see, for instance, [3]. Although a coupling between physical and stochastic variables yield more complicated structures than those arising from deterministic FE methods, the CBS constants can be estimated for SGM in some cases. Theoretical proofs were obtained for problems with coefficients which depend linearly on random variables [29, 30]. In the present paper, we focus on problems with lognormally distributed parameters [5, 23, 32, 37, 38]. Up to our knowledge, this is the first time when the CBS constant for a hierarchical two-by-two matrix splitting for problems with lognormally distributed parameters has been theoretically estimated. There are two main assumptions used in this paper: the coefficients of the problem are assumed to be constant in every element of the physical domain, and stochastic approximation spaces are in the form of a tensor product of orthogonal multivariate polynomials. While the former assumption represents a usual assumption for the approximation of data, the latter could be quite restrictive. We must stress that, up to now, the results for the other frequently used discretization, complete polynomials, have not been available. Still, the introduced results present a new direction in the preconditioning of SGM and show how the original AML ideas can be applied to new problems. In addition, we show that the technique used for estimating the CBS constant, a kind of element-wise computation, can be adopted to obtain the upper bounds for block preconditioning for SGM. Let us also mention that the theoretical upper bounds for the CBS constants may yield efficient two-sided and guaranteed a posteriori estimates of the algebraic error of an approximate solution.

In the following section, we introduce our model problem: a differential equation and the parametrization of random input data. The elements of the SGM matrix are derived in section 3 for tensor products of polynomials used for the discretization of stochastic components of the solution. For theoretical purposes, matrix elements are evaluated using shifted Hermite orthogonal polynomials on each element separately. In section 4, we recall the AML strategy for a hierarchical two-by-two block splitting of a matrix, or, equivalently, for the splitting of approximation finite-dimensional discretization spaces. We present a lemma which enables us to evaluate the upper bounds for the CBS constants for SGM from a matrix associated with a single reference element only. We show how the upper bound for the CBS constant indicates the efficiency of some AML preconditioning methods. We recall the formula for two-sided a posteriori error estimates. In section 5, we formulate the main results of the paper, the guaranteed upper bounds for the CBS constants for elliptic differential equations with isotropic lognormally distributed parameters using tensor product polynomials for the discretization of the stochastic component of the solution and the upper bound for the condition number for the block diagonal preconditioning. Simple numerical experiments in section 6 illustrate the obtained theoretical results and the paper is concluded by a short discussion.

2 Problem setting

Let $(\Omega, \mathcal{F}, \mu)$ be a complete probability space where Ω is the sample space with σ-algebra $\mathcal{F}$ of its subsets and with the probability measure $\mu : \mathcal{F} \to \langle 0, 1\rangle$. For a real valued random variable ξ in $(\Omega, \mathcal{F}, \mu)$, we denote the expectation by

$$E[\xi] = \int_\Omega \xi(\omega)\,\mathrm{d}\mu(\omega) = \int_{\mathcal{R}} z\rho(z)\,\mathrm{d}z,$$

where $\rho : \mathcal{R} \to \mathcal{R}$ is the associated probability density function. Let $D \subset \mathcal{R}^d$, $d = 1, 2$ or 3, be a bounded Lipschitz domain and let ∂D be the boundary of D. We solve a model elliptic boundary value problem in almost sure sense [5, 23, 39]

$$-\nabla \cdot (a(x,\omega)\nabla u(x,\omega)) = f(x) \tag{1}$$

in $D \times \Omega$ with the homogeneous Dirichlet condition on the boundary $\partial D \times \Omega$. The techniques and methods developed in this paper can be directly applied to problems with non-homogeneous boundary conditions or with a combination of Dirichlet and Neumann boundary conditions, cf. also section 4.1. The variable $x \in D$ denotes physical coordinates and the gradient symbol denotes differentiation with respect to x. The function $f \in L^2(D)$ is deterministic, but the results of this paper extend to f dependent on ω.

Our interest is devoted to problems where the diffusion coefficient $a(x,\omega)$ is a lognormally distributed scalar random field. Let $y = (y_1, \ldots, y_N) : \Omega \to \Gamma$ be a vector of random variables which are independent and normally distributed with zero mean and unit variance and with the same probability density $\rho(y_i) = \mathrm{e}^{-y_i^2/2}/\sqrt{2\pi}$. Here, $\Gamma = \mathcal{R}^N$ is the joint image and $\bar\rho(y) = \prod_{i=1}^N \rho(y_i)$ is the joint probability density function of y_i, $i = 1, \ldots, N$. We consider the diffusion coefficient $a(x,\omega) :\ D \times \Omega \to \mathcal{R}$ of our model problem in the form

$$a(x,\omega) = \exp\left(a_0(x) + \sum_{i=1}^N a_i(x) y_i(\omega)\right). \tag{2}$$

The logarithm of $a(x,\omega)$ may represent, for example, the truncated Karhunen-Loève expansion of some random process. Then $a_i(x) \in L^2(D)$ would be normalized and pair-wise orthogonal eigenfunctions of some covariance operator $C(x,x')$ defined on D^2, which would be multiplied by the square roots of the corresponding eigenvalues [7, 8, 18, 23, 37]. Since the eigenvalues form a sequence of real non-negative decreasing numbers which tend to zero, the norms of a_k in (2) tend to zero as well. In what follows, we do not specify where the functions a_k come from, we only suppose that they are constant on the interiors of the elements used for the discretization of D. The Doob-Dynkin lemma [5, 10, 16] says that u satisfying (1) with a defined by (2) can be expressed as a random field for every $x \in D$, i.e. $u = u(x,y)$, and

$$\int_\Omega g(y(\omega))\,\mathrm{d}\mu(\omega) = \int_\Gamma g(y)\bar\rho(y)\,\mathrm{d}y$$

can be used to evaluate all integrals over the sample space Ω for all Borel functions g of y. As a consequence, we can introduce the weak formulation of (1). Let us define the bilinear form $\mathcal{A} : H \times H \to \mathcal{R}$,

$$\mathcal{A}(u,v) = \int_\Gamma \int_D a\nabla u \cdot \nabla v \bar\rho\,\mathrm{d}x\mathrm{d}y, \tag{3}$$

where $H = H_0^1(D) \otimes L^2_{\bar\rho}(\Gamma)$, where $L^2_{\bar\rho}(\Gamma)$ is the weighted Sobolev space of square-integrable functions defined on Γ with the weight function $\bar\rho$. $\mathcal{A}(u,v)$ is the energy scalar product inducing the energy norm in a usual way [5, 6, 7, 18, 28]

$$\|u\|_{\mathcal{A}} = \int_\Gamma \int_D a\,|\nabla u|^2\,\bar\rho\,\mathrm{d}x\,\mathrm{d}y.$$

For f from the dual space H' and for $v \in H$, let

$$\langle f, v \rangle = \int_\Gamma \int_D f v \bar{\rho} \, \mathrm{d}x \, \mathrm{d}y.$$

The weak formulation of problem (1) then reads: find $u \in H$ such that

$$\mathcal{A}(u, v) = \langle f, v \rangle \tag{4}$$

for all $v \in H$. Due to the normal distribution of y_i in (2), $i = 1, \ldots, N$, the coefficient a is not bounded away from zero and from infinity. In other words, there do not exist any constants $c_1, c_2 \in \mathcal{R}$, such that

$$0 < c_1 \leq a(x, \omega) \leq c_2 < \infty$$

almost sure and almost everywhere in D. However, following Lemma 1.2 in [5], we can prove the existence and the uniqueness of the solution u to problem (4) and the boundedness of moments of u up to the order k if $f \in L^2(D) \otimes L^{2k}(\Omega)$. In our setting, $f \in L^2(D)$ is deterministic, then this assumption is trivially fulfilled. Another approach to prove the well posedness of (4) was introduced by Gittelson in [23]. To avoid difficulties in numerical computation, one can also use the truncated normal probability distribution of y_i and corresponding sets of orthogonal polynomials as the bases of approximation spaces, see [37].

3 Stochastic Galerkin method

The discretization of model problem (4) can be obtained by its projection into some finite-dimensional subspace V of H and by using the truncated generalized polynomial chaos expansion [8] of an approximate solution

$$u(x, y) = \sum_{j=1}^{M} u_j(x) \Phi_j(y), \tag{5}$$

where $\{\Phi_j(y)\}_{j=1}^M$ is a set of N-variate polynomials, which are orthogonal with respect to the scalar product of $L^2_{\bar{\rho}}(\Gamma)$. This method is known as the stochastic Galerkin method (SGM) or as the spectral Galerkin approach [5, 6, 11, 21]. Let $u_j \in V_x \subset H_0^1(D)$, $j = 1, \ldots, M$, where V_x is a span of (bi)linear finite element basis functions $\psi_r(x)$, $r = 1, \ldots, F$, defined on D. Thus,

$$u(x, y) = \sum_{j=1}^{M} \sum_{r=1}^{F} u_{jr} \psi_r(x) \Phi_j(y), \tag{6}$$

$u_{jr} \in \mathcal{R}$. Let $\Phi_j(y)$ be products of univariate polynomials

$$\Phi_j(y) = \phi_{j_1}(y_1) \phi_{j_2}(y_2) \ldots \phi_{j_N}(y_N), \tag{7}$$

where ϕ_k denotes a polynomial of the order $k = 0, 1, \ldots$, and let V_y be the span of $\Phi_1(y), \ldots, \Phi_M(y)$. Two main types of sets of N-variate polynomials $\Phi_j(y)$ are usually considered for the approximation off u: complete polynomials of the total degree less than or equal to a given constant P, i.e. $0 \leq j_1 + \ldots + j_n \leq P$, or tensor products of univariate polynomials of orders which do not exceed the given constants P_k, i.e. $j_k \leq P_k$ for $k = 1, \ldots, N$. Complete polynomials are recommended to be used, for example, if the norms of functions a_k, $k = 1, \ldots, N$, in (2) are approximately the same, because it can be expected that the solution can be well approximated by polynomials $\Phi_i(y)$ where the variables y_i have almost the same influence. The tensor product polynomials can be more appropriate, for example, if the norms of a_k decay rapidly with the growing k. In this case, the maximal degrees of polynomials $\phi_{j_k}(y_k)$ of the expansion of a decay with a similar

rate if k increases, see [18], and we may expect that the solution can be accurately and efficiently approximated by $\Phi_i(y)$ where the maximal orders of $\phi_{j_k}(y_k)$ get smaller when k increases. Thus, it can be more efficient to use tensor product polynomials with decreasing P_k than a set of complete polynomials. In this paper, we only consider tensor product polynomials. The dimension of a set of tensor product polynomials of N variables of degrees P_k, $k = 1, \ldots, N$, is

$$M = \prod_{i=1}^{N} (P_i + 1).$$

The finite-dimensional solution space of (4) is then the tensor product space $V = V_x \otimes V_y$ of the dimension $M \times F$. The convergence analysis and a priori error estimates for these discretization spaces can be found, for example, in [5, 6, 11, 12, 23]. The discretized problem is represented by a system of $M \times F$ linear equations of $M \times F$ unknowns

$$Au = B.$$

We denote a function $u \in V$ and its vector representation with respect to some fixed basis of V with the same letter. Let us denote the coordinates of A, u, and B by pairs of indices, say j, r, such that the first index j is the coordinate with respect to the basis functions $\Phi_j(y)$, $j = 1, \ldots, M$, of V_y, and the second index r is the coordinate with respect to the FE basis functions $\psi_r(x)$ of V_x. Let us order the basis functions $\Phi_j(y)\psi_r(x)$ of V lexicographically, i.e. in such a way that the second subscript is changing faster than the first one. Let the N-variate polynomials $\Phi_j(y)$ be ordered in such a manner that $j < m$ whenever $\Phi_j(y) = \phi_{j_1}(y_1) \ldots \phi_{j_N}(y_N)$, $\Phi_m(y) = \phi_{m_1}(y_1) \ldots \phi_{m_N}(y_N)$ and $(j_1, \ldots, j_N) < (m_1, \ldots, m_N)$ according to the anti-lexicographical ordering, which means, for example, that the first element is changing fastest.

Let us denote by A_{jk} the block of A of the size $F \times F$ associated with the basis polynomials Φ_j and Φ_k, and let $(A_{jk})_{rs}$ be an element of this block corresponding to the finite element basis functions ψ_r and ψ_s. Let us order the elements of the solution vector u and of the right hand side B adequately. Then the elements of A and B are

$$\begin{aligned} (A_{jk})_{rs} &= \frac{1}{(2\pi)^{N/2}} \int_{\mathcal{R}^N} \int_D \exp\left(a_0(x) + \sum_{i=1}^{N} a_i(x) y_i \right) \\ &\quad \times \nabla\psi_r(x) \cdot \nabla\psi_s(x) \Phi_j(y) \Phi_k(y) \mathrm{e}^{-(y_1^2 + \ldots + y_N^2)/2} \, \mathrm{d}x \, \mathrm{d}y \\ B_{jr} &= \frac{1}{(2\pi)^{N/2}} \int_{\mathcal{R}^N} \int_D f(x) \psi_r(x) \Phi_j(y) \mathrm{e}^{-(y_1^2 + \ldots + y_N^2)/2} \, \mathrm{d}x \, \mathrm{d}y. \end{aligned} \tag{8}$$

Note that $B_{jr} = 0$ for $j > 0$ whenever f is deterministic. Since we focus on the spectral properties of preconditioning of A, our starting point is to find suitable formulas for evaluating the integral

$$\int_{\mathcal{R}^N} \exp\left(a_0(x) + \sum_{i=1}^{N} a_i(x) y_i \right) \Phi_j(y) \Phi_k(y) \mathrm{e}^{-(y_1^2 + \ldots + y_N^2)/2} \, \mathrm{d}y,$$

or, especially,

$$\int_{\mathcal{R}} \exp\left(a_i(x) y_i \right) \phi_{j_i}(y_i) \phi_{k_i}(y_i) \mathrm{e}^{-y_i^2/2} \, \mathrm{d}y_i,$$

for $i = 1, \ldots, N$, provided that $a_i(x)$ are constant on every element.

3.1 Shifted Hermite polynomials

Let Φ_j, $j = 1, \ldots, M$, be the multivariate orthogonal polynomials defined by (7) where the univariate polynomials ϕ_k are normalized Hermite polynomials orthogonal with respect to the weighted

scalar product of $L^2_\rho(\mathcal{R})$, where $\rho(z) = \mathrm{e}^{-z^2/2}/\sqrt{2\pi}$. The normalized Hermite polynomials ϕ_k can be obtained recursively from $\phi_0(z) = 1$, $\phi_1(z) = z$, and

$$\sqrt{k+1}\phi_{k+1}(z) = z\phi_k(z) - \sqrt{k}\phi_{k-1}(z). \tag{9}$$

for $k \geq 2$ [19, 31]. Then

$$\frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} (\phi_k(z))^2 \, \mathrm{e}^{-z^2/2} \, \mathrm{d}z = 1$$

and

$$\frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}} z\phi_k(z)\phi_{k+1}(z) \, \mathrm{e}^{-z^2/2} \, \mathrm{d}z = \sqrt{k+1}. \tag{10}$$

One can easily prove by induction that a shifted Hermite polynomial is given by

$$\phi_n(z+c) = \sum_{k=0}^{n} \binom{n}{k} \sqrt{\frac{k!}{n!}} c^{n-k} \phi_k(z). \tag{11}$$

Then, for $c \in \mathcal{R}$, we have

$$\begin{aligned}
\int_{\mathcal{R}} \phi_j(z)\phi_k(z)\mathrm{e}^{cz}\mathrm{e}^{-z^2/2} \, dz &= \int_{\mathcal{R}} \phi_j(z)\phi_k(z)\mathrm{e}^{c^2/2}\mathrm{e}^{-(z-c)^2/2} \, dz \\
&= \mathrm{e}^{c^2/2} \int_{\mathcal{R}} \phi_j(z)\phi_k(z)\mathrm{e}^{-(z-c)^2/2} \, dz \\
&= \mathrm{e}^{c^2/2} \int_{\mathcal{R}} \phi_j(z+c)\phi_k(z+c)\mathrm{e}^{-z^2/2} \, dz \\
&= \mathrm{e}^{c^2/2} \int_{\mathcal{R}} \sum_{m=0}^{j} \binom{j}{m} \sqrt{\frac{m!}{j!}} c^{j-m} \phi_i(z) \sum_{n=0}^{k} \binom{k}{n} \sqrt{\frac{n!}{k!}} c^{k-n} \phi_n(z) \mathrm{e}^{-z^2/2} \, dz \\
&= \mathrm{e}^{c^2/2} \int_{\mathcal{R}} \sum_{m=0}^{j} \sum_{n=0}^{k} \binom{j}{m} \binom{k}{n} \sqrt{\frac{m!}{j!}\frac{n!}{k!}} c^{j+k-m-n} \phi_m(z)\phi_n(z) \mathrm{e}^{-z^2/2} \, dz \\
&= \mathrm{e}^{c^2/2} \sum_{m=0}^{j} \sum_{n=0}^{k} \binom{j}{m} \binom{k}{n} \sqrt{\frac{m!}{j!}\frac{n!}{k!}} c^{j+k-m-n} \int_{\mathcal{R}} \phi_m(z)\phi_n(z) \mathrm{e}^{-z^2/2} \, dz \\
&= \mathrm{e}^{c^2/2} \sum_{m=0}^{j} \sum_{n=0}^{k} \binom{j}{m} \binom{k}{n} \sqrt{\frac{m!}{j!}\frac{n!}{k!}} c^{j+k-m-n} \delta_{nm} \sqrt{2\pi} \\
&= \mathrm{e}^{c^2/2} \sqrt{2\pi} \sum_{m=0}^{\min(j,k)} \binom{j}{m} \binom{k}{m} \frac{m!}{\sqrt{j!k!}} c^{j+k-2m}.
\end{aligned} \tag{12}$$

We make use of the last expression in the subsequent sections.

3.2 Stochastic Galerkin matrices for lognormally distributed parameters

The elements of A defined by (8) can be approximated in several different ways. In [17, 32], the standard polynomial chaos expansion of $a(x, y)$ in terms of N-variate Hermite orthogonal polynomials [20] is considered

$$a(x,y) = \exp\left(a_0(x) + \sum_{i=1}^{N} a_i(x) y_i\right) \approx \sum_{k=0}^{\tilde{M}} \tilde{a}_k(x)\Phi_k(y), \tag{13}$$

where

$$\tilde{a}_k(x) = \frac{\mathrm{E}\left[a(x,y)\Phi_k(y)\right]}{\mathrm{E}\left[\Phi_k^2(y)\right]} = \frac{\mathrm{E}\left[\Phi_k(y+h(x))\right]}{\mathrm{E}\left[\Phi_k^2(y)\right]} \exp\left(a_0(x) + \frac{1}{2}\sum_{i=1}^{N} a_i^2(x)\right),$$

where $h(x)$ is the deterministic vector $h(x) = (a_1(x), \ldots, a_N(x))$. The truncated polynomial chaos expansion (13) of $a(x, y)$ can be then substituted into (2) and (8). Note that $\mathrm{E}\left[\Phi_k(y + h(x))\right]$ could be evaluated making use of shifted Hermite polynomials for every $x \in D$. In the evaluation of (8), it is then necessary to integrate the products of triples of orthogonal polynomials

$$\int_{\mathcal{R}} \phi_i(z)\phi_j(z)\phi_k(z)\mathrm{e}^{-z^2/2}\,\mathrm{d}z,$$

which is guaranteed to be zero if, for example, $k > i + j$. This means that $\Phi_k(y)$ in the expansion (13) of $a(x, y)$ must be considered up to the degree $2P + 1$ in all variables y_i whenever one uses polynomials of the degrees P for the solution and test spaces. Otherwise, the matrix A may become semidefinite or indefinite, cf. [27]. Thus, $\tilde{M}$ in (13) must be sufficiently large. This form of the expansion and evaluation of a allows us to compute and store only component blocks (of sizes $M \times M$ or $F \times F$) instead of the whole Kronecker product matrix A.

In our paper, we use a formally different way to evaluate the entries of A. However, if all $a_i(x)$ are constant on each element, the resulting values of the entries of A are the same as if they were computed in the classical way presented above. The reason why we express the entries of A in a different form is purely theoretical: this provides a better insight into the structure of A. In particular, we will see that if $a_i(x)$, $i = 1, \ldots, N$, are constant on every element, the energy scalar product (3) of two basis functions of V integrated only over a single reference element has the form of Kronecker product of matrices with simple structures. The entries of A given by (8) are

$$\begin{aligned}(A_{jk})_{rs} &= \frac{1}{(2\pi)^{N/2}} \int_D \mathrm{e}^{a_0(x)} \nabla\psi_r(x) \cdot \nabla\psi_s(x) \\ &\quad \cdot \int_{\mathcal{R}^N} \mathrm{e}^{a_1(x)y_1 + \ldots + a_n(x)y_n} \Phi_j(y)\Phi_k(y)\mathrm{e}^{-(y_1^2+\ldots+y_N^2)/2}\,\mathrm{d}y\,\mathrm{d}x.\end{aligned}$$

Let us denote

$$\begin{aligned}\tilde{G}_{ijk}(x) &= \frac{1}{\sqrt{2\pi}} \int_{\mathcal{R}^N} \mathrm{e}^{a_i(x)z} \phi_j(z)\phi_k(z)\mathrm{e}^{-z^2/2}\,\mathrm{d}z \\ &= \mathrm{e}^{a_i(x)^2/2} \sum_{m=0}^{\min(j,k)} \binom{j}{m}\binom{k}{m} \frac{m!}{\sqrt{j!k!}}\, a_i(x)^{j+k-2m}, \end{aligned} \tag{14}$$

where the last equality follows from (12). Then

$$\begin{aligned}(A_{jk})_{rs} &= \int_D \prod_{i=1}^{N} \tilde{G}_{ij_ik_i}(x)\, \mathrm{e}^{a_0(x)} \nabla\psi_r(x) \cdot \nabla\psi_s(x)\,\mathrm{d}x, \\ &= \int_D G_{jk}(x)\, \mathrm{e}^{a_0(x)} \nabla\psi_r(x) \cdot \nabla\psi_s(x)\,\mathrm{d}x, \end{aligned} \tag{15}$$

for $r, s = 1, \ldots, F$, $i = 1, \ldots, N$, $j, k = 1, \ldots, M$, where

$$G_{jk}(x) = \prod_{i=1}^{N} \tilde{G}_{ij_ik_i}(x),$$

and where $\Phi_j(y) = \phi_{j_1}(y_1) \ldots \phi_{j_N}(y_N)$, $\Phi_k(y) = \phi_{k_1}(y_1) \ldots \phi_{k_N}(y_N)$.

Let us introduce some observations. The matrix A is symmetric and positive definite for any set of approximation polynomials $\Phi_j(y)$, $j \in I \subset \{1, \ldots, M\}$. This means that every diagonal block A_{jj} of A is symmetric and positive definite. Each $F \times F$ block A_{jk} of A corresponding to $\Phi_j(y)$ and $\Phi_k(y)$ is sparse and its non-zero pattern is the same as the non-zero pattern of the stiffness matrix of the corresponding deterministic problem.

The main goal of this paper is to study the spectral properties of block-diagonal and hierarchical preconditioning of A. We will consider only such a splitting of V that corresponds to some collection of subsets of polynomials $\Phi_j(y)$. It means that we do not use any partitioning with respect to physical unknowns. Some details of a general analysis of the two-by-two hierarchical preconditioning can be found in the next section.

4 Algebraic multilevel preconditioning

The AML preconditioning is based on a hierarchical decomposition of the finite-dimensional solution spaces. The main quantity which can serve as a "measure of efficiency" is the strengthened Cauchy-Bunyakowsky-Schwarz (CBS) constant [2, 13, 24]. If there is a guaranteed upper bound for the CBS constant which is sufficiently small and if the ratios of dimensions of any two consecutive hierarchical spaces are sufficient, we can efficiently employ some types of algebraic multilevel preconditioning.

Let us consider an elliptic differential equation with Dirichlet or Dirichlet and Neumann boundary conditions, possibly with random input data. Let $\mathcal{A}(u,v)$ be the energy scalar product induced by the weak form of the problem and defined on some appropriate function space H. Let $V \subset H$ be some finite-dimensional discretization space which can be expressed as the direct sum of its subspaces U and W of positive dimensions, $U \cap W = 0$,

$$V = U \oplus W.$$

Let A be the matrix of the resulting system of linear equations accordingly decomposed into the two-by-two block form,

$$Au = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = B. \tag{16}$$

Then A, A_{11} and A_{22} are symmetric positive definite matrices. We again do not distinguish between the function $u \in V$ and its vector representation with respect to some fixed basis of V. Thus, we can write $v^T A u$ instead of $\mathcal{A}(u,v)$ for some real vectors or for $u, v \in V$, respectively.

4.1 CBS constant

The CBS constant $\gamma \in \langle 0, 1)$ [2, 13, 24] for subspaces U and W with respect to a symmetric and positive semidefinite bilinear form $\mathcal{A}(u,v)$ is the smallest real $\gamma \geq 0$ satisfying

$$\mathcal{A}(u,w)^2 \leq \gamma^2 \, \mathcal{A}(u,u) \, \mathcal{A}(w,w), \qquad u \in U, \; w \in W,$$

or, equivalently,

$$(v_1^T A_{12} v_2)^2 \leq \gamma^2 \, v_1^T A_{11} v_1 \; v_2^T A_{22} v_2,$$

where v_1 and v_2 are any real vectors of appropriate dimensions.

Making use of specific discretization spaces, like, for example, hierarchical FE spaces, a guaranteed upper bound for γ can be easily obtained after the "restriction" of V, U, W and $\mathcal{A}(u,v)$ on a single reference "macro-element", and from the splitting of such a "local" space according to $V = U \oplus W$. This is a classical tool used in the AML theory, which appears, for example, in [2, 3, 4, 13, 24]. We introduce a variant of this proposition, which is general enough and thus applicable to SGM. Let us consider a set of symmetric positive semidefinite bilinear forms $\mathcal{A}_i(\cdot,\cdot) : V \times V \to \mathcal{R}$, $i = 1, \ldots, n$, and let $\mathcal{A}(\cdot,\cdot)$ be a symmetric positive definite bilinear form (scalar product) defined by

$$\mathcal{A}(u,v) = \sum_{i=1}^{n} \mathcal{A}_i(u,v), \tag{17}$$

for all $u, v \in H$.

Lemma 4.1 *Let $\mathcal{A}(\cdot,\cdot) : V \times V \to \mathcal{R}$ be a scalar product, $V = U \oplus W$, $U \cap W = 0$ and let (17) hold. Denote by $\gamma_i \in \langle 0, 1)$ the CBS constants for $\mathcal{A}_i(\cdot,\cdot)$ and for subspaces U and W, i.e.*

$$\mathcal{A}_i(u,w)^2 \le \gamma_i^2 \, \mathcal{A}_i(u,u) \, \mathcal{A}_i(w,w), \qquad u \in U, \; w \in W, \qquad i = 1, \dots, n.$$

Then for the CBS constant γ for $\mathcal{A}(\cdot,\cdot)$ and for U and W, we have $\gamma \le \max_{i=1,\dots,n} \gamma_i$.

Proof. Denote $\beta = \max_{i=1,\dots,n} \gamma_i$. Then from $\mathcal{A}(u,w) = \sum_{i=1}^n \mathcal{A}_i(u,w)$ and from the Hölder inequality, we have for any $u \in U$ and $w \in W$,

$$\begin{aligned}
\mathcal{A}(u,w)^2 &= \left(\sum_{i=1}^n \mathcal{A}_i(u,w) \right)^2 \le \left(\sum_{i=1}^n |\mathcal{A}_i(u,w)| \right)^2 \\
&\le \left(\sum_{i=1}^n \gamma_i \sqrt{\mathcal{A}_i(u,u)\mathcal{A}_i(w,w)} \right)^2 \\
&\le \beta^2 \left(\sum_{i=1}^n \sqrt{\mathcal{A}_i(u,u)\mathcal{A}_i(w,w)} \right)^2 \\
&\le \beta^2 \left(\sum_{i=1}^n \mathcal{A}_i(u,u) \right) \left(\sum_{j=1}^n \mathcal{A}_j(w,w) \right) \\
&= \beta^2 \mathcal{A}(u,u)\mathcal{A}(w,w),
\end{aligned}$$

which completes the proof.

Lemma 4.1 allows us to estimate the CBS constant of the energy scalar product (3) using the integration only over individual elements of the domain D. In this case $\mathcal{A}(\cdot,\cdot)$ corresponds to the integration over D and $\mathcal{A}_i(\cdot,\cdot)$ corresponds to the integration over some element D_i provided the interiors of the elements are pair-wise disjoint and their union equals D. Moreover, for estimating γ_i it is enough to consider only a general reference element. Note that $\mathcal{A}_i(\cdot,\cdot)$ is positive definite if D_i coincides with a part of the boundary where the Dirichlet boundary condition is applied, and $\mathcal{A}_i(\cdot,\cdot)$ is positive semi-definite, otherwise. The latter case corresponds to the Neumann problem on D_i and it is also used for the reference element to obtain the upper bound for the CBS constants γ_i. Since we consider the splitting of the approximation spaces only with respect to stochastic variables in this paper, neither the shape nor the size of the elements influence this estimate. Let us mention that the technique of estimating the CBS constant of SGM and of estimating the condition number for the block diagonal preconditioning is the main contribution of this paper. This technique is based on considering (3) on a single reference element.

4.2 Algebraic multilevel methods

In this paper, we define the condition number $\hat{\kappa}(B^{-1}A)$ of the product of a symmetric matrix A and its symmetric approximate inverse B^{-1} by the ratio of the maximal and minimal eigenvalues of $B^{-1}A$, which is identical to the common definition of the condition number of $B^{-1/2}AB^{-1/2}$, $\kappa(B^{-1/2}AB^{-1/2})$ (but it is not the same as $\kappa(B^{-1}A)$, because the eigenvalues of $B^{-1}A^2B^{-1}$ and of $B^{-1/2}AB^{-1}AB^{-1/2}$ may be different if A and B do not commute). Let us now recall the principles of AML preconditioning methods, which are fully described e.g. in [2, 13, 24]. Let A be considered in a two-by-two block form of (16), and let us assume that the block A_{11} can be again hierarchically

decomposed into two-by-two blocks. Let for any such hierarchical two-by-two block partitioning the CBS constant be not greater than $\gamma \in \langle 0, 1)$, i.e.

$$(u^T A_{12} v)^2 \leq \gamma^2 u^T A_{11} u \, v^T A_{22} v,$$

for all vectors u and v of appropriate sizes. Let us further assume that it is relatively easy to solve a set of linear equations with the matrix A_{22} and that it is relatively costly to solve a system with A_{11}. The matrices A and A^{-1} can be decomposed into

$$\begin{aligned} A &= \begin{pmatrix} I & A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ A_{22}^{-1}A_{21} & I \end{pmatrix} \\ A^{-1} &= \begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix} \begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}. \end{aligned}$$

Denote the Schur complement of A_{22} by $S = A_{11} - A_{12}A_{22}^{-1}A_{21}$. In AML methods, S^{-1} is approximated by

$$Z = M_{11}^{-1} Q \left(A_{11} M_{11}^{-1} \right),$$

where M_{11} is either equal to A_{11}, or is an approximation to it. Here Q should be a polynomial of a low order fulfilling $Q(1) = 1$. Thus, for the preconditioner of A, we can take

$$M^{-1} = \begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix} \begin{pmatrix} M_{11}^{-1}Q(A_{11}M_{11}^{-1}) & 0 \\ 0 & A_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}. \tag{18}$$

For example, if we use $M_{11}^{-1} = A_{11}^{-1}/(1-\gamma^2)$, then we get for the spectrum of $M^{-1}A$

$$\sigma(M^{-1}A) = \{1/(1-\gamma^2)\} \cup \sigma(A_{11}^{-1}S/(1-\gamma^2)) \subset \langle 1, 1/(1-\gamma^2) \rangle. \tag{19}$$

Thus, if we use for preconditioning only the blocks A_{11} and A_{22}, i.e. if $M_{11} = A_{11}$, the resulting method corresponds exactly to the two-by-two block Gauss-Seidel preconditioning [33, 42] and $\hat{\kappa}(M^{-1}A) \leq 1/(1-\gamma^2)$. When we substitute $M_{11}^{-1} (\approx A_{11}^{-1})$ recursively in the same manner as M^{-1} in (18), we obtain the V-cycle type of the AML preconditioning. This method will be called AML-V in our numerical examples in section 6. The condition number is then bounded by $1/(1-\gamma^2)^{L-1}$, where L is the number of levels, cf. [24]. If the CBS constants differ on different levels, the resulting condition number is bounded by $1/((1-\gamma_1^2)\cdots(1-\gamma_{L-1}^2))$, where γ_k are particular CBS constants. Numerical examples will be presented in section 6.

We can also only use the two-by-two block diagonal matrix with blocks A_{11} and A_{22} for the preconditioning of A. Then the resulting condition number is bounded by

$$\hat{\kappa}\left(\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}^{-1} A \right) \leq \frac{1+\gamma}{1-\gamma}, \tag{20}$$

see [24]. This method will be called B2 in section 6.

Let us assume that $\sigma(A_{11}M_{11}^{-1}) \subset \langle 1, \lambda \rangle$. Then, there exists a polynomial Q of the first degree such that

$$\sigma\left(A_{11}M_{11}^{-1}Q(A_{11}M_{11}^{-1})\right) \subset \left\langle 1, \frac{(1+\lambda)^2}{4\lambda} \right\rangle.$$

Indeed, we can choose $Q(t) = \alpha(\beta - t)$ where $\beta = 1 + \lambda$ and $\alpha = 1/\lambda$. Moreover, we have

$$\sigma\left(\frac{1}{1-\gamma^2} A_{11}^{-1}(A_{11} - A_{12}A_{22}^{-1}A_{21}) \right) \subset \frac{1}{1-\gamma^2} \langle 1-\gamma^2, 1 \rangle = \left\langle 1, \frac{1}{1-\gamma^2} \right\rangle.$$

Let us come back to the preconditioning defined by (18), where the exact term S^{-1} is substituted by $M_{11}^{-1}Q(A_{11}M_{11}^{-1})$. If we started with $\sigma(M_{11}^{-1}A_{11}) \subset \langle 1, \lambda \rangle$, we would like to get $\sigma(M^{-1}A) \subset \langle 1, \lambda \rangle$ again. Thus, we need

$$\frac{(1+\lambda)^2}{4\lambda} \frac{1}{1-\gamma^2} \leq \lambda. \tag{21}$$

It can be verified that the solution of this inequality $\lambda > 1$ exists if $\gamma^2 < 3/4$. In such a case,

$$\lambda \geq \frac{1}{2\sqrt{1-\gamma^2}-1}. \tag{22}$$

Then

$$\sigma\left(\frac{1}{1-\gamma^2}\frac{y^T M^{-1} y}{y^T A^{-1} y}\right) \in \langle 1, \lambda \rangle$$

and

$$\hat{\kappa}\left(M^{-1}A\right) \leq \lambda.$$

Thus, we can prove recursively that for $\gamma^2 < 3/4$ the condition number for this type of the multilevel AML method (recursive two-by-two splitting of the first block of A and using (18)) is less than or equal to $\lambda = 1/(2\sqrt{1-\gamma^2}-1)$. This method will be called AML-W in section 6.

Remark 4.2 If a in (2) is linear in y and if y_i are uniformly or normally distributed, then under some conditions and for some splitting of the solution space, $\gamma^2 < 1/2$ and thus (21) can be fulfilled for $\lambda = 1 + \sqrt{2}$, see [30].

Remark 4.3 Due to the twofold solution of linear systems with M_{11} on every level, the AML-W scheme has a form of the W-cycle. It means that in every loop and on every level $k = 1, 2, \ldots, L$, there must be at least $2^{(L-k)}$ matrix-vector multiplications (of different dimensions according to the size of the problem on the level k) performed. Thus, the AML preconditioning can be optimal only if the reduction ratio between the dimensions of A and A_{11} is sufficient on every level. In such a case, the resulting condition number is uniformly bounded and the number of operations of a single step of the AML-W preconditioning can be linearly proportional to the number of elements of A, cf. [24]. On the other hand, within the AML-V method, only two systems with matrices corresponding to the A_{22} block must be solved on levels $1, \ldots, L-1$ and one system must be solved on the level L.

Remark 4.4 Let us summarize that applying the two-by-two block triangular preconditioner M, the resulting condition number of the matrix is bounded from above by $1/(1-\gamma^2)$, see (19). If we only use the two-by-two block diagonal preconditioner, the resulting condition number is bounded by $(1+\gamma)/(1-\gamma)$, see (20). If we use the hierarchical AML-W method with the polynomial Q of the degree one, and if $\gamma^2 < 3/4$ then the resulting condition number is bounded by $\lambda = 1/(2\sqrt{1-\gamma^2}-1)$. The resulting condition number is bounded by $1/((1-\gamma_1^2)\cdots(1-\gamma_{L-1}^2))$ for the AML-V method, where γ_k is the CBS constant on level k. Computational complexities of these methods depend on particular problems.

Remark 4.5 Apart from preconditioning, the AML approach also leads to reliable a posteriori estimates of the algebraic error, see, for example [1, Chapter 5.4]. Having γ sufficiently small, taking the residual r of some approximation of u and noticing that $\sigma(M^{-1}A) \subset \langle 1, \lambda \rangle$, we get

$$\frac{1}{\lambda} r^T M^{-1} r \leq r^T A^{-1} r \leq r^T M^{-1} r.$$

This together with $r^T A^{-1} r = e^T A e$ where $e = A^{-1} r$ is the algebraic error of u, means that the energy norm of the error can be estimated by the relatively easily computable term $r^T M^{-1} r$. This estimate is two-sided, guaranteed and its computational cost corresponds to evaluating $M^{-1} r$.

5 Block preconditioning and CBS constants for the stochastic Galerkin method for lognormally distributed parameters

In this section, we introduce the main results of this paper: (i) guaranteed upper bounds for the CBS constants for hierarchical two-by-two splitting of the matrix, and (ii) guaranteed upper bounds for the condition numbers for the block-diagonal preconditioning. Both results are obtained for the SGM matrix A for problems with lognormally distributed parameters defined by (8). We consider only tensor products of Hermite polynomials as bases for the approximation of stochastic components of the solution. The hierarchical splitting and the choice of blocks for the block-diagonal preconditioning is made only with respect to the degrees of approximation polynomials of stochastic variables. In the same way as in section 3, let us denote by M_{jk} and by $(M_{jk})_{rs}$ the blocks of M and their elements, respectively. The sizes of the blocks will always be clear from the context. In the following, we use the common consensus that $\prod_{i=k}^{m} c_i = 1$ for any $c_i \in \mathcal{R}$ whenever $m < k$.

Let $V = V_x \otimes V_y$ be the approximation space defined in section 3. According to Lemma 4.1 and to section 4.1, to obtain the uniform upper bound for the CBS constant, we can restrict our problem to a single reference element, which we call E. Let us also assume that all functions $a_i(x)$, $i = 0, 1, \ldots, N$, defined by (2) are piecewise constant, and, thus, they are constant on E. Let the span of all FE basis functions which are non-zero on E be denoted by $V_{x;E}$. Then to determine the CBS constant for some splitting of $V = V_x \otimes V_y$, we can only study the adequate splitting of $V_E = V_{x;E} \otimes V_y$. Let F_E be the number of basis functions of $V_{x,E}$ and let A_E be the matrix of the corresponding Neumann problem on E. Then A_E is positive semidefinite, cf. section 4.1.

Denote by $V_{y;P_1,\ldots,P_N}$ the span of polynomials $\Phi_j(y) = \phi_{j_1}(y_1) \ldots \phi_{j_N}(y_N)$ where $\phi_{j_i}(y_i)$ are Hermite orthogonal polynomials (defined by (9)) of the degree less than or equal to P_i, $i = 1, \ldots, N$. Let $W_{y;N;P_1,\ldots,P_N}$ be the span of polynomials $\Phi_j(y) = \phi_{j_1}(y_1) \ldots \phi_{j_N}(y_N)$ where the degree of $\phi_{j_i}(y_i)$ is less than or equal to P_i, $i = 1, \ldots, N-1$ and the degree of $\phi_{j_N}(y_N)$ is equal to P_N. We consider the splitting of the discretization space V into the direct sum of two subspaces

$$V = V_x \otimes V_{y;P_1,\ldots,P_N} = (V_x \otimes V_{y;P_1,\ldots,P_N-1}) \oplus (V_x \otimes W_{y;N;P_1,\ldots,P_N}). \tag{23}$$

It means that the set of $\Phi_j(y)$ is decomposed into two parts according to the degree of the polynomials of the "last" random variable y_N. The corresponding splitting of the discretization space V_E is then

$$V_E = V_{x;E} \otimes V_{y;P_1,\ldots,P_N} = (V_{x;E} \otimes V_{y;P_1,\ldots,P_N-1}) \oplus (V_{x;E} \otimes W_{y;N;P_1,\ldots,P_N}). \tag{24}$$

The dimensions of these spaces are

$$\begin{aligned}
\dim\left(V_{x;E} \otimes V_{y;P_1,\ldots,P_N}\right) &= F_E \cdot \prod_{k=1}^{N} (P_k + 1), \\
\dim\left(V_{x;E} \otimes V_{y;P_1,\ldots,P_N-1}\right) &= F_E \cdot P_N \cdot \prod_{k=1}^{N-1} (P_k + 1), \\
\dim\left(V_{x;E} \otimes W_{y;N;P_1,\ldots,P_N}\right) &= F_E \cdot \prod_{k=1}^{N-1} (P_k + 1).
\end{aligned}$$

Collecting the findings of previous sections and paragraphs, we may now approach to the main results of this paper.

Theorem 5.1 *Let us consider the problem (2)-(4) with piecewise constant functions $a_k(x)$, $k = 0, 1, \ldots, N$, which are constant on the interiors of all elements used for some FE discretization of the problem. Let the two-by-two splitting of the approximation space $V = V_x \otimes V_y$ be considered as a direct sum of the form (23). Then the CBS constant γ for this splitting and for the scalar product $\mathcal{A}(\cdot,\cdot)$ is bounded by*

$$\gamma^2 \leq \frac{\beta - 1}{\beta},$$

where

$$\beta = \sum_{k=0}^{P_N} \binom{P_N}{k} \frac{1}{k!} c^{2k}$$

and $c = \|a_N(x)\|_\infty$ on D.

Proof. According to Lemma 4.1, section 4.1 and the previous paragraphs of section 5, we can only study the corresponding Neumann problem on a single reference element E. Let the corresponding matrix A_E be split in accordance with (24). This means that we can evaluate (8), (14) and (15) only on E, where $a_i(x) = a_i$, $i = 0, 1, \ldots, N$, are constant (and possibly negative). Then $\tilde{G}_{ijk}(x) = \tilde{G}_{ijk}$ and $G_{jk}(x) = G_{jk}$ are constant on E as well, and

$$\tilde{G}_{ijk} = \mathrm{e}^{a_i^2/2} \sum_{m=0}^{\min(j,k)} \binom{j}{m}\binom{k}{m} \frac{m!}{\sqrt{j!k!}} a_i^{j+k-2m}, \tag{25}$$

$$G_{jk} = \prod_{i=1}^{N} \tilde{G}_{ij_ik_i}, \tag{26}$$

$j, k = 1, \ldots, M$. Let K_E be the stiffness matrix of the deterministic Neumann problem with the coefficient e^{a_0} on E. Then $A_E = G \otimes K_E$ and the elements of A_E are

$$(A_{E\,jk})_{rs} = G_{jk} \cdot (K_E)_{rs}, \tag{27}$$

$j, k = 1, \ldots, M$, $r, s = 1, \ldots, F_E$. Then, for instance, for $N = 1$ and $P = 2$ we have $\Phi_1(y) = \phi_0(y_1)$, $\Phi_2(y) = \phi_1(y_1)$, $\Phi_3(y) = \phi_2(y_1)$, and the matrix $A_E = A_E^{[N,P]} = A_E^{[1,2]}$ is

$$\begin{aligned} A_E^{[1,2]} &= \mathrm{e}^{a_1^2/2} \begin{pmatrix} K_E & a_1 K_E & (a_1^2/\sqrt{2}) K_E \\ a_1 K_E & (1+a_1^2) K_E & (\sqrt{2}a_1 + a_1^3/\sqrt{2}) K_E \\ (a_1^2/\sqrt{2}) K_E & (\sqrt{2}a_1 + a_1^3/\sqrt{2}) K_E & (1 + 2a_1^2 + a_1^4/2) K_E \end{pmatrix} \\ &= \mathrm{e}^{a_1^2/2} \begin{pmatrix} 1 & a_1 & a_1^2/\sqrt{2} \\ a_1 & 1+a_1^2 & \sqrt{2}a_1 + a_1^3/\sqrt{2} \\ a_1^2/\sqrt{2} & \sqrt{2}a_1 + a_1^3/\sqrt{2} & 1 + 2a_1^2 + a_1^4/2 \end{pmatrix} \otimes K_E, \end{aligned}$$

where $\otimes$ stands for the Kronecker product of matrices. If, for example, $N = 2$, $P_1 = P_2 = 2$ then according to the ordering of $\Phi_j(y)$ defined in section 3, we have $\Phi_1(y) = \phi_0(y_1)\phi_0(y_2)$, $\Phi_2(y) = \phi_1(y_1)\phi_0(y_2)$, $\Phi_3(y) = \phi_2(y_1)\phi_0(y_2)$, $\Phi_4(y) = \phi_0(y_1)\phi_1(y_2)$, $\ldots$, $\Phi_9(y) = \phi_2(y_1)\phi_2(y_2)$, and $A_E = A_E^{[N,P]} = A_E^{[2,2]}$ is

$$A_E^{[2,2]} = \mathrm{e}^{a_2^2/2} \begin{pmatrix} 1 & a_2 & a_2^2/\sqrt{2} \\ a_2 & 1+a_2^2 & \sqrt{2}a_2 + a_2^3/\sqrt{2} \\ a_2^2/\sqrt{2} & \sqrt{2}a_2 + a_2^3/\sqrt{2} & 1 + 2a_2^2 + a_2^4/2 \end{pmatrix} \otimes A_E^{[1,2]}.$$

As shown in this small example, due to (15) and (27), the matrix A_E has the Kronecker product structure. The numbering and the tensor product form of approximation polynomials $\Phi_j(y)$ yield the block structure of A_E, where each block is associated with a single degree of $\phi_{j_N}(y_N)$. Then A_E is built from $(P_N+1)\times(P_N+1)$ blocks where all of these blocks are equal up to a multiplicative

constant. The sizes of all of these blocks are $F_E \prod_{i=1}^{N-1}(P_i+1) \times F_E \prod_{i=1}^{N-1}(P_i+1)$. Then, finally, for estimating the CBS constant for the splitting (24), we may study only the splitting of a $(P_N+1)\times(P_N+1)$ matrix $\tilde{A}$, the entries of which are

$$(\tilde{A})_{j+1,k+1} = \mathrm{e}^{-a_N^2/2}\tilde{G}_{Njk}, \tag{28}$$

for $j,k = 0,\ldots,P_N$. We have

$$(\tilde{A})_{j+1,k+1} = \sum_{m=0}^{\min(j,k)} \binom{j}{m}\binom{k}{m}\frac{m!}{\sqrt{j!k!}}a_N^{j+k-2m} = \sum_{m=0}^{\min(j,k)} \binom{j}{m}\sqrt{\frac{m!}{j!}}a_N^{j-m}\binom{k}{m}\sqrt{\frac{m!}{k!}}a_N^{k-m}.$$

It can be noticed that $\tilde{A}$ is the sum of $(P_N+1)\times(P_N+1)$ rank-one matrices,

$$\tilde{A} = \sum_{m=0}^{P_N} \bar{v}_m \bar{v}_m^T,$$

where $\tilde{v}_m \in \mathcal{R}^{P_N+1}$, $(\bar{v}_m)_i = 0$ if $i < m$ and

$$(\bar{v}_m)_{i+1} = \binom{i}{m}\sqrt{\frac{m!}{i!}}a_N^{i-m} = \frac{\sqrt{i!}}{(i-m)!\sqrt{m!}}a_N^{i-m},$$

otherwise. For example, for $P_N = 2$ we have

$$\tilde{A} = \begin{pmatrix} 1 \\ a_N \\ \frac{1}{\sqrt{2}}a_N^2 \end{pmatrix}\begin{pmatrix} 1 & a_N & \frac{1}{\sqrt{2}}a_N^2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ \sqrt{2}a_N \end{pmatrix}\begin{pmatrix} 0 & 1 & \sqrt{2}a_N \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

Let us consider the matrix $\tilde{A}$ in the two-by-two block form

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{12}^T & \tilde{A}_{22} \end{pmatrix}, \tag{29}$$

where $\tilde{A}_{11}$ is of the type $P_N \times P_N$ and $\tilde{A}_{22}$ is a scalar. Let $u = (u_1,\ldots,u_{P_N}) \in \mathcal{R}^{P_N}$ and $w \in \mathcal{R}$ be arbitrary, and let us study the CBS constant γ in

$$(u^T\tilde{A}_{12}w)^2 \le \gamma^2 u^T\tilde{A}_{11}u\, w^T\tilde{A}_{22}w.$$

We have

$$\begin{aligned}
u^T\tilde{A}_{11}u &= \sum_{k=0}^{P_N}\left(\sum_{i=k+1}^{P_N} u_i(\tilde{v}_k)_i\right)^2 = \sum_{k=0}^{P_N-1}\left(\sum_{i=k+1}^{P_N} u_i(\tilde{v}_k)_i\right)^2, \\
w^T\tilde{A}_{22}w &= \sum_{k=0}^{P_N}(\tilde{v}_k)_{P_N+1}^2 w^2, \\
u^T\tilde{A}_{12}w &= \sum_{k=0}^{P_N}\sum_{i=k+1}^{P_N} u_i(\tilde{v}_k)_i(\tilde{v}_k)_{P_N+1}w = \sum_{k=0}^{P_N-1}\sum_{i=k+1}^{P_N} u_i(\tilde{v}_k)_i(\tilde{v}_k)_{P_N+1}w,
\end{aligned}$$

where the second and last equalities follow from $(\tilde{v}_{P_N})_i = 0$ for $i = 1,\ldots,P_N$. Then from Hölder's inequality

$$\begin{aligned}
(u^T\tilde{A}_{12}w)^2 &= \left(\sum_{k=0}^{P_N-1}\sum_{i=k+1}^{P_N} u_i(\tilde{v}_k)_i(\tilde{v}_k)_{P_N+1}w\right)^2 \\
&\le \sum_{k=0}^{P_N-1}\left(\sum_{i=k+1}^{P_N} u_i(\tilde{v}_k)_i\right)^2 \sum_{k=0}^{P_N-1}(\tilde{v}_k)_{P_N+1}^2 w^2 \\
&= \frac{\sum_{k=0}^{P_N-1}(\tilde{v}_k)_{P_N+1}^2}{\sum_{k=0}^{P_N}(\tilde{v}_k)_{P_N+1}^2}\, u^T\tilde{A}_{11}u\, w^T\tilde{A}_{22}w.
\end{aligned}$$

Finally, this yields

$$\begin{aligned}
\gamma^2 &\leq \frac{\sum_{k=0}^{P_N-1}(\tilde{v}_k)^2_{P_{N+1}}}{\sum_{k=0}^{P_N}(\tilde{v}_k)^2_{P_{N+1}}} = \frac{\sum_{k=0}^{P_N-1}\binom{P_N}{k}^2\frac{k!}{P_N!}a_N^{2P_N-2k}}{\sum_{k=0}^{P_N}\binom{P_N}{k}^2\frac{k!}{P_N!}a_N^{2P_N-2k}} = \frac{\sum_{k=0}^{P_N-1}\binom{P_N}{k}^2\frac{k!}{P_N!}a_1^{2P_N-2k}}{1+\sum_{k=0}^{P_N-1}\binom{P_N}{k}^2\frac{k!}{P_N!}a_N^{2P_N-2k}} \\
&= \frac{\sum_{k=0}^{P_N-1}\binom{P_N}{k}\frac{1}{(P_N-k)!}a_N^{2P_N-2k}}{1+\sum_{k=0}^{P_N-1}\binom{P_N}{k}\frac{1}{(P_N-k)!}a_N^{2P_N-2k}} = \frac{\sum_{k=0}^{P_N}\binom{P_N}{P_N-k}\frac{1}{(P_N-k)!}a_N^{2P_N-2k}-1}{\sum_{k=0}^{P_N}\binom{P_N}{P_N-k}\frac{1}{(P_N-k)!}a_N^{2P_N-2k}} \\
&= \frac{\sum_{k=0}^{P_N}\binom{P_N}{k}\frac{1}{k!}a_N^{2k}-1}{\sum_{k=0}^{P_N}\binom{P_N}{k}\frac{1}{k!}a_N^{2k}}. \qquad (30)
\end{aligned}$$

Note that the upper bound for γ^2 depends on $|a_N|$ and that it increases if $|a_N|$ grows. Then taking $c = \|a_N(x)\|_\infty$ instead of a_N in (30) completes the proof of the theorem.

Remark 5.2 The studied problem is isotropic and the coefficient a in (3) is a scalar function. We can treat similarly a more general problem, where a and a_k are $d \times d$ positive definite matrices. In such a case, instead of scalars $\exp(a_k y_k)$ we would consider matrices, but in the analysis we would deal with scalars $\exp(\lambda_{k_m} y_k)$, where λ_{k_m} is the mth eigenvalue of the matrix a_k. This analysis is rather technical and does not seem to bring anything new to the presented theory therefore we do not develop it in this direction.

Theorem 5.3 *Let us consider the problem (2)-(4) with piecewise constant functions $a_k(x)$, $k = 0, 1, \ldots, N$, which are constant on the interiors of all elements used for the FE discretization of the problem. Let the block-diagonal preconditioning be considered for the system of linear equations with the matrix A defined by (8), where the preconditioning diagonal blocks are of the sizes $s = F\prod_{k=1}^{N-1}(P_k+1)$ and correspond to the approximation subspaces $W_{y;N;P_1,\ldots,P_{N-1},k}$, where $k = 0, 1, \ldots, P_N$, respectively. Then the condition number $\hat{\kappa}(A_{\mathrm{BD}})$ of the resulting system is bounded by*

$$\hat{\kappa}(A_{\mathrm{BD}}) \leq \frac{\max_{z\in\langle -c,c\rangle}\lambda_{\max}(\tilde{G}_{N;d}(z))}{\min_{z\in\langle -c,c\rangle}\lambda_{\min}(\tilde{G}_{N;d}(z))}$$

where $c = \|a_N\|_\infty$ on D, $\lambda_{\min}(\tilde{G}_{N;d}(z))$ and $\lambda_{\max}(\tilde{G}_{N;d}(z))$ are the extremal eigenvalues of $\tilde{G}_{N;d}(z)$, which is the diagonally preconditioned matrix $\tilde{G}_N(z)$ of the size $(P_N+1)\times(P_N+1)$, where

$$(\tilde{G}_N(z))_{j+1,k+1} = \sum_{m=0}^{\min(j,k)}\binom{j}{m}\binom{k}{m}\frac{m!}{\sqrt{j!k!}}z^{j+k-2m}.$$

Proof. Let us denote by E_t the elements of D used by the FE method for the physical component of the solution, $D = \cup_{t=1}^{N_D}\bar{E}_t$. Then we can consider $a(u,u) = u^T A u$ as a sum of integrals over individual elements E_t and we can look at the structure of A from yet another perspective,

$$A = \sum_{t=1}^{N_D}\tilde{G}_N(z_t)\otimes\tilde{G}_{N-1}(z_t)\otimes\cdots\tilde{G}_1(z_t)\otimes K_{E_t}$$

where $z_t = a_N(x_t)$ and x_t is any point from the interior of E_t, $\tilde{G}_i(z)$ is of the type $(P_i+1)\times(P_i+1)$, K_{E_t} is of the type $F\times F$ and

$$\begin{aligned}
\left(\tilde{G}_i(z)\right)_{j+1,k+1} &= \mathrm{e}^{z^2/2}\sum_{m=0}^{\min(j,k)}\binom{j}{m}\binom{k}{m}\frac{m!}{\sqrt{j!k!}}z^{j+k-2m}, \quad i = 1,\ldots,N \\
(K_{E_t})_{rs} &= \int_{K_{E_t}}\mathrm{e}^{a_0(x_t)}\nabla\psi_r(x)\cdot\nabla\psi_s(x)\,\mathrm{d}x.
\end{aligned}$$

Note that the matrix $\tilde{G}_N(z)$ equals $\tilde{A}$ from the proof of Theorem 5.1 if $z = a_N$. Let $\tilde{G}_N^{\mathrm{diag}}(z)$ be the diagonal part of $\tilde{G}_N(z)$. Denote $c = \|a_N\|_\infty$. Note that $c = \max_{t=1,\ldots,N_D} |a_N(x_t)|$. Let us denote the extremal eigenvalues of the diagonally preconditioned matrix $\tilde{G}_N(z)$ by

$$\begin{aligned}\alpha_1 &= \min_{z\in\langle -c,c\rangle} \lambda_{\min}\left(\left(\tilde{G}_N^{\mathrm{diag}}(z)\right)^{-1}\tilde{G}_N(z)\right)\\ \alpha_2 &= \max_{z\in\langle -c,c\rangle} \lambda_{\max}\left(\left(\tilde{G}_N^{\mathrm{diag}}(z)\right)^{-1}\tilde{G}_N(z)\right).\end{aligned}$$

Since $\tilde{G}_N(z)$ is positive definite for $z \in \mathcal{R}$ and $c < \infty$, then $0 < \alpha_1 \le \alpha_2 < \infty$. We have

$$u^T A u = \sum_{t=1}^{N_D} u^T \tilde{G}_N(z_t) \otimes \tilde{G}_{N-1}(z_t) \otimes \cdots \tilde{G}_1(z_t) \otimes K_{E_t} u$$

and thus,

$$\begin{aligned}u^T A u &\ge \alpha_1 \sum_{t=1}^{N_D} u^T \tilde{G}_N^{\mathrm{diag}}(z_t) \otimes \tilde{G}_{N-1}(z_t) \otimes \cdots \tilde{G}_1(z_t) \otimes K_{E_t} u\\ u^T A u &\le \alpha_2 \sum_{t=1}^{N_D} u^T \tilde{G}_N^{\mathrm{diag}}(z_t) \otimes \tilde{G}_{N-1}(z_t) \otimes \cdots \tilde{G}_1(z_t) \otimes K_{E_t} u.\end{aligned}$$

Since for the block diagonal matrix M with blocks of A corresponding to the approximation subspaces $W_{y;N;P_1,\ldots,P_{N-1},k}$, $k = 0, 1, \ldots, P_N$ we have

$$u^T M u = \sum_{t=1}^{N_D} u^T \tilde{G}_N^{\mathrm{diag}}(z_t) \otimes \tilde{G}_{N-1}(z_t) \otimes \cdots \tilde{G}_1(z_t) \otimes K_{E_t} u,$$

we get

$$\alpha_1\, u^T M u \le u^T A u \le \alpha_2\, u^T M u$$

and therefore the resulting condition number is not greater than α_2/α_1, which completes the proof.

Remark 5.4 Theorem 5.3 transforms the question about the condition number for the block-diagonal preconditioning of the stochastic Galerkin matrix A to estimating the maximal and minimal eigenvalues of the $(P_N+1)\times(P_N+1)$ matrices $\tilde{G}_{N;d}(z) = \left(\tilde{G}_N^{\mathrm{diag}}(z)\right)^{-1}\tilde{G}_N(z)$ for $z \in \langle -c, c\rangle$ where $c = \max_{t=1,\ldots,N_D} |a_N(x_t)|$. We assume that $a_N(x)$ is constant on every element. Although we are not able to evaluate the extremal eigenvalues of $\tilde{G}_{N;d}(z)$ exactly, we can estimate them easily, for example, from graphical plots of these eigenvalues depending on $z \in \langle -c, c\rangle$. The elements of $\tilde{G}_{N;d}(z)$ are also given by

$$\begin{aligned}\left(\tilde{G}_{N;d}(z)\right)_{j+1,k+1} &= \sum_{m=0}^{\min(j,k)} \binom{j}{m}\binom{k}{m}\frac{m!}{\sqrt{j!k!}} z^{j+k-2m}\\ &= \frac{\int_{\mathcal{R}} \phi_i(y)\phi_j(y)\mathrm{e}^{zy+\frac{y^2}{2}}\,\mathrm{d}y}{\sqrt{\int_{\mathcal{R}} \phi_i^2(y)\; e^{zy+\frac{y^2}{2}}\,\mathrm{d}y \int_{\mathcal{R}} \phi_j^2(y)\mathrm{e}^{zy+\frac{y^2}{2}}\,\mathrm{d}y}},\end{aligned}$$

where ϕ_j are normalized Hermite orthogonal polynomials. Obviously, $\alpha_1(0) = \alpha_2(0) = 1$, because $\tilde{G}_{N;d}(0)$ is the identity matrix. Interestingly, for example, for $P_N \le 4$ the functions $\alpha_1(z) = \lambda_{\min}(\tilde{G}_{N;d}(z))$ and $\alpha_2(z) = \lambda_{\max}(\tilde{G}_{N;d}(z))$ are both even and they monotonically decrease and increase, respectively, on $\langle 0, c\rangle$. But a theoretical proof of

$$\begin{aligned}\alpha_1(c) &= \min_{z\in\langle -c,c\rangle}\lambda_{\min}(\tilde{G}_{N;d}(z))\\ \alpha_2(c) &= \max_{z\in\langle -c,c\rangle}\lambda_{\max}(\tilde{G}_{N;d}(z))\end{aligned}$$

for any P_N is not known to the author. Some examples comparing the exact values of condition numbers and their upper bounds obtained by Theorem 5.3 for block diagonal preconditioning of A for $P_N = 2, 3$ or 4 are presented in Table 4 in the next section.

6 Numerical experiments

In this section, we introduce some simple numerical experiments where we apply several preconditioning methods and compare the resulting condition numbers with their theoretical upper bounds derived in previous sections. We also compare some numerical values of the CBS constants γ with their theoretical upper bounds. We consider problem (4) with $d = 1$, $D = (0, 1)$, $f = 1$ and a given by (2) where and $a_0(x) = 1$, $a_k(x) = c_k \sin(k\pi x)$, where $c_k = 1/2$ in Tables 1 and 3, or $c_k = 1/k$ in Tables 2 and 4, or $c_k = 1$ in Table 4, $k = 1, \ldots, N$. The matrix A is defined by (8), where, $N = 1, 2, 3$ or 4, the number F of finite element basis functions is variable.

In Tables 1 and 2, we present the squares of the computed CBS constants and their theoretical upper bounds obtained from Theorem 5.1. The results in Table 1 illustrate that for different problems, the upper bound for γ depends only on P_N when $c_N = \|a_N\|_\infty$ does not change. The results of Table 2 show the influence of c_N together with P_N. The theoretical upper bounds for γ^2 are almost reached even if the examples are low-dimensional: $d = 1$ and $F = 10$. We show in Table 3 that γ^2 may change for different values of F but is still bounded by its theoretical upper bound. In general, the upper bounds for condition numbers depend neither on the number of deterministic degrees of freedom nor on the condition number of the corresponding deterministic problem, see also [29, 30].

Many types of preconditioning methods for SGM can be examined and compared. Especially, the methods for the log-transformed problem may be competitive, see, for example, [37]. In this paper we restrict ourselves to a small number of preconditioning methods which are of a similar nature as the introduced AML methods. We compare six types of methods which are either well known or described in section 4.2:

1. Diagonal preconditioning (D). This is the classical diagonal scaling of the matrix A: multiplying A by D^{-1}, where D is the diagonal of A. We expect that this method is weak and we introduce it just to show the conditioning of A (almost) unpreconditioned.

2. Block diagonal mean based preconditioning (BDm). The preconditioning block diagonal matrix is composed of $M = \prod_{k=1}^{N}(P_k + 1)$ diagonal blocks which are identical and equal to the first diagonal block of A of the size F. The complexity of one step of this method is the same as solving M problems of the sizes F with the same matrix but with different right hand sides. The BDm preconditioning can be parallelized.

3. Block diagonal preconditioning (BD). We use $P_N + 1$ diagonal blocks of A of sizes $s = F\prod_{k=1}^{N-1}(P_k + 1)$. The complexity of one step of this method is thus given by solving $P_N + 1$ problems of the size s. This method can be parallelized. Note that the blocks are larger than those in the BDm method. The blocks are different in the case of lognormally distributed a, as opposed to the case where a depends linearly on random variables, where the diagonal blocks are the same. The upper bounds for condition numbers are given by Theorem 5.3.

4. Two-by-two block preconditioning (B2). The preconditioning matrix is composed of two diagonal blocks of A of sizes sP_N and s, respectively. Due to the large size of the first block, this method is rather demanding and we introduce it mainly for theoretical purposes. The resulting condition number is bounded by $\hat{\kappa}_{\mathrm{B2}} \leq (1+\gamma)/(1-\gamma)$, where γ is the CBS constant for the corresponding splitting of A defined by (23), see (20) in section 4.2.

5. Algebraic multilevel V-cycle (AML-V). We use P_N+1 levels of hierarchy with respect to the polynomial degree of the stochastic variable y_N. In one step of this method, $2P_N-1$ systems of equations of the size s are solved. The resulting condition number can be estimated by $\hat{\kappa}_V \leq 1/((1-\gamma_1^2)\cdots(1-\gamma_{P_N}^2))$, see section 4.2.

6. Algebraic multilevel W-cycle (AML-W). We use P_N+1 levels of hierarchy with respect to the polynomial degree of the stochastic variable y_N. In one step of this method one solves approximately (depending on the implementation of the recursions for very low levels of hierarchy) 2^{P_N} sets of linear equations of sizes s, see section 4.2 or [24]. The resulting condition number can be estimated by $\hat{\kappa}_W = \lambda = 1/(2\sqrt{1-\gamma^2}-1)$ if Q is of the degree one and $\gamma^2 < 3/4$, see (22). In our experiments, we only use the degree of Q equal to one, and thus, we do not apply this method to the case $\gamma^2 \geq 3/4$. If the CBS constants differ on different levels of hierarchy, a more accurate estimate for λ is possible, but the evaluation is more involved. We do not treat this case in this paper.

In Table 4, the condition numbers, their theoretical upper bounds and the numbers of steps of the CG method are shown for four numerical experiments for the problem introduced at the beginning of this section. In all experiments we use $F = 100$ and $P_k = 2, 3$ or 4 for $k = 1, \ldots, N$. In experiments (a) and (b) we set $N = 1$, and in (c) and (d) we set $N = 3$. In (a) we set $c_1 = 1/3$ while in (b) we set $c_1 = 1$. In (c) we set $c_k = 1/k$, while in (d) we set $c_k = 1$, for $k = 1, 2, 3$. We can see that if c_N or P_N increase, the effect of all preconditioning methods gets worse although with different rates. Since we use the splitting with respect to polynomials of the variable y_N only, it is natural that only methods D and BDm are influenced by the number of stochastic variables N. Methods BD and BDm are almost equivalent for moderate c_k, but they differ, if c_k increases. As recalled in section 4.2, the W-cycle AML algorithm can be applied only if $\gamma^2 < 3/4$, therefore, the AML-W is not applied if $\gamma^2 \geq 3/4$ in our experiments. Interestingly, since γ changes for different levels of hierarchy, namely, it is smaller for lower levels, the condition numbers of AML-V and AML-W are comparable in our experiments. However, the upper bounds for condition numbers overestimate the true values for AML-V in some experiments.

Taking into account the computational effort, we can observe that more work means better preconditioning in general, with the following exception: in the BD method P_N+1 systems of the size s are solved in every preconditioning step, while in the AML-V method $2P_N+1$ systems of the same sizes s are solved in every step. In Table 5 we compare the overall numbers of systems of the size s which must be solved before reaching the prescribed relative error 1E-8. It seems that the AML-V method is slightly more efficient than the BD method and the difference becomes more remarkable if the conditioning of the matrix A gets worse. On the other hand, parallelization can be better applied to the BD than to the AML-V method.

Table 1: The CBS constants γ^2 obtained for the block splitting of A defined by (23), and their theoretical upper bounds obtained for $a_0(x) = 1$, $c_k = \|a_k\|_\infty = 1/2$, $F = 10$, $P_1 = P_2 = P_3 = P_4 = 1, 2, 3, 4$ or 5, and $N = 1, 2, 3$ or 4.

$P_k =$	1	2	3	4	5
A:					
$N = 1$	0.1965	0.3417	0.4523	0.5387	0.6074
$N = 2$	0.1874	0.3283	0.4370	0.5229	0.5918
$N = 3$	0.1871	0.3274	0.4357	0.5210	0.5896
$N = 4$	0.1938	0.3406	0.4521	0.5388	0.6077
upper bounds:					
$N = 1, 2, \ldots$	0.2000	0.3469	0.4584	0.5451	0.6138

Table 2: The CBS constants γ^2 obtained for the block splitting of A defined by (23) and their theoretical upper bounds obtained for $a_0(x) = 1$, $c_k = \|a_k\|_\infty = 1/k$, $F = 10$, $P_1 = P_2 = P_3 = P_4 = 1, 2, 3, 4$ or 5, and $N = 1, 2, 3$ or 4.

$P_k =$	1	2	3	4	5
A:					
$N = 1$	0.4944	0.7088	0.8190	0.8815	0.9195
$N = 2$	0.1878	0.3294	0.4390	0.5255	0.5949
$N = 3$	0.0928	0.1734	0.2463	0.3126	0.3715
$N = 4$	0.0567	0.1100	0.1589	0.2039	0.2454
upper bounds:					
$N = 1$	0.5000	0.7143	0.8235	0.8852	0.9224
$N = 2$	0.2000	0.3469	0.4584	0.5351	0.6138
$N = 3$	0.1000	0.1859	0.2604	0.3254	0.3826
$N = 4$	0.0588	0.1127	0.1621	0.2075	0.2495

Table 3: The CBS constants γ^2 obtained for the block splitting of A defined by (23) for $N = 3$, $P_1 = P_2 = P_3 = 2$, $c_3 = \|a_3\|_\infty = 1/2$, and $F = 5, 10, 25, 50$ and 100. The theoretical upper bound for γ^2 is 0.3469 (see Table 1).

$F =$	5	10	25	50	100
A:	0.2050	0.3274	0.3461	0.3469	0.3467

7 Discussion

We study block preconditioning methods for SGM for elliptic problems with lognormally distributed coefficients. We focus especially on AML methods and on block-diagonal preconditioning methods. The physical part of the solution is approximated by FE functions and the stochastic part is approximated by the tensor products of polynomials. Both preconditioning methods use only a (hierarchical) splitting of the approximation spaces of the stochastic part of the solution. We prove guaranteed upper bounds for the resulting condition numbers. Some of these bounds are based on the strengthened CBS constants. We introduce a methodology for obtaining guaranteed upper bounds for these constants. Let us recall that having these bounds, we can also use the AML approach to obtain guaranteed two-sided a posteriori estimates of the algebraic error, which can be employed, for example, in adaptive algorithms. Our estimates of the CBS constant do not depend on the type of FE used for the physical part of the solution and they hold true if the numbers of stochastic degrees of freedom are different in different parts of the physical domain. The main limitation of this paper is that a proof of the upper bound for the CBS constant for complete polynomials is missing. Preliminary numerical experiments made by the author indicate that the upper bounds for the CBS constants obtained for complete polynomials of the total order P are greater than the bound for tensor product polynomials up to the degree $P_k = P$. However, no theoretical estimates have been available yet.

It should be mentioned that there is probably no general strategy to be used for the preconditioning of SGM. Practical problems may vary significantly in both physical and stochastic parts of the input data and therefore different (combinations of) preconditioning methods may lead to the best efficiency for particular problems. In our paper, we demonstrate that block preconditioners deteriorate with different rates when the coefficients of variation and/or degrees of approximation polynomials grow. But the outcomes of this paper may help to decide on a theoretical basis

Table 4: Condition numbers ($\hat{\kappa}$) and their theoretical upper bounds ($\hat{\kappa}_{\rm est}$), if they are available, and numbers of CG iterations (it) to get the relative error 1E-8 for $F = 100$, $N = 1$ or 3, and for different $c_k = \|a_k\|_\infty$, $P_k = 2, 3$ or 4 (numerical experiments (a)-(d)). The minus sign denotes that the number of steps of the CG method exceeds 1000 and the slash sign denotes that the W-cycle cannot be performed with Q of degree one because of $\gamma^2 \geq 0.75$ (cf. section 4.2).

	$\hat{\kappa} \leq \hat{\kappa}_{\rm est}$	it	$\hat{\kappa} \leq \hat{\kappa}_{\rm est}$	it	$\hat{\kappa} \leq \hat{\kappa}_{\rm est}$	it
$P_k =$	2		3		4	
(a) $N = 1$, $c_1 = 1/3$						
$\gamma^2 =$	0.19		0.26		0.32	
D	9.20E+03	191	1.22E+04	272	1.57E+04	363
BDm	3.16	15	4.71	18	6.66	21
BD	$3.12 \leq 3.12$	14	$4.65 \leq 4.65$	18	$6.58 \leq 6.58$	21
B2	$2.52 \leq 2.52$	11	$3.08 \leq 3.08$	11	$3.66 \leq 3.66$	10
AML-V	$1.23 \leq 1.36$	6	$1.39 \leq 1.85$	7	$1.57 \leq 2.74$	8
AML-W	$1.23 \leq 1.24$	6	$1.36 \leq 1.39$	7	$1.50 \leq 1.56$	7
(b) $N = 1$, $c_1 = 1$						
$\gamma^2 =$	0.71		0.82		0.89	
D	4.21E+04	279	1.03E+05	468	2.30E+05	723
BDm	28.20	48	90.71	75	250.61	125
BD	$22.64 \leq 22.65$	42	$70.69 \leq 70.73$	72	$191.43 \leq 191.54$	106
B2	$11.91 \leq 11.92$	29	$20.61 \leq 20.62$	35	$32.79 \leq 32.80$	41
AML-V	$4.00 \leq 7.00$	17	$8.60 \leq 39.67$	25	$18.54 \leq 345.43$	36
AML-W	$3.96 \leq 14.48$	17	/	/	/	/
(c) $N = 3$, $c_k = 1/k$						
$\gamma^2 =$	0.19		0.26		0.32	
D	1.53E+05	-	5.69E+05	-	1.80E+06	-
BDm	138.41	104	770.28	229	3.42E+03	454
BD	$3.12 \leq 3.12$	15	$4.65 \leq 4.65$	18	$6.57 \leq 6.58$	22
B2	$2.52 \leq 2.52$	12	$3.08 \leq 3.08$	12	$3.65 \leq 3.66$	11
AML-V	$1.23 \leq 1.36$	7	$1.38 \leq 1.85$	8	$1.57 \leq 2.74$	8
AML-W	$1.23 \leq 1.24$	7	$1.36 \leq 1.39$	7	$1.50 \leq 1.56$	7
(d) $N = 3$, $c_k = 1$						
$\gamma^2 =$	0.71		0.82		0.89	
D	1.04E+06	-	7.50E+06	-	5.85E+07	-
BDm	4.50E+03	537	8.49E+04	-	1.09E+06	-
BD	$22.62 \leq 22.65$	44	$70.60 \leq 70.73$	76	$191.11 \leq 191.54$	120
B2	$11.90 \leq 11.92$	30	$20.60 \leq 20.62$	37	$32.76 \leq 32.80$	44
AML-V	$4.00 \leq 7.00$	17	$8.60 \leq 39.67$	26	$18.51 \leq 345.43$	37
AML-W	$3.95 \leq 14.48$	17	/	/	/	/

Table 5: Numbers of sets of equations of the size s solved within the solution by the BD and AML-V methods, respectively, and ratios of these numbers for settings (a) - (d).

	(a)			(b)			(c)			(d)		
P_k:	2	3	4	2	3	4	2	3	4	2	3	4
BD	42	72	105	126	288	530	45	72	110	132	304	600
AML-V	30	49	72	85	175	324	35	56	72	85	182	333
BD / AML-V	1.4	1.5	1.5	1.5	1.6	1.7	1.3	1.3	1.5	1.6	1.7	1.8

which strategy of preconditioning can be acceptable. Our results of [30] justify the convergence of experiments of [33], while our new estimates may lead to explaining the numerical experiments of [32].

Acknowledgement. This work was supported by the ERC-CZ project LL1202.

References

[1] M. Ainsworth, J. T. Oden, *A Posteriori Error Estimation in Finite Element Analysis*, 2011, John Wiley & Sons.

[2] O. Axelsson, *Iterative Solution Methods*. Cambridge University Press, 1996.

[3] O. Axelsson, R. Blaheta, M. Neytcheva, Preconditioning for boundary value problems using elementwise Schur complements, SIAM Journal on Matrix Analysis and Applications 31, 2009, pp. 767-789.

[4] R. Blaheta, S. Margenov, M. Neytcheva, Uniform estimate of the constant in the strengthened CBS inequality for anisotropic nonconforming FEM systems, Numerical Linear Algebra with Applications 11, 2004, pp. 309-326, DOI: 10.1002/nla.350.

[5] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equation with random input data, SIAM Review 2, 2010, pp. 317-355.

[6] I. Babuška, R. Tempone, G. E. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, SIAM Journal on Numerical Analysis 42, 2004, pp. 800-825.

[7] A. Bespalov, C. E. Powell, D. Silvester, Energy norm a posteriori error estimation for parametric operator equations, SIAM Journal on Scientific Computing 36, 2014, pp. A339-A363.

[8] D. Bosq, *Linear Processes in Function Spaces, Theory and Applications*, Lecture Notes in Statistics 149, Springer, 2000.

[9] M. Brezina, A. Doostan, T. Manteuffel, S. McCormick, J. Ruge, Smoothed aggregation algebraic multigrid for stochastic PDE problems with layered materials, Numerical Linear Algebra with Applications 21, 2014, pp. 239-255.

[10] F. Brezzi, M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, 1991.

[11] M. K. Deb, I. M. Babuška, J. T. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques, Computer Methods in Applied Mechanics and Engineering 190, 2001, pp. 6359-6372, 2001.

[12] M. Eigel, C. J. Gittelson, Ch. Schwab, E. K. Zander, Adaptive stochastic Galerkin FEM, Computer Methods in Applied Mechanics and Engineering 270, 2014, pp. 247-269.

[13] V. Eijkhout, P. Vassilevski, The role of the strengthened C.B.S.-inequality in multi-level methods, SIAM Review 33, 1991, pp. 405-419.

[14] H. C. Elman, O. G. Ernst, D. P. O'Leary, M. Stewart, Efficient iterative algorithms for the stochastic finite element method with applications to acoustic scattering, Computer Methods in Applied Mechanics and Engineering 194, 2006, pp. 1037-1055.

[15] H. Elman, D. Furnival, Solving the stochastic steady-state diffusion problem using multigrid, IMA Journal of Numerical Analysis 27, 2007, pp. 675-688.

[16] O. G. Ernst, C. E. Powell, D. J. Silvester, E. Ullmann, Efficient solvers for a linear stochastic Galerkin mixed formulation of diffusion problems with random data, SIAM Journal on Scientific Computing 31, 2009, pp. 1424-1447.

[17] O. G. Ernst, E. Ullmann, Stochastic Galerkin matrices, SIAM Journal on Matrix Analysis and Applications 31, 2010, pp. 1848-1872.

[18] P. Frauenfelder, Ch. Schwab, R. A. Todor, Finite elements for elliptic problems with stochastic coefficients, Computer Methods in Applied Mechanics and Engineering 194, 2005, pp. 205-228.

[19] W. Gautschi, *Orthogonal Polynomials, Computation and Approximation*, Oxford University Press, 2004.

[20] R. Ghanem, The Nonlinear Gaussian Spectrum of Log-Normal Stochastic Processes and Variables, Journal of Applied Mechanics 66, 1999, pp. 964-973.

[21] R. G. Ghanem, P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, 1991 Springer Verlag, New York, 1991.

[22] M. B. Giles, Multilevel Monte Carlo Methods, Springer Proceedings in Mathematics & Statistics 65, 2013, pp. 83-103.

[23] C. Gittelson, Stochastic Galerkin discretization of the log-normal isotropic diffusion problem, Mathematical Models and Methods in Applied Sciences 20, 2010, pp. 237-263.

[24] J. Kraus, S. Margenov, *Robust Algebraic Multilevel Methods and Algorithms*, Radon Series on Computational and Applied Mthematics 5, RICAM, Walter de Gruyter, 2009, Germany.

[25] B. Lee, Parallel preconditioners and multigrid solvers for stochastic polynomial chaos discretizations of the diffusion equation at the large scale, Numerical Linear Algebra with Applications 23, 2016, pp. 5-36.

[26] O. P. Le Maître, O. M. Knio, B. J. Debusschere, H. N. Najm, R. G. Ghanem, A multigrid solver for two-dimensional stochastic diffusion equations, Computer Methods in Applied Mechanics and Engineering 192, 2003, pp. 4723-4744.

[27] H. G. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Computer Methods in Applied Mechanics and Engineering 194, 2005, pp. 1295-1331.

[28] C. E. Powell, H. C. Elman, Block-diagonal preconditioning for spectral stochastic finite-element systems, IMA Journal of Numerical Analysis 29, 2009, pp. 350-375

[29] I. Pultarová, Adaptive Algorithm for stochastic Galerkin method, Applications of Mathematics 60, 2015, pp. 551-571.

[30] I. Pultarová, Hierarchical preconditioning for the stochastic Galerkin method: upper bounds to the strengthened CBS constants, Computer and Mathematics with Applications 71, 2016, pp. 949-964.

[31] A. Ralston, *Základy numerické matematiky (A First Course in Numerical Analysis)*, Academia Praha, 1973 (in Czech).

[32] B. Sousedík, R. G. Ghanem, Truncated hierarchical preconditioning for the stochastic Galerkin FEM, International Journal for Uncertainty Quantification 4, 2014, pp. 333-348.

[33] B. Sousedík, R. G. Ghanem and E. T. Phipps, Hierarchical Schur complement preconditioner for the stochastic Galerkin methods, Numerical Linear Algebra with Applications 21, 2013, pp. 136-151.

[34] W. Subber, S. Loisel, Schwarz preconditioners for stochastic elliptic PDEs, Computer Methods in Applied Mechanics and Engineering 272, 2014, pp. 34-57.

[35] W. Subber, A. Sarkar, A domain decomposition method of stochastic PDEs: An iterative solution techniques using a two-level scalable preconditioner, Journal of Computational Physics 257 2014, pp. 298-317.

[36] E. Ullmann, A Kronecker product preconditioner for stochastic Galerkin finite element discretizations, SIAM Journal on Scientific Computing 32, 2010, pp. 923-946.

[37] E. Ullmann, H. C. Elman, O. G. Ernst, Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems, SIAM Journal on Scientific Computing 34, 2012, pp. A659-A682.

[38] E. Ullmann, C. E. Powell, Solving log-transformed random diffusion problems by stochastic Galerkin mixed finite element methods, SIAM/ASA Journal on Uncertainty Quantification 3, 2015, pp. 509-534.

[39] D. Xiu, *Numerical Methods for Stochastic Computations*, Princeton University Press, Princeton, New Jersey, 2010.

[40] D. Xiu, J. S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM Journal on Scientific Computing 27, 2005, pp. 1118-1139.

[41] D. Xiu, J. Shen, Efficient stochastic Galerkin methods for random diffusion equations, Journal of Computational Physics 228, 2009, pp. 266-281.

[42] B. Zheng, G. Lin, J. Xu, Block triangular preconditioning for stochastic Galerkin method, arXiv:1304.1755v1 [math.NA] 5 Apr 2013.