

# Sparse tensor discretization of elliptic sPDEs\*

M. Bieri, R. Andreev and C. Schwab

Research Report No. 2009-07  
February 2009

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

---

\*Research supported in part by the Swiss National Science Foundation under Grant No. 200021-120290/1

# Sparse Tensor Discretization of elliptic sPDEs \*

Marcel Bieri, Roman Andreev, Christoph Schwab<sup>†</sup>

February 11, 2009

## Abstract

We propose and analyze *sparse deterministic-stochastic tensor Galerkin finite element methods (sparse sGFEMs)* for the numerical solution of elliptic partial differential equations (PDEs) with random coefficients in a bounded physical domain  $D \subset \mathbb{R}^d$ . The sparse sGFEMs are based on a separation of stochastic and deterministic input variables by Karhunen-Loève expansion of the random diffusion coefficient. In tensor product sGFEM, the variational solution to the boundary value problem is approximated in tensor product finite element spaces  $V^\Gamma \otimes V^D$ , where  $V^\Gamma$  and  $V^D$  denote suitable finite dimensional subspaces of the stochastic and deterministic function spaces, respectively.

These approaches lead to sGFEM algorithms of complexity  $O(N_\Gamma N_D)$ , where  $N_\Gamma$  and  $N_D$  denote the number of stochastic and deterministic Degrees of Freedom (DoFs), respectively. Stochastic collocation algorithms proposed e.g. in [39] exhibit the same complexity (with smaller constant).

In this work, we use *tensor products of hierarchic sequences*  $V_1^\Gamma \subset V_2^\Gamma \subset \dots$ ,  $V_1^D \subset V_2^D \subset \dots$  of finite dimensional spaces to approximate the law of the random solution. The hierarchies of approximation spaces allows to define *sparse tensor product spaces*  $V_\ell^\Gamma \hat{\otimes} V_\ell^D$ ,  $\ell = 1, 2, \dots$  yielding an algorithm of  $O(N_\Gamma \log N_D + N_D \log N_\Gamma)$  work and memory.

We estimate the convergence rate of sGFEM for algebraic decay of the input random field's KL coefficients. We give a algorithm for an input adapted selection of deterministic and stochastic discretization spaces. The convergence rate (in terms of the total number of degrees of freedom) of the proposed method is superior to Monte-Carlo approximations.

Numerical examples illustrate the theoretical results and indicate superiority of the sparse tensor product discretization proposed here versus the full tensor product approach with Smolyak type subspaces  $V^\Gamma$  proposed elsewhere.

**Key words.** Stochastic partial differential equations, uncertainty quantification, stochastic finite element methods, multilevel approximations, sparse tensor products

**AMS subject classifications.** 35R60, 60H15, 65C20, 65N12, 65N15, 65N30

## 1 Introduction

### 1.1 Random field inputs to PDEs in engineering

Efficient numerical solution of partial differential equations (PDEs) with random field inputs is a key ingredient in Uncertainty Quantification (UQ) in engineering.

---

\*Research supported in part by the Swiss National Science Foundation under Grant No. 200021-120290/1

<sup>†</sup>Seminar for Applied Mathematics, ETH Zentrum, Zürich, Switzerland

Random field inputs are spatially inhomogeneous, in general non-stationary and non-ergodic, input data of partial differential equations, such as coefficients (e.g., in subsurface flow models) and loadings (e.g., in seismic analysis); however, also the problem domain of definition can be subject to uncertainty due to, e.g., inaccurate manufacturing processes or incomplete knowledge about the computational domain.

Development of efficient numerical solvers of PDEs with random field input requires advances in several areas of applied mathematics and numerical analysis:

- first, the *mathematical description of uncertain, spatially inhomogeneous inputs as random fields* is required. This is, in principle, well-known, see, e.g. [2, 35]; of particular interest are *parametric, deterministic representations of random field input data*: there, as in parametric statistics, one is interested in *parsimonious* parametric descriptions of these inputs. This is, however, not always possible and in certain cases (as, e.g., spatially inhomogeneous input data with small correlation length) a possibly countable number of parameters must be admitted to allow for descriptions of realistic input fields,
- second, the *mathematical formulation of PDEs with random field input* leads to stochastic partial differential equations (sPDEs),
- third, development of efficient, deterministic numerical solvers of PDEs with random field input and, more generally, the
- *efficient numerical sensitivity analysis of parametrized deterministic PDEs on possibly high-dimensional parameter spaces*.

A key requirement in e.g. risk-analysis for engineering systems modelled by PDEs with random field input data is the knowledge of the joint probability density functions (joint pdf's) of the solution (which is, in turn, also a random field). Note carefully that, contrary to the belief of some, incomplete knowledge of input data does *not* imply sufficiency of low solver accuracy of the “forward” numerical solver. To the contrary, in our view highly accurate numerical solution of PDEs with random field input is required for the quantification of uncertainty. Consider, for example, predictions of so-called *extremal events*: here, sensitivity of the “tail-behavior” of the random solution’s joint pdf’s in terms of input data needs to be quantified. This, in turn, requires highly accurate computations of the parametric dependence of the tails of solution pdf’s on the random field input data which is parametrized on a possibly high-dimensional parametric domain.

Numerical solution strategies for PDEs in a physical domain  $D$  with random field inputs can be grouped into two broad classes.

First, so-called *nonintrusive solvers*: here, existing deterministic solvers of the PDE of interest are used *without any modification* as a building block in an outer “UQ-loop” where some form of sampling of the random parameter space is used to generate a set of particular input realizations to be processed by the deterministic forward PDE solver, leading to corresponding outputs of the random solution from which the joint pdf’s are recovered. In this class, we find the Monte-Carlo (MC) and quasi-Monte-Carlo (QMC) sampling strategies of the random inputs as well as certain high-order “polynomial chaos” methods which are based on spectral representations of the random fields’ in- and outputs. Due to the generally slow convergence of MC and QMC sampling strategies and the generally high computational cost of each “sample” (amounting to one deterministic PDE solve in, possibly, spatial dimension 3), the computational efficiency of these approaches is low. Throughout the paper,  $N_\Gamma$  denotes the number of “samples” from the probability domain and  $N_D$  denotes the number of deterministic degrees of freedom used in the physical domain  $D$  by the deterministic ‘forward’ solver.

Higher than MC accuracy per “sample” was achieved recently by *stochastic collocation* (see [40, 38, 37, 4, 24, 23] and the references there); in both MC/QMC and stochastic collocation, however, the random solutions’ approximate law is parametrically represented by  $N_\Gamma$  times  $N_D$  degrees of freedom (DoFs). Due to the high cost of one deterministic PDE solve, attention has been focused in recent years on minimizing the number  $N_\Gamma$  of deterministic solves by judicious choice of collocation points; see, e.g. [5, 15, 20, 21, 38, 37, 40] and the references there. In [24, 23], *nonuniformly distributed* sets of collocation points of low cardinality  $N_\Gamma$  obtained by a Smolyak-style tensorization of univariate sequences of interpolation points were proposed and found superior to straightforward tensorization as in [40, 38, 37, 4].

For complex engineering systems where the smallest  $N_D$  necessary for resolution of all spatial length scales of interest is close to the limit of available computing resources, nonintrusive computational approaches to UQ with a large number  $N_\Omega$  of “samples” are today beyond reach.

In the present paper, we consider therefore a novel, *intrusive* class of solvers: we show how (by now well established) *adaptive multilevel or multiresolution solvers* of deterministic elliptic PDE’s can be intertwined with *adaptive hierarchic parametric representations* of the solution random fields. We demonstrate that in this fashion UQ algorithms of complexity  $O(N_\Omega \log N_D + N_D \log N_\Omega)$  can be obtained.

The UQ algorithms proposed and analyzed here have several characteristic features: first, they are *data-compressive* at all stages of the algorithm. Specifically, uncertainty in the numerical solution’s contributions from different scales of discretization (i.e., from different levels of its multilevel representation) is resolved only to the extent necessary to maintain overall accuracy.

Second, a spectral analysis (i.e. Karhunen-Loève expansion) of the input random fields sharp a-priori estimates are used to determine, by an algorithm inspired by number theoretic results [9] on integer factorizations, a-priori the essential components of the unknown random solution to be computed to achieve a prescribed error threshold  $\varepsilon > 0$  with minimal work.

## 1.2 Model Problem

As in [4, 5, 17, 7, 15, 20, 38, 37, 40], our model problem is an elliptic diffusion problem with stochastic diffusion coefficient  $a(\omega, \mathbf{x})$ . We assume that the physical domain  $D \subset \mathbb{R}^d$  is a known, bounded open set with Lipschitz boundary  $\partial D$  where  $d$  is a positive integer.

Denote by  $(\Omega, \mathcal{B}, P)$  a probability space where  $\Omega$  denotes the outcomes,  $\mathcal{B} \subset 2^\Omega$  the sigma algebra of possible events and  $P : \mathcal{B} \rightarrow [0, 1]$  a probability measure.

In order to guarantee existence, uniqueness and well-posedness of the model problem introduced below, we have to make the following assumption on the diffusion coefficient  $a(\omega, \mathbf{x})$ , following [5].

**Assumption 1.1.** Let  $a \in L^\infty(D \times \Omega)$  be strictly positive, with lower and upper bound

$$P \left\{ \omega \in \Omega : a_{\min} \leq \operatorname{ess\,inf}_{\mathbf{x} \in D} a(\omega, \mathbf{x}) \wedge \operatorname{ess\,sup}_{\mathbf{x} \in D} a(\omega, \mathbf{x}) \leq a_{\max} \right\} = 1 \quad (1.1)$$

where the essential infimum and supremum are taken with respect to the Lebesgue measure in  $D \subset \mathbb{R}^d$ .

The model problem under consideration here can then be written as follows: given  $f \in L^2(D)$  and  $a \in L^\infty(D \times \Omega)$  satisfying Assumption 1.1, find  $u$  such that

$$\begin{cases} -\operatorname{div}(a(\omega, \mathbf{x}) \nabla u(\omega, \mathbf{x})) = f(\mathbf{x}) & \text{in } D, \\ u(\omega, \mathbf{x})|_{\mathbf{x} \in \partial D} = 0, \end{cases} \quad (1.2)$$

for  $P$ -a.e.  $\omega \in \Omega$ , where the diffusion coefficient  $a(\omega, \mathbf{x})$  and the solution  $u(\omega, \mathbf{x})$  are random fields in the spatial domain  $D$ .

Given a deterministic source term  $f(\mathbf{x}) \in L^2(D)$ , one is often interested in deterministic statistics of  $u(\omega, \mathbf{x})$  rather than in the random solution itself.

In order to compute statistical moments of the random solution  $u(\mathbf{x}, \omega)$  as well as, e.g., probabilities of interest in UQ, MC and QMC approaches have been proposed for the numerical solution of (1.2), see, e.g., [36] for a convergence analysis.

MC and QMC methods, however, suffer from the need to restart the (expensive) sampling process for essentially every statistical quantity to be computed.

To remedy this, in recent years several authors have proposed the computation of parametric representations of the joint pdf's of the random solution  $u(\omega, \mathbf{x})$  of (1.2). Here, a *parsimonious representation* of these joint pdf's is desirable. There are, roughly speaking, two main classes of methods to achieve this: first, perturbation methods [27, 28] where the random solution's fluctuations about the (unknown) mean field are approximated by sparse tensor product approximations, and, second, representations of the unknown random solution in terms of  $V^D$ -valued polynomials of the (independent) sources of randomness of the input data, where  $V^D$  denotes a Banach space in which the deterministic version of (1.2) is well-posed.

In this paper, we shall focus on the latter approach and analyze the convergence rates of projections of random solution's joint pdf's onto sparse tensor products of spaces spanned by polynomials of the (independent) sources of randomness of the input data and hierarchic multilevel subspaces of  $V^D$ . The projection is mean-square in probability and the  $V^D$ -projection in physical space; we term this projection accordingly "sparse tensor product stochastic Galerkin FEM".

In [7] we presented an efficient numerical sGFEM algorithm to solve (1.2) by a stochastic Galerkin FEM based on a sparse polynomial chaos (SPC) expansion of the solution. The work [7] revealed that SPC is a very effective and accurate tool to parametrically approximate the solution's random behavior while using considerably less degrees of freedom than e.g. generalized polynomial chaos (gPC) [20, 40] or multi-element gPC (ME-gPC) [38, 37] approximations. In the present paper, we will follow a best  $N$ -term PC approach and derive, in addition, a *sparse tensor product construction* between stochastic and deterministic variables. This will allow us to treat the case of algebraically decaying Karhunen-Loève coefficients, see Section 2.3 ahead, in an efficient fashion.

### 1.3 Outline

The paper is structured as follows. In Section 2 we will formulate the stochastic Galerkin FEM and introduce the main notations used throughout this paper. We also derive the sparse tensor product formulation of the sGFEM and present a first theoretical result related to the complexity of the sparse tensor product space.

In Section 3 we present hierarchic sequences of stochastic and deterministic approximation spaces and prove algebraic convergence rates for the sparse tensor discretizations in terms of the total number of degrees of freedom under the assumption of algebraically decaying Karhunen-Loève coefficients.

An algorithm to find suitable stochastic discretization spaces will then be presented in Section 4. Issues related to the implementation and complexity of this algorithm are discussed.

In Section 5 we will present numerical examples comparing sparse sGFEM to 'full' sGFEM and give some concluding remarks.

The Appendix A finally contains the proofs of the results stated in Section 3.

## 2 Galerkin Formulation

We recapitulate a variational formulation of the model problem (1.2) which will be the basis for the sparse tensor Galerkin discretization.

### 2.1 Variational formulation

The computational stochastic Galerkin FEM is based on a *variational formulation* of (1.2) in both, physical as well as in probability space. Specifically, denote by  $V^D$  a Banach space in which the deterministic problems in (1.2) are well-posed. For example, for the Dirichlet problem for the diffusion equation,  $V^D = H_0^1(D)$ . By  $L_P^2(\Omega; V^D)$  we denote the Bochner space of mappings  $\Omega \ni \omega \rightarrow u(\omega, \cdot) \in V^D$ . Note that  $L_P^2(\Omega; V^D) \simeq L_P^2(\Omega) \otimes V^D$  where  $L_P^2(\Omega) := \{\xi(\omega) \mid \int_{\Omega} \xi^2(\omega) dP(\omega) < \infty\}$ .

To derive the variational formulation, we multiply (1.2) with a test function  $v$ , integrate by parts over the physical domain  $D$  and take expectations on both sides. This yields the *variational formulation* of (1.2): find  $u \in L_P^2(\Omega; V^D)$  such that for all  $v \in L_P^2(\Omega; V^D)$

$$b(u, v) := \mathbb{E} \left[ \int_D a(\omega, \mathbf{x}) \nabla u(\omega, \mathbf{x}) \cdot \nabla v(\omega, \mathbf{x}) d\mathbf{x} \right] = \mathbb{E} \left[ \int_D f(\mathbf{x}) v(\omega, \mathbf{x}) d\mathbf{x} \right]. \quad (2.1)$$

From Assumption 1.1 it follows that the bilinear form  $b(\cdot, \cdot)$  is coercive and continuous, i.e. there exists a constant  $\gamma > 0$  such that for all  $u, v \in L_P^2(\Omega; H_0^1(D))$  holds

$$b(v, v) \geq \gamma \|v\|_{L_P^2(\Omega, H_0^1(D))} \quad \text{and} \quad |b(u, v)| \leq \gamma^{-1} \|u\|_{L_P^2(\Omega, H_0^1(D))} \|v\|_{L_P^2(\Omega, H_0^1(D))}. \quad (2.2)$$

For every  $f \in H^{-1}(D) = (V^D)^*$ , the unique solvability of (2.1) follows from the Lax-Milgram Lemma.

### 2.2 Continuous Dependence on Data

A main requirement for the stable numerical approximation of (1.2) is continuous dependence of the solution  $u$  on the random input data  $a(\omega, \mathbf{x})$ : let  $a(\omega, \mathbf{x})$  and  $\tilde{a}(\omega, \mathbf{x})$  be two random diffusion coefficients satisfying (1.1), and denote by  $u, \tilde{u} \in L_P^2(\Omega; H_0^1(D))$  the corresponding unique weak solutions. By a Strang-type perturbation argument, it holds

$$\|u - \tilde{u}\|_{L_P^2(\Omega, H_0^1(D))} \leq C \|a - \tilde{a}\|_{L_P^\infty(\Omega, L^\infty(D))} \|f\|_{H^{-1}(D)}. \quad (2.3)$$

### 2.3 Separation of stochastic and deterministic variables

To formulate the stochastic finite element method we separate stochastic and deterministic variables in the diffusion coefficient  $a(\omega, x)$ . One way to achieve this is the so-called Karhunen-Loève expansion defined below. We assume that the known information about the diffusion coefficient includes its mean field

$$\mathbb{E}_a(\mathbf{x}) = \int_{\Omega} a(\omega, \mathbf{x}) dP(\omega), \quad \mathbf{x} \in D, \quad (2.4)$$

and covariance

$$C_a(\mathbf{x}, \mathbf{x}') = \int_{\Omega} (a(\omega, \mathbf{x}) - \mathbb{E}_a(\mathbf{x}))(a(\omega, \mathbf{x}') - \mathbb{E}_a(\mathbf{x}')) dP(\omega), \quad \mathbf{x}, \mathbf{x}' \in D. \quad (2.5)$$

**Definition 2.1.** A covariance function  $C_a(\mathbf{x}, \mathbf{x}') \in L^2(D \times D)$  given by (2.5) is said to be *admissible* if it is symmetric and positive definite in the sense that

$$0 \leq \sum_{k=1}^n \sum_{j=1}^n a_k C_a(\mathbf{x}_k, \mathbf{x}_j) \bar{a}_j \quad \forall \mathbf{x}_k, \mathbf{x}_j \in D, a_k, a_j \in \mathbb{C} \quad (2.6)$$

We refer to [26] for basic results on positive definite functions and examples.

The covariance operator of a random permeability  $a(\omega, \mathbf{x}) \in L^2(\Omega \times D)$  is

$$\mathcal{C}_a : L^2(D) \longrightarrow L^2(D), \quad (\mathcal{C}_a u)(\mathbf{x}) := \int_D C_a(\mathbf{x}, \mathbf{x}') u(\mathbf{x}') d\mathbf{x}'. \quad (2.7)$$

Given an admissible covariance function  $V_a(\mathbf{x}, \mathbf{x}')$  in the sense of Definition 2.1, the associated covariance operator  $\mathcal{C}_a$  is a symmetric, non-negative and compact integral operator from  $L^2(D)$  to  $L^2(D)$ .

It therefore has a countable sequence of eigenpairs  $(\lambda_m, \varphi_m)_{m \geq 1}$

$$\mathcal{C}_a \varphi_m = \lambda_m \varphi_m \quad m = 1, 2, \dots \quad (2.8)$$

where the sequence (real and positive) of KL-eigenvalues  $\lambda_m$  is enumerated with decreasing magnitude and is either finite or tends to zero as  $m \rightarrow \infty$ , i.e.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq 0 \text{ ( with multiplicity counted ),} \quad \lambda_m \rightarrow 0 \quad \text{as } m \rightarrow \infty. \quad (2.9)$$

The KL eigenfunctions  $\varphi_m(\mathbf{x})$  are assumed to be  $L^2(D)$ -orthonormal, i.e.

$$\int_D \varphi_m(\mathbf{x}) \varphi_n(\mathbf{x}) d\mathbf{x} = \delta_{mn}, \quad m, n = 1, 2, \dots \quad (2.10)$$

**Definition 2.2 (Karhunen-Loève expansion).** The Karhunen-Loève (KL) expansion of a random field  $a(\omega, \mathbf{x})$  with finite mean (2.4) and covariance (2.5) which is admissible in the sense of Definition 2.1 is given by

$$a(\omega, \mathbf{x}) = \mathbb{E}_a(\mathbf{x}) + \sum_{m \geq 1} \sqrt{\lambda_m} \varphi_m(\mathbf{x}) Y_m(\omega). \quad (2.11)$$

The family of random variables  $(Y_m)_{m \geq 1}$  is given by

$$Y_m(\omega) = \frac{1}{\sqrt{\lambda_m}} \int_D (a(\omega, \mathbf{x}) - \mathbb{E}_a(\mathbf{x})) \varphi_m(\mathbf{x}) d\mathbf{x} : \Omega \rightarrow \mathbb{R}. \quad (2.12)$$

One immediately verifies that

$$\mathbb{E}[Y_m] = 0, \quad \mathbb{E}[Y_m \cdot Y_n] = \delta_{mn}, \quad \forall m, n \geq 1, \quad (2.13)$$

i.e. the  $Y_m(\omega)$  are centered with unit variance and pairwise uncorrelated.

The following Proposition provides a bound on the eigenvalues and eigenfunctions of (2.8) related to the regularity of the covariance operator  $\mathcal{C}_a(\mathbf{x}, \mathbf{x}')$ .

**Proposition 2.3.** *Assume  $C_a \in H^{t,t}(D \times D) := H^t(D) \otimes L^2(D) \cap L^2(D) \otimes H^t(D)$  with  $t > d/2$ . Then the eigenvalues of (2.8) admit the bound*

$$\lambda_m \leq C m^{-t/d} \quad m \geq 1 \quad (2.14)$$

with a constant  $C > 0$  independent of  $m$ . Moreover,  $\varphi_m \in H^t(D)$  and for every  $\varepsilon \in (0, t - d/2]$  there exists a constant  $C > 0$ , depending on  $\varepsilon, d$  but not on  $m$ , such that

$$\|\varphi_m\|_{L^\infty(D)} \leq C \lambda_m^{-(d/2+\varepsilon)/t}, \quad m = 1, 2, 3, \dots \quad (2.15)$$

*Proof.* The estimate on the eigenvalues follows directly from [29], Proposition 2.21. It was also shown in [29], Proposition A.2 that  $\varphi_m \in H^t(D)$ .

The Sobolev embedding theorem [1] for all  $t^* \in (d/2, t]$  yields

$$\|\varphi_m\|_{C^0(\bar{D})} \leq C \|\varphi_m\|_{H^{t^*}(D)} \quad \forall m \geq 1 \quad (2.16)$$

with a positive constant  $C$  depending on  $t^*, d$ . It remains to bound the norm on the right hand side of the inequality. It is well-known that fractional order Sobolev spaces  $H^{t^*}(D)$  can be obtained by complex interpolation [34] between  $L^2(D)$  and  $H^t(D)$  for any  $t^* < t$  and we therefore have the Riesz-Thorin-type inequality (see [34] Paragraph 1.9.3)

$$\|u\|_{H^{t^*}(D)} \leq \|u\|_{H^t(D)}^{t^*/t} \|u\|_{L^2(D)}^{1-t^*/t} \quad (2.17)$$

for any function  $u \in H^t(D)$ . Choosing now  $u = \varphi_m$  we have, due to the  $L^2$ -normalization that

$$\|\varphi_m\|_{H^{t^*}(D)} \leq \|\varphi_m\|_{H^t(D)}^{t^*/t} \quad (2.18)$$

From the eigenvalue equation

$$\varphi_m(\mathbf{x}) = \frac{1}{\lambda_m} \int_D C_a(\mathbf{x}, \mathbf{x}') \varphi_m(\mathbf{x}') d\mathbf{x}'$$

it follows by differentiating and using the Cauchy-Schwartz inequality

$$\|\partial^\alpha \varphi_m\|_{L^\infty(D)} \leq C(\alpha) \lambda_m^{-1} \quad (2.19)$$

Combining now (2.19), (2.18) and (2.16) yields

$$\|\varphi_m\|_{L^\infty(D)} \leq C \lambda_m^{-t^*/t} \quad (2.20)$$

which completes the proof.  $\square$

*Remark 2.4.* In the case where the Karhunen-Loève eigenvalues decay exponentially, i.e. like  $\lambda_m \lesssim e^{-cm^{1/d}}$ , we have from Theorem 2.24 that

$$\|\varphi_m\|_{L^\infty(D)} \leq C \lambda_m^{-\delta} \quad (2.21)$$

for any  $\delta > 0$ . This coincides with Proposition 2.3, since the Karhunen-Loève eigenvalues tend to zero fast the algebraic decay (2.14). Hence we can choose  $t$  as large as possible in (2.15) which implies (2.20).

## 2.4 Parametric Deterministic Problem

To parametrize the stochastic input, we make the following assumption on the random variables  $Y_m$  in the KL-representation (2.11) of  $a$ .

**Assumption 2.5.** a) The family  $(Y_m)_{m \geq 1} : \Omega \rightarrow \mathbb{R}$  is *independent*,

b) the KL-expansion (2.11) of the input data is *finite*, i.e. there exists  $\bar{M} < \infty$  such that  $Y_m \equiv 0$  for all  $m > \bar{M}$ ,

c) with each  $Y_m(\omega)$  in (2.11), (2.12) is associated a complete probability space  $(\Omega_m, \Sigma_m, P_m)$ ,  $m \in \mathbb{N}$  with the following properties:

(i) the range of  $Y_m$ ,  $\Gamma_m := \text{Ran}(Y_m) \subseteq \mathbb{R}$ , is compact and, after scaling, equal to  $[-1, 1]$  for all  $m$ , and

(ii) the probability measure  $P_m$  admits a probability density function  $\rho_m : \Gamma_m \rightarrow [0, \infty)$  such that  $dP_m(\omega) = \rho_m(y_m) dy_m$ ,  $m \in \mathbb{N}$ ,  $y_m \in \Gamma_m$ , and



- (iii) the sigma algebras  $\Sigma_m$  are subsets of the Borel sets of the interval  $\Gamma_m$ , i.e.  $\Sigma_m \subseteq \mathcal{B}(\Gamma_m)$ .

In order to numerically treat the KL-expansion (2.2), it is truncated after  $M < \bar{M}$  terms (with  $M$  to be determined as part of the discretization). We define

$$a_M(\omega, \mathbf{x}) := \mathbb{E}_a(\mathbf{x}) + \sum_{m=1}^M \sqrt{\lambda_m} \varphi_m(\mathbf{x}) Y_m(\omega). \quad (2.22)$$

With the finite expansion (2.22), we replace in (1.2) the random input  $a(\omega, \mathbf{x})$  by its  $M$ -term truncated Karhunen-Loève expansion  $a_M(\omega, \mathbf{x})$  in (2.22). We obtain an  $M$ -dimensional parametric, deterministic variational formulation which is equivalent to the stochastic problem (1.2) with input that is approximated by the truncated Karhunen-Loève expansion (2.22).

To state this formulation as well as for the stochastic Galerkin FEM (sGFEM) discretization below, we introduce the following function spaces: For

$$\Gamma := \Gamma_1 \times \Gamma_2 \times \dots, \quad \mathbf{y} = (y_1, y_2, \dots) \in \Gamma$$

we denote by  $L_\rho^2(\Gamma)$  the space of all mappings  $v : \Gamma \rightarrow \mathbb{R}$  which are square integrable with respect to the measure  $\rho(d\mathbf{y}) = \rho(\mathbf{y})d\mathbf{y}$ .

We shall assume that  $\rho(d\mathbf{y})$  is a probability measure on  $\Gamma$ . If  $H$  denotes any separable Hilbert space with norm  $\|\cdot\|_H$ , by  $L_\rho^2(\Gamma; H)$  we denote the Bochner space of functions  $v : \Gamma \rightarrow H$  for which  $\|v(\mathbf{y}, \cdot)\|_H : \mathbf{y} \rightarrow \mathbb{R}$  belongs to  $L_\rho^2(\Gamma)$ .

We shall require in particular the space

$$L_\rho^2(H_0^1) := L_\rho^2(\Gamma; H_0^1(D)) \simeq L_\rho^2(\Gamma) \otimes H_0^1(D) \quad (2.23)$$

where  $\otimes$  denotes the tensor product between separable Hilbert spaces.

By the independence assumption 2.5 a), the probability density  $\rho(\mathbf{y})$  is separable, i.e.

$$\rho(\mathbf{y}) = \prod_{m \geq 1} \rho_m(y_m)$$

with  $\rho_m(y_m)$  as in Assumption 2.5 c) i).

The parametric, deterministic formulation of (1.2), (2.11) reads, in variational form: find

$$u_M \in L_\rho^2(H_0^1) : \quad b_M(u_M, v) = l(v) \quad \forall v \in L_\rho^2(H_0^1) \quad (2.24)$$

with

$$\begin{aligned} b_M(u_M, v) &= \mathbb{E} \left[ \int_D a_M(\mathbf{y}, \mathbf{x}) \nabla u_M(\mathbf{y}, \mathbf{x}) \cdot \nabla v(\mathbf{y}, \mathbf{x}) \, d\mathbf{x} \right] \\ &= \int_\Gamma \int_D a_M(\mathbf{y}, \mathbf{x}) \nabla u_M(\mathbf{y}, \mathbf{x}) \cdot \nabla v(\mathbf{y}, \mathbf{x}) \rho(\mathbf{y}) \, d\mathbf{x} d\mathbf{y} \end{aligned} \quad (2.25)$$

and

$$l(v) = \mathbb{E} \left[ \int_D f(\mathbf{x}) v(\mathbf{y}, \mathbf{x}) \, d\mathbf{x} \right] = \int_\Gamma \int_D f(\mathbf{x}) v(\mathbf{y}, \mathbf{x}) \rho(\mathbf{y}) \, d\mathbf{x} d\mathbf{y}. \quad (2.26)$$

Here, based on (2.22),

$$a_M(\mathbf{y}, \mathbf{x}) = a_0(\mathbf{x}) + \sum_{m=1}^M y_m \psi_m(\mathbf{x}), \quad a_0 = \mathbb{E}_a, \quad \psi_m(\mathbf{x}) := \lambda_m^{1/2} \varphi_m(\mathbf{x}). \quad (2.27)$$

*Remark 2.6.* In what follows, the functions  $\psi_m$ , or more precisely  $L^\infty$ -estimates on those, will play a crucial role in the motivation of the algorithm presented in Section 4 to find a representation of the random solution  $u_M$  with as few terms as possible. Combining (2.14) and (2.15) we obtain

$$\forall m \in \mathbb{N} : \quad b_m := \|\psi_m\|_{L^\infty(D)} = \sqrt{\lambda_m} \|\varphi_m\|_{L^\infty(D)} \leq c_1 m^{-s} \quad (2.28)$$

where

$$s = \frac{1}{2d}(t - d - \varepsilon) \quad (2.29)$$

for an arbitrary  $\varepsilon \in (0, t - d/2]$  and where the constant  $c_1$  in (2.28) depends on  $\varepsilon, d$ .

*Remark 2.7.* The unique solvability of (2.24) uniformly in the number  $M$  of terms retained in the truncated Karhunen-Loève -expansion follows from Assumption 1.1 on the coefficient  $a(\omega, \mathbf{x})$  and from Proposition 2.3, i.e. from (2.14), (2.15). Specifically, under Assumption 1.1 and Proposition 2.3 (see also Remark 2.6) with sufficiently small  $C > 0$  in (2.14), there exists  $M_0 > 0$  and  $\gamma > 0$  (depending on  $a_{\min}, a_{\max}$  in (1.1), but independent of  $M$ ) such that

$$\forall M \geq M_0 \quad \forall v \in L_\rho^2(\Gamma; H_0^1(D)) : \quad b_M(v, v) \geq \gamma \|v\|_{L_\rho^2(\Gamma; H_0^1(D))}^2, \quad (2.30)$$

$$\forall v, w \in L_\rho^2(\Gamma; H_0^1(D)) : \quad |b_M(v, w)| \leq \gamma^{-1} \|v\|_{L_\rho^2(\Gamma; H_0^1(D))} \|w\|_{L_\rho^2(\Gamma; H_0^1(D))} \quad (2.31)$$

by the Lax-Milgram Lemma.

For  $M \geq M_0$ ,  $a_M$  satisfies (1.1) and the corresponding random solution  $u_M \in L_P^2(\Omega, V^D)$  exists, is unique and, by (2.3), there exists  $C > 0$  independent of  $M$  such that

$$\|u - u_M\|_{L_\rho^2(\Gamma, H_0^1(D))} \leq C \|a - a_M\|_{L^\infty(\Gamma, L^\infty(D))} \|f\|_{H^{-1}(D)}. \quad (2.32)$$

## 3 Sparse Tensor Discretization

### 3.1 Hierarchic subspace sequences

In sGFEM, we discretize the variational formulation (2.24) by Galerkin projection onto a hierarchic sequence of finite dimensional subspaces of  $L_\rho^2(\Gamma; H_0^1(D))$  in (2.23). Specifically, we choose two hierarchic families of finite dimensional subspaces,

$$V_0^\Gamma \subset V_1^\Gamma \subset \dots \subset V_{l_1}^\Gamma \subset V_{l_1+1}^\Gamma \subset \dots \subset L_\rho^2(\Gamma) \quad (3.1)$$

and

$$V_0^D \subset V_1^D \subset \dots \subset V_{l_2}^D \subset V_{l_2+1}^D \subset \dots \subset H_0^1(D) \quad (3.2)$$

with  $l_1, l_2$  being the *levels of refinement*. We introduce *detail spaces*  $W_{l_1}^\Gamma$  and  $W_{l_2}^D$  such that

$$W_0^\Gamma := V_0^\Gamma \quad \text{and} \quad V_{l_1}^\Gamma = V_{l_1-1}^\Gamma \oplus W_{l_1}^\Gamma \quad \text{for } l_1 = 1, 2, \dots \quad (3.3)$$

and

$$W_0^D := V_0^D \quad \text{and} \quad V_{l_2}^D = V_{l_2-1}^D \oplus W_{l_2}^D \quad \text{for } l_2 = 1, 2, \dots \quad (3.4)$$

where the sums are direct so that the (finite-dimensional) approximation spaces  $V_L^\Gamma$  and  $V_L^D$  admit a multilevel decomposition

$$V_L^\Gamma = \bigoplus_{l_1=0}^L W_{l_1}^\Gamma \quad \text{and} \quad V_L^D = \bigoplus_{l_2=0}^L W_{l_2}^D. \quad (3.5)$$

Furthermore, let

$$\Psi_{l_1}^\Gamma = \{\psi_{l_1 k}^\Gamma : k \in \nabla_{l_1}^\Gamma\} \quad \Psi_{l_2}^D = \{\psi_{l_2 k}^D : k \in \nabla_{l_2}^D\}, \quad (3.6)$$

with  $\nabla_{l_1}^\Gamma, \nabla_{l_2}^D$  being appropriate index sets, be bases of the increment spaces  $W_{l_1}^\Gamma$  and  $W_{l_2}^D$ , respectively. The two sets

$$\Psi^\Gamma = \{\psi_{l_1 k}^\Gamma : l_1 \geq 0, k \in \nabla_{l_1}^\Gamma\}, \quad \Psi^D = \{\psi_{l_2 k}^D : l_2 \geq 0, k \in \nabla_{l_2}^D\} \quad (3.7)$$

are assumed to be Riesz bases of  $L_\rho^2(\Gamma)$  and of  $H_0^1(D)$ , respectively.

In the next two sections we present examples of hierarchic sequences of subspaces for both components of  $\mathcal{H}_0^1 = L_\rho^2(\Gamma; H_0^1(D))$ : for  $L_\rho^2(\Gamma)$ , we use spaces of so-called ‘chaos-polynomials’, as was done e.g. in [4, 5, 7, 15, 20, 38, 37, 40]. This will be described in the following section. For  $H_0^1(D)$ , we use a multilevel family of Finite-Element type subspaces. As will become clear, we need a stable multilevel splitting so that we opt to use a spline-wavelet Finite-Element basis, which we present in the next section.

### 3.2 Hierarchic polynomial approximation in $L_\rho^2(\Gamma; H)$

The crucial part in the sGFEM is the choice of the sequence  $\{V_l^\Gamma\}_{l=0}^\infty$  of subspaces in (3.1), due to the possibly arbitrary high dimension  $M$  of the parameter domain  $\Gamma$ .

It is therefore imperative to reduce the number of stochastic degrees of freedom to a minimum while still retaining accuracy. Since, by Assumption 2.5, the  $\Gamma_m$  are compact we assume w.l.o.g. that  $\Gamma_m = [-1, 1], \forall m$ . We further assume that the probability densities  $\rho_m$  are constant, i.e. we will restrict our discussion to the case of a uniform probability density  $\rho_m(y_m) = \frac{1}{2}$ .

By  $\mathbb{N}_c^\mathbb{N}$ , we denote the set of all multiindices of compact support, i.e.. all sequences  $\alpha = (\alpha_1, \alpha_2, \dots) \subset \mathbb{N}_0$  of nonnegative integers whose ‘support’, i.e. the set  $\text{supp}(\alpha) := \{j \in \mathbb{N} : \alpha_j \neq 0\}$ , is finite: for  $\alpha \in \mathbb{N}_c^\mathbb{N}$ , we denote by  $|\alpha|_0 = \#\text{supp}(\alpha)$ .

For  $\alpha \in \mathbb{N}_c^\mathbb{N}$ , we denote the *tensorized Legendre polynomial of degree  $\alpha$*  by

$$\mathcal{L}_\alpha(\mathbf{y}) := L_{\alpha_1}(y_1)L_{\alpha_2}(y_2) \cdots L_{\alpha_M}(y_M) \cdots \quad \mathbf{y} \in \Gamma \quad (3.8)$$

where  $L_{\alpha_m}(y_m)$  denotes the Legendre polynomial of degree  $\alpha_m$  of  $y_m \in [-1, 1]$  normalized such that  $L_{\alpha_m}(1) = 1$ . Then,  $L_\rho^2(\Gamma) = \overline{\text{span}\{\mathcal{L}_\alpha : \alpha \in \mathbb{N}_c^\mathbb{N}\}}$  where the closure is w.r. to  $L_\rho^2(\Gamma)$  and the solution  $u_M$  to the parametric, deterministic problem (2.24) can be represented in terms of Legendre polynomials as

$$u_M(\mathbf{y}, \mathbf{x}) = \sum_{\alpha \in \mathbb{N}_c^\mathbb{N}} u_\alpha(\mathbf{x}) \mathcal{L}_\alpha(\mathbf{y}), \quad (3.9)$$

where the Legendre ‘coefficient’  $u_\alpha \in H_0^1(D)$  is given by

$$u_\alpha = \left( \prod_{m=1}^M 2\alpha_m + 1 \right) \int_\Gamma u_M(\mathbf{y}, \cdot) \mathcal{L}_{M, \alpha}(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y} \quad (3.10)$$

where the equality has to be read in  $H_0^1(D)$ . Given any index set  $\Lambda \subset \mathbb{N}_0^M$  of cardinality  $|\Lambda| < \infty$ , we denote

$$u_{M, \Lambda}(\mathbf{y}, \mathbf{x}) = \sum_{\alpha \in \Lambda} u_\alpha(\mathbf{x}) \mathcal{L}_{M, \alpha}(\mathbf{y}). \quad (3.11)$$

the  $\Lambda$ -truncated Legendre expansion (3.10).

In the Galerkin approximation below, we aim at finding an increasing sequence  $\Lambda(1) \subset \Lambda(2) \subset \Lambda(3) \subset \cdots \subset \mathbb{N}_c^\mathbb{N}$  of index sets  $\Lambda(l)$  such that the approximations  $u_{M, \Lambda(l)}$  converge at a certain rate  $0 < r \in \mathbb{R}$  in terms of  $|\Lambda(l)|$ . We adopt  $|\Lambda(l)|$  as

(rough) measure of work (construction of  $u_{M,\Lambda}$  in (3.11) requires computation of  $|\Lambda|$  many functions  $u_\alpha \in H_0^1(D)$ ).

For exponential eigenvalue decay (2.9) in the KL-expansion (2.11), i.e. if  $\lambda_m \lesssim e^{-cm^\kappa}$ ,  $\kappa := \frac{1}{d}$ , this is indeed possible with arbitrary high algebraic convergence rates  $r$  as it has been shown in [7, 33]. There, an algorithm has been presented to find index sets  $\Lambda$ , based only on the input data  $a, f$ , such that there exists a constant  $C(r) > 0$ , independent of  $l$  and  $M$  and

$$\|u_M - u_{M,\Lambda_l}\|_{L_\rho^2(\Gamma, H_0^1(D))} \leq C(r)|\Lambda(l)|^{-r} \quad \forall r > 0 \quad (3.12)$$

The case of algebraic decay (2.14) of the Karhunen-Loève eigenvalues however, is more difficult due to the possibly slow decay of the “coefficients” in (3.11).

We use best  $N$ -term approximations of  $u_M$  in (3.9) as a benchmark and prove a similar result to (3.12) in the case of algebraic Karhunen-Loève eigenvalue decay (2.14), (2.28), for a certain range of rates  $r$  by providing sufficient conditions on the input data.

To this end, for any finite set  $\Lambda \subset \mathbb{N}_c^{\mathbb{N}}$  and for any Hilbert space  $H$ , we denote by

$$\Pi(\Lambda; H) := \left\{ v = \sum_{\alpha \in \Lambda} v_\alpha \mathcal{L}_\alpha : v_\alpha \in H, \alpha \in \Lambda \right\} \subset L_\rho^2(\Gamma; H) \quad (3.13)$$

the linear space of  $H$ -valued polynomials which can be expressed as a finite linear combination of  $|\Lambda|$  many  $\rho(d\mathbf{y})$  orthogonal polynomials  $\mathcal{L}_\alpha(\mathbf{y})$  (if  $\Lambda = \emptyset$ , we define  $\Pi(\Lambda; H) = \{0\}$ ). They are a countable  $\rho(d\mathbf{y})$ -orthonormal basis of  $L_\rho^2(\Gamma)$  and any  $v \in L_\rho^2(\Gamma; H)$  is determined by its coefficient vector  $(v_\alpha : \alpha \in \mathbb{N}_c^{\mathbb{N}}) \subset H$  via  $v = \sum_{\alpha \in \mathbb{N}_c^{\mathbb{N}}} v_\alpha \mathcal{L}_\alpha$ . Moreover, we have the isometry

$$\|v\|_{L_\rho^2(\Gamma; H)}^2 = \sum_{\alpha \in \mathbb{N}_c^{\mathbb{N}}} \|v_\alpha\|_H^2 = \|\{ \|v_\alpha\|_H : \alpha \in \mathbb{N}_c^{\mathbb{N}} \}\|_{\ell_2}^2.$$

The *best  $N$ -term semidiscrete approximation error* of  $u \in L_\rho^2(H)$  is defined as

$$\sigma_N(u) := \inf_{\Lambda \subset \mathbb{N}_c^{\mathbb{N}}, |\Lambda| \leq N} \inf_{v \in \Pi(\Lambda; H)} \|u - v\|_{L_\rho^2(H)}. \quad (3.14)$$

Note that  $\sigma_N(u)$  is uniquely defined even though  $v$  might not be unique.

As is well-known, best  $N$ -term approximations are generally not computationally accessible. We therefore present in Section 5 an algorithm to locate for certain sequences  $\{\lambda_m\}_{m=1}^\infty$  of Karhunen-Loève -eigenvalues corresponding sequences  $\{\Lambda_l\}_{l=0}^\infty$  of index sets  $\Lambda_l \subset \mathbb{N}_c^{\mathbb{N}}$  of “active” indices in the “chaos” expansions (3.11) which realize “quasi-best  $N$ -term approximations”.

We define  $\mathcal{A}_r(H)$  as the set of those  $u \in L_\rho^2(\Gamma; H)$  for which

$$|u|_{\mathcal{A}_r(H)} := \sup_{N \geq 1} N^r \sigma_N(u) \quad (3.15)$$

is finite. This class consists of functions in  $L_\rho^2(\Gamma; H)$  which can be best  $N$ -term approximated at a rate  $r$ . For a parameter  $\gamma > 0$  we introduce the index sets

$$\Lambda_\gamma(l_1) := \arg \max_{\substack{\Lambda \subset \mathbb{N}_0^{\mathbb{N}} \\ |\Lambda| = \lceil 2^{\gamma l_1} \rceil}} \left( \sum_{\alpha \in \Lambda} \|u_\alpha\|_H \right) \subset \mathbb{N}_0^{\mathbb{N}}, \quad l_1 = 0, 1, 2, \dots \quad (3.16)$$

of the  $\lceil 2^{\gamma l_1} \rceil$  coefficients  $u_\alpha$  in (3.9) with the largest  $H$ -norm. Thus  $\Lambda_\gamma(l_1)$  provides a best  $\lceil 2^{\gamma l_1} \rceil$ -term approximation of  $u \in L_\rho^2(\Gamma; H)$ .

To apply these concepts to (1.2), we choose  $H = H_0^1(D)$ . The choice of the parameter  $\gamma > 0$  will be discussed in Section 3.7 and be based on the expected

convergence rate of the spatial discretization of (1.2). The corresponding approximation spaces are then given by

$$V_{l_1}^\Gamma := \Pi(\Lambda_\gamma(l_1); V). \quad (3.17)$$

Since the Karhunen-Loève eigenvalue sequence  $\{\lambda_m\}$  in (2.8) is monotonically decreasing, the index sets  $\Lambda_\gamma(l_1)$  are nested for a fixed  $\gamma > 0$  and hence also the  $V_{l_1}^\Gamma$  in the sense of (3.1). The detail spaces  $W_{l_1}^\Gamma$  then consist exactly of the multivariate Legendre polynomials corresponding to the indices in  $\Lambda_\gamma(l_1) \setminus \Lambda_\gamma(l_1 - 1)$ , i.e.

$$W_{l_1}^\Gamma = \Pi(\Lambda(l_1) \setminus \Lambda(l_1 - 1); V) \quad (3.18)$$

Further define the  $L_\rho^2$  projection  $P_{l_1}^\Gamma : L_\rho^2(\Gamma) \rightarrow V_{l_1}^\Gamma$  by  $P_{l_1}^\Gamma u_M := u_{M, \Lambda_\gamma(l_1)}$  as in (3.11). Then the following approximation properties hold.

**Proposition 3.1.** *Let  $s > 1$  be the decay rate of  $\psi_m(\mathbf{x})$  given by (2.29). If  $u_M$  solves (2.24) then for each  $0 < r < s - \frac{3}{2}$  exists a constant  $C(r)$  such that for every  $\gamma > 0$  and for the sequence of projections  $P_{l_1}^\Gamma$  corresponding to the index sets  $\Lambda_\gamma(l_1)$  in (3.16) it holds*

$$\|u_M - P_{l_1}^\Gamma u_M\|_{L_\rho^2(\Gamma; H_0^1(D))} \leq C(r) (N_{l_1}^\Gamma)^{-r} |u_M|_{\mathcal{A}_r(H_0^1(D))} \quad (3.19)$$

where  $N_{l_1}^\Gamma := |\Lambda_\gamma(l_1)| \rightarrow \infty$  as  $l_1 = 0, 1, 2, \dots$ . Since  $N_{l_1}^\Gamma = \lceil 2^{\gamma l_1} \rceil$ , we can express (3.19) equivalently as

$$\|u_M - P_{l_1}^\Gamma u_M\|_{L_\rho^2(\Gamma; H_0^1(D))} \leq C(r) 2^{-l_1 \gamma r} |u_M|_{\mathcal{A}_r(H_0^1(D))} \quad (3.20)$$

Here, the constant  $C(r)$  is independent of  $M$  (see Appendix A for a proof).

*Remark 3.2.* To have  $r > 0$  in (3.19) requires  $s > \frac{3}{2}$  in (2.29). Such a decay can be expected, for example, if  $C_a \in H^{t,t}(D \times D)$  with  $t > 4d$  as it can be easily derived from Remark 2.6. In other words, the regularity of the 2-point correlation  $C_a$  implies the range of possible rates  $r$  of best  $N$ -term approximations of  $u_M$ .

*Remark 3.3.* Proposition 3.1 indicates only the existence and the convergence rate  $r$  of a *semidiscrete* best- $N^\Gamma$ -term approximation. These approximations are not constructive: the proof uses the (a-priori unknown) values of  $\|u_\alpha\|_{H_0^1(D)}$ . However, in Section 4 we will present a strategy based on the Karhunen-Loève eigenvalues of the input data to a-priori locate sequences of sets  $\Lambda_\gamma(l_1)$  that appear to be close to the optimal sets in numerical experiments. Moreover, we will prove that the time and memory used to find those sets is linear (up to logarithms) in the size of the index sets.

### 3.3 Hierarchic Multilevel Approximation in $D$

As hierarchical basis for the spatial discretization in the physical domain  $D$  we choose finite element wavelet bases. In the following we will briefly present a construction of such finite element wavelet bases and recapitulate basic approximation results. For a more detailed introduction into wavelets we refer to [11, 13]. The construction presented here is based on [22] which, in principle, allows us to construct wavelets on a regular simplicial triangulation of  $D \subset \mathbb{R}^d$ . Here, we will treat the case of  $d = 1$  and  $d = 2$  explicitly.

Let  $\{\mathcal{T}_{l_2}\}_{l_2}$  be a nested sequence of regular simplicial triangulations of  $D$  and denote by  $\mathcal{I}(\mathcal{T}_{l_2})$  the index set of vertices of the mesh  $\mathcal{T}_{l_2}$ , denoted by  $\mathcal{C}(\mathcal{T}_{l_2})$ , and by  $\hat{\mathcal{I}}(\mathcal{T}_{l_2+1})$  the index set of vertices of the mesh  $\mathcal{T}_{l_2+1}$  which do not belong to  $\mathcal{T}_{l_2}$ . We then define

$$V_{l_2}^D := S^1(D, \mathcal{T}_{l_2}) = \{u \in H_0^1(D) : u|_T \in \mathcal{P}_1(T) \text{ for } T \in \mathcal{T}_{l_2}\}, \quad (3.21)$$

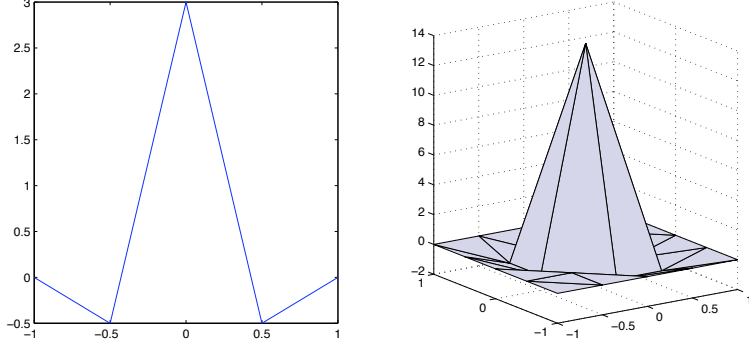


Figure 3.1: The function  $\theta_{l_2}^i$  on the reference element  $[-1, 1]^d$  for  $d = 1$  (left) and  $d = 2$  (right)

i.e. the space of continuous piecewise linear functions on the triangulation  $\mathcal{T}_{l_2}$ . Clearly, the spaces  $V_{l_2}^D$  are hierarchic in the sense of (3.2). Denote by  $\phi_{l_2}^k(\mathbf{x})$ ,  $k \in \mathcal{I}(\mathcal{T}_{l_2})$  the standard hat functions on the mesh  $\mathcal{T}_{l_2}$ , i.e. the piecewise linear polynomials with value 1 at the vertex  $k$  and zero at the other nodes. The scaling functions on the level  $l_2 = 0$  are then defined as the hat functions on the coarsest mesh  $\mathcal{T}_0$ , i.e.  $\psi_{0k} = \phi_0^k$  and  $\nabla_0 = \mathcal{I}(\mathcal{T}_0)$ . The construction of wavelets on a higher level  $l_2 > 0$  is based on the meshes  $\mathcal{T}_{l_2}$  and  $\mathcal{T}_{l_2-1}$ . First, we construct a family of functions  $\theta_{l_2}^i(\mathbf{x}) \in S^1(D, \mathcal{T}_{l_2})$ ,  $i \in \mathcal{I}(\mathcal{T}_{l_2-1})$  satisfying  $(\theta_{l_2}^i, \phi_{l_2-1}^k)_{L^2(D)} \simeq \delta_{ik}$ . Such functions are given in  $d = 1$  by

$$\theta_{l_2}^i(\mathbf{v}) = \begin{cases} 3 & \mathbf{v} = \mathbf{v}_i \in \mathcal{C}(\mathcal{T}_{l_2-1}) \\ -\frac{1}{2} & \mathbf{v} \in \mathcal{C}(\mathcal{T}_{l_2}) \text{ is neighbor of } \mathbf{v}_i \\ 0 & \text{any other } \mathbf{v} \in \mathcal{C}(\mathcal{T}_{l_2}) \end{cases} \quad (3.22)$$

and in  $d = 2$  by

$$\theta_{l_2}^i(\mathbf{v}) = \begin{cases} 14 & \mathbf{v} = \mathbf{v}_i \in \mathcal{C}(\mathcal{T}_{l_2-1}) \\ -1 & \mathbf{v} \in \mathcal{C}(\mathcal{T}_{l_2}) \text{ is neighbor of } \mathbf{v}_i \\ 0 & \text{any other } \mathbf{v} \in \mathcal{C}(\mathcal{T}_{l_2}) \end{cases} \quad (3.23)$$

See Figure 3.3 for an example of such a  $\theta_{l_2}^i$  in one and two dimensions.

The ensemble of functions  $\{\theta_{l_2}^i : i \in \mathcal{I}(\mathcal{T}_{l_2-1})\} \cup \{\phi_{l_2}^i : i \in \hat{\mathcal{I}}(\mathcal{T}_{l_2})\}$  forms a  $L^2$ -Riesz basis of  $V_{l_2}^D$  satisfying  $(\theta_{l_2}^i, \phi_{l_2-1}^k)_{L^2(D)} = 0$  if  $i \neq k$ . The wavelets on level  $l_2$  are then obtained by the formula

$$\psi_{l_2 i}(\mathbf{x}) = \phi_{l_2}^i(\mathbf{x}) - \sum_{k \in \mathcal{I}(\mathcal{T}_{l_2-1})} \frac{(\phi_{l_2}^i, \phi_{l_2-1}^k)_{L^2(D)}}{(\theta_{l_2}^k, \phi_{l_2-1}^k)_{L^2(D)}} \theta_{l_2}^k(\mathbf{x}), \quad i \in \hat{\mathcal{I}}(\mathcal{T}_{l_2}) \quad (3.24)$$

An example of such a wavelet on a triangular mesh in  $D \subset \mathbb{R}^2$  can be seen in Figure 3.3

The functions  $\psi_{l_2 i}$ ,  $i \in \nabla_{l_2} := \hat{\mathcal{I}}(\mathcal{T}_{l_2})$  form a uniform  $L^2$ -Riesz basis for  $W_{l_2}^D$ , see [22], Proposition 3.2.10. Scaling the wavelets with a factor  $2^{-l_2}$  i.e. defining  $\tilde{\psi}_{l_2 i}(\mathbf{x}) = 2^{-l_2} \psi_{l_2 i}(\mathbf{x})$  then form a Riesz basis for  $H^1(D)$ . In case of homogeneous Dirichlet boundary conditions, the above construction can be modified as follows: For  $\mathbf{v}_i \in \mathcal{C}(\mathcal{T}_{l_2})$ , the corresponding  $\phi_{l_2-1}^i$ ,  $\phi_{l_2}^i$  and  $\theta_{l_2}^i$  are excluded from the ensembles and the resulting wavelets  $\tilde{\psi}_{l_2 j}$  then form a uniform Riesz bases for  $H_0^1(D)$ .

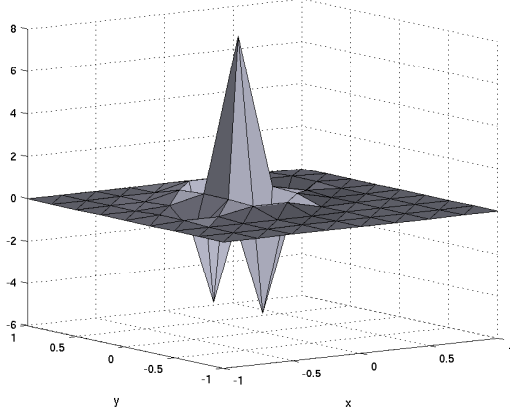


Figure 3.2: A piecewise linear finite element wavelet on a triangular mesh in 2-d

Now define the  $H_0^1(D)$ -projection  $P_{l_2}^D : L^2(D) \rightarrow V_{l_2}^D$ . For functions  $u \in H^{1+t}(D)$   $t \in [0, 1]$  and  $l_2 \geq 0$  the following approximation property is proved in [22], Lemma 2.3.1:

$$\|u - P_{l_2}^D u\|_{H^1(D)} \lesssim 2^{-l_2 t} \|u\|_{H^{1+t}(D)}, \quad t \in [0, 1] \quad (3.25)$$

Noting that  $N_{l_2}^D := \dim(V_{l_2}^D) \simeq 2^{l_2 d}$  we then obtain the following convergence rates with respect to the number of degrees of freedom:

$$\|u - P_{l_2}^D u\|_{H^1(D)} \lesssim (N_{l_2}^D)^{-t/d} \|u\|_{H^{1+t}(D)}, \quad t \in [0, 1]. \quad (3.26)$$

### 3.4 Sparse Tensor Product Spaces

Having specified the subspace hierarchies  $\{V_\ell^\Gamma\}_{\ell \geq 0} \subset L_\rho^2(\Gamma)$  and  $\{V_\ell^D\}_{\ell \geq 0} \subset H_0^1(D)$  in (2.23) which admit splittings (3.3) and (3.4), we denote by

$$V_L^\Gamma \otimes V_L^D = \bigoplus_{0 \leq l_1, l_2 \leq L} W_{l_1}^\Gamma \otimes W_{l_2}^D \subset L_\rho^2(\Gamma) \otimes H_0^1(D) \quad (3.27)$$

the (full) tensor product space of the finite dimensional component subspaces  $V_L^D$  and  $V_L^\Gamma$ , respectively. Unlike earlier work, we analyze below the approximate solution of the parametric deterministic problem (2.24) by Galerkin projection onto *sparse tensor product spaces* defined by

$$V_L^\Gamma \hat{\otimes} V_L^D := \bigoplus_{0 \leq l_1 + l_2 \leq L} W_{l_1}^\Gamma \otimes W_{l_2}^D, \quad (3.28)$$

see also Figure 3.6 for an illustration of  $V_L^\Gamma \hat{\otimes} V_L^D$  in terms of the component detail spaces. Every function  $u \in V_L^\Gamma \hat{\otimes} V_L^D$  can therefore be decomposed as

$$u = \sum_{l_2=0}^L \sum_{l_1=0}^{L-l_2} \sum_{k \in \nabla_{l_2}^D} \sum_{k' \in \nabla_{l_1}^\Gamma} d_{l_1 l_2 k k'} \psi_{l_1 k'}^\Gamma \psi_{l_2 k}^D. \quad (3.29)$$

### 3.5 Sparse Tensor sGFEM discretization

The stochastic Galerkin FEM discretization of the elliptic sPDE (1.2), based on its deterministic, parametric reformulation (2.24), can be written as:

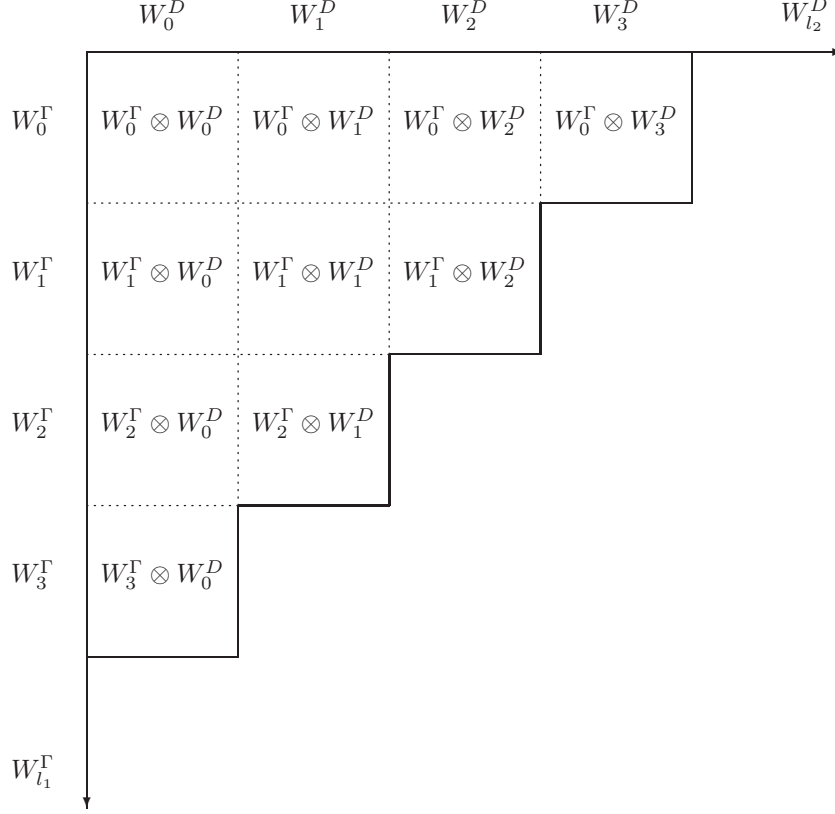


Figure 3.3: Illustration of the sparse tensor product space  $V_L^\Gamma \hat{\otimes} V_L^D$

Find

$$\hat{u}_M \in V_L^\Gamma \hat{\otimes} V_L^D : \quad b_M(\hat{u}_M, v) = l(v) \quad \forall v \in V_L^\Gamma \hat{\otimes} V_L^D \quad (3.30)$$

where  $b_M(\cdot, \cdot)$  and  $l(\cdot)$  are given by (2.25) and (2.26), respectively. As a consequence of (2.30), (2.31), if  $M$  is sufficiently large (3.30) defines, for every  $L \geq 0$ , a unique sGFEM approximation  $\hat{u}_M \in V_L^\Gamma \hat{\otimes} V_L^D$  which is a quasioptimal approximation in  $L^2_\rho(\Gamma; H_0^1(D))$  of  $u_M$  defined in (2.24):

$$\|u_M - \hat{u}_M\|_{L^2_\rho(\Gamma; H_0^1(D))} \leq C \|u_M - \hat{v}\|_{L^2_\rho(\Gamma; H_0^1(D))} \quad \forall \hat{v} \in V_L^\Gamma \hat{\otimes} V_L^D. \quad (3.31)$$

### 3.6 Complexity Estimate

In contrast to full tensor product discretizations, where the complexity of sGFEM and sCFEM algorithms is  $O(N_\Gamma^L N_D^L)$ , we will show log-linear complexity in either  $N_\Gamma^L$  or  $N_D^L$  for the corresponding sparse tensor product Galerkin discretizations.

**Lemma 3.4.** *Assume that the dimensions of the detail spaces  $W_{l_1}^\Gamma$  and  $W_{l_2}^D$  grow exponentially with respect to the levels  $l_1, l_2$ , i.e. there exist two geometric sequences  $b_\Gamma^{l_1}$  and  $b_D^{l_2}$ ,  $l_1, l_2 = 1, 2, 3, \dots$ , with bases  $b_\Gamma$  and  $b_D$  such that*

$$\dim(W_{l_1}^\Gamma) \sim b_\Gamma^{l_1} \quad \text{and} \quad \dim(W_{l_2}^D) \sim b_D^{l_2}$$

*Then the sparse tensor product space  $V_L^\Gamma \hat{\otimes} V_L^D$  is of cardinality*

$$\dim(V_L^\Gamma \hat{\otimes} V_L^D) \sim O(L^\theta \max(b_\Gamma, b_D)^{L+1}) \quad (3.32)$$



where  $\theta = 1$  if  $b_\Gamma = b_D$  and zero otherwise.

*Proof.*

$$\dim(V_L^\Gamma \hat{\otimes} V_L^D) \leq c \sum_{l_1=0}^L \sum_{l_2=0}^{L-l_1} b_\Gamma^{l_1} b_D^{l_2} \leq c \sum_{l_1=0}^L b_\Gamma^{l_1} b_D^{L-l_1+1} \quad (3.33)$$

In the case where  $b_\Gamma = b_D$  therefore have

$$\dim(V_L^\Gamma \hat{\otimes} V_L^D) \leq c \sum_{l_1=0}^L b_\Gamma^{L+1} = c(L+1)b_\Gamma^{L+1} \quad (3.34)$$

which leads to (3.32) with  $\theta = 1$ . Now we assume w.l.o.g.  $b_D > b_\Gamma$  and it follows from (3.33)

$$\begin{aligned} \dim(V_L^\Gamma \hat{\otimes} V_L^D) &\leq c \sum_{l_1=0}^L b_\Gamma^{l_1} b_D^{L-l_1+1} \\ &= c b_D^{L+1} \sum_{l_1=0}^L b_\Gamma^{l_1} \left( \frac{b_\Gamma}{b_D} \right)^{l_2} \\ &= c b_D^{L+1} \frac{(b_D/b_\Gamma)^{L+1} - 1}{(b_D/b_\Gamma) - 1} \end{aligned}$$

Since the denominator is positive it can be absorbed in the constant and we end up with

$$\dim(V_L^\Gamma \hat{\otimes} V_L^D) \leq c b_D^{L+1} \quad (3.35)$$

This completes the proof.  $\square$

### 3.7 Convergence rates of sparse tensor sGFEM

In Sections 3.2 and 3.3, convergence rates of the ‘polynomial chaos’ type stochastic and the multilevel spatial discretizations of the model diffusion problem (1.2) on  $\Gamma \times D$  have been discussed.

In the present section, we prove algebraic convergence rates of the sparse sGFEM discretization (3.30) of the parametric, deterministic formulation (2.24) of the sPDE (1.2), under appropriate regularity hypotheses for the unknown random solution. To obtain these convergence rate estimates for the sGFEM approximation  $\hat{u}_M$  defined in (3.30) of  $u$ , by (3.31) it remains to estimate the best-approximation error of  $u_M$  from the sparse tensor product space  $V_L^\Gamma \hat{\otimes} V_L^D$  and to combine this bound with the error introduced by truncation of the Karhunen-Loève series (2.32).

To this end, we define the sparse tensor product projection operator  $\hat{P}_L : L_\rho^2(\Gamma) \otimes H_0^1(D) \rightarrow V_L^\Gamma \hat{\otimes} V_L^D$  by

$$(\hat{P}_L v)(\mathbf{y}, \mathbf{x}) := \sum_{0 \leq l_1 + l_2 \leq L} (P_{l_1}^\Gamma - P_{l_1-1}^\Gamma) \otimes (P_{l_2}^D - P_{l_2-1}^D) v(\mathbf{y}, \mathbf{x}) \quad (3.36)$$

The error of the sparse sGFEM approximation can be decomposed into the error stemming from truncating the KL-expansion and from the additional discretization error resulting from approximation of the gpc coefficient  $u_\alpha$  on  $V_L^\Gamma \hat{\otimes} V_L^D$  rather than on  $L_\rho^2(\Gamma) \otimes H_0^1(D)$ .

To simplify notation we denote by  $\|\cdot\|$  the norm taken w.r.t.  $L_\rho^2(\Gamma; H_0^1(D))$  unless otherwise indicated. Denoting by  $u$  the solution to the original problem

(1.2), by  $u_M$  the solution to the truncated problem (2.24) and by  $\hat{u}_M$  the discrete solution to (3.30) we have, by the triangle inequality,

$$\underbrace{\|u - \hat{u}_M\|}_{\mathbf{e}} \leq \underbrace{\|u - u_M\|}_{\mathbf{e}_M} + \underbrace{\|u_M - \hat{P}_L u_M\|}_{\hat{\mathbf{e}}_L}. \quad (3.37)$$

To estimate the discretization error  $\hat{\mathbf{e}}_L$  we start from the definition of  $\hat{P}_L$  in (3.36) and write

$$\begin{aligned} \|v - \hat{P}_L v\| &\leq \left\| \sum_{l_1=0}^L \sum_{l_2=L-l_1+1}^{\infty} (P_{l_1}^{\Gamma} - P_{l_1-1}^{\Gamma}) \otimes (P_{l_2}^D - P_{l_2-1}^D) v \right\| \\ &\quad + \left\| \sum_{l_1=L+1}^{\infty} \sum_{l_2=0}^{\infty} (P_{l_1}^{\Gamma} - P_{l_1-1}^{\Gamma}) \otimes (P_{l_2}^D - P_{l_2-1}^D) v \right\| \\ &= \left\| \sum_{l_1=0}^L (P_{l_1}^{\Gamma} - \text{Id} + \text{Id} - P_{l_1-1}^{\Gamma}) \otimes (\text{Id} - P_{L-l_1+1}^D) v \right\| \\ &\quad + \left\| (\text{Id} - P_L^{\Gamma}) \otimes (\text{Id}) v \right\| \\ &\leq \sum_{l_1=0}^L \left( \left\| (\text{Id} - P_{l_1}^{\Gamma}) \otimes (\text{Id} - P_{L-l_1+1}^D) v \right\| \right. \\ &\quad \left. + \left\| (\text{Id} - P_{l_1-1}^{\Gamma}) \otimes (\text{Id} - P_{L-l_1+1}^D) v \right\| \right) + C(r) 2^{-L\gamma r} \|v\|_{\mathcal{A}_r(H^{1+t}(D))} \\ &\leq (C(r, t) L 2^{-L \min(\gamma r, t)} + C(r) 2^{-L\gamma r}) \|v\|_{\mathcal{A}_r(H^{1+t}(D))} \end{aligned}$$

Note that we used the approximation results (3.20) and (3.25) in the last two lines of the estimates above. We equilibrate the two error bounds by choosing

$$\gamma := \frac{t}{r} \quad (3.38)$$

and we obtain for the error  $\hat{\mathbf{e}}_L$  the estimate

$$\hat{\mathbf{e}}_L = \|u_M - \hat{P}_L u_M\|_{L^2_{\rho}(\Gamma; H_0^1(D))} \leq C(r, t) L 2^{-Lt} \|u_M\|_{\mathcal{A}_r(H^{1+t}(D))} \quad (3.39)$$

with a constant possibly depending on the stochastic approximation rate  $r$ .

It remains to estimate the Karhunen-Loève -truncation error  $\mathbf{e}_M$  and to choose  $M$  s.t.  $\mathbf{e}_M \sim \hat{\mathbf{e}}_L$ .

Due to the orthonormality of the eigenfunctions in (2.11) we have that

$$\mathbf{e}_M^2 = \sum_{m>M} \lambda_m \lesssim \sum_{m>M} m^{-s} \leq \int_M^{\infty} x^{-s} = s M^{-(s-1)} \quad (3.40)$$

Hence choosing  $M \sim L^{-2/(s-1)} 2^{2Lt/(s-1)}$  leads to

$$\|u - \hat{u}_M\|_{L^2_{\rho}(\Gamma; H_0^1(D))} \leq C(t) L 2^{-Lt} \|u_M\|_{\mathcal{A}_r((H_0^1 \cap H^{1+t})(D))} \quad (3.41)$$

Since the cardinality of the detail spaces  $W_{l_1}^{\Gamma}$ ,  $W_{l_2}^D$  both satisfy the assumptions of Lemma 3.4 we arrive at our main result on convergence rates for the sGFEM discretization of sPDEs.

**Proposition 3.5.** *Let the solution  $u$  to the model problem (1.2) satisfy*

$$u \in \mathcal{A}_r((H^{1+t} \cap H_0^1)(D)) \quad \text{forsome} \quad 0 < r < s - \frac{3}{2}, \quad 0 < t \leq 1 \quad (3.42)$$

with  $s > \frac{3}{2}$  as in (2.28). Let  $\hat{u}_M$  denote the sGFEM solution to the problem (3.30) with the sparse tensor product spaces sequence  $V_L^\Gamma \hat{\otimes} V_L^D$ , defined in (3.28) with the sequence  $V_\ell^\Gamma$  in (3.1) given by (3.17) with the index sets (3.16).

Then there exists a constant  $c_1 > 0$ , independent of  $M$  and  $L$ , such that by choosing  $M = \lceil c_1 L^{-1/(s-1)} 2^{Lt/(s-1)} \rceil$  we have

$$\|u - \hat{u}_M\|_{L^2_\rho(\Gamma; H_0^1(D))} \leq C(\beta) L^{1+\beta} (\hat{N}_L)^{-\beta} \|u\|_{\mathcal{A}_r(H^{1+t}(D))} \quad (3.43)$$

where  $\hat{N}_L := \dim(V_L^\Gamma \hat{\otimes} V_L^D)$  and  $\beta = \min(r, t/d)$ .

*Remark 3.6 (Higher order elements).* So far we only considered linear finite element wavelets in  $D$ , i.e.  $p = 1$ . Considering, more generally, elements of order  $p \geq 1$  results in convergence rates

$$\|u - \hat{u}_M\|_{L^2_\rho(\Gamma; H_0^1(D))} \lesssim L^{1+\beta} (\hat{N}_L)^{-\beta} \|u\|_{\mathcal{A}_r(H^{t+1}(D))} \quad (3.44)$$

with  $\beta = \min(r, t/d)$  if

$$u \in \mathcal{A}_r((H^{1+t} \cap H_0^1)(D)) \quad \text{for some } 0 < r < s - \frac{3}{2}, 0 < t \leq p$$

We see from this that by choosing  $p = \lceil rd \rceil$  we can obtain the maximum possible rate  $r$ .

*Remark 3.7 (full tensor product).* The main motivation for using the sparse tensor product (3.28) between hierarchic sequences of spatial and stochastic discretization spaces lies in the reduction of degrees of freedom to  $O(L^\theta \max(\dim(V_L^\Gamma), \dim(V_L^D)))$ , shown in Lemma 3.4, as opposed to  $O(\dim(V_L^\Gamma) \cdot \dim(V_L^D))$  in a full tensor approach. Using the approximation results (3.20) and (3.25) one can easily derive the full tensor convergence rates

$$\|u - P_L^D \otimes P_L^\Gamma u\| \lesssim (N_L)^{-\bar{\beta}} \|u\|_{\mathcal{A}_r((H_0^1 \cap H^{p+1})(D))} \quad (3.45)$$

with  $\bar{\beta} = (d/p + 1/r)^{-1}$  and where  $P_L^D$  and  $P_L^\Gamma$  denote the  $L^2$  projections onto the spatial and stochastic discretization spaces, respectively, as described in Subsections 3.2 and 3.3. A proof of the regularity (3.42) for  $p = 1$  is given in [10].

## 4 Localization of gPC modes

In the present section, we address an implementation of the sparse tensor sGFEM discretization of (2.1).

It is based on the observation, proved in Appendix A, that the decay of the KL-coefficients of the input random field determines the decay of the coefficients  $u_\alpha$  in the expansion (3.10) of the random solution which are to be computed approximately in the sGFEM. Specifically, we use (cf. (A.15), (A.8) in Appendix A) that for any  $\alpha \in \mathbb{N}_c^{\mathbb{N}}$  it holds

$$\|u_\alpha\|_H \lesssim \eta^{-\alpha} := \prod_{m \geq 1} \eta_m^{-\alpha_m}, \quad \eta_m = r_m + \sqrt{1 + r_m^2}, \quad r_m = \frac{a_{\min}^*}{C(\delta) \|\psi_m\|_{L^\infty(D)} m^{1+\delta}} \quad (4.1)$$

where  $\psi_m$  is defined in (2.27) being the spatial basis functions in the KL-expansion of  $a$ . In particular it follows from (4.1) that  $\eta_m^{-1} \lesssim \|\psi_m\|_{L^\infty} m^{1+\delta}$  and further by Remark 2.6 that  $\eta_m^{-1} \lesssim m^{-s+(1+\delta)}$ . In order to obtain a viable algorithm, efficient identification of quasi-optimal index sets  $V_{l_1}^\Gamma$  of ‘‘active’’ gpc-coefficients of possibly minimal cardinality is necessary. To this end we present an algorithm which identifies those ‘‘active’’ gpc-coefficients  $u_\alpha$  based on the sequence of upper bounds  $\{\|\psi_m\|_{L^\infty}\}_{m \geq 1}$ . We start by introducing some notation.

**Definition 4.1.** Let  $c_0$  be the collection of strictly decreasing sequences of positive real numbers bounded above by one and tending to zero:

$$c_0 = \{\mu = (\mu_1, \mu_2, \dots) \in \mathbb{R}_+^{\mathbb{N}} \mid \lim_{m \rightarrow \infty} \mu_m = 0 \text{ and } 1 > \mu_1 \geq \mu_2 \geq \dots\}. \quad (4.2)$$

By  $\mathbb{N}_0^{\mathbb{N}}$  denote the set of sequences of nonnegative integers “with compact support”, i.e. all sequences, for which only a finite number of terms are different from zero. As before, for  $\mu \in c_0$  and  $\alpha \in \mathbb{N}_c^{\mathbb{N}}$ , we denote by  $\mu^\alpha = \mu_1^{\alpha_1} \mu_2^{\alpha_2} \dots$ . Note that the product is well-defined due to  $\alpha \in \mathbb{N}_c^{\mathbb{N}}$  having only a finite number of nonzero elements. For any threshold  $\epsilon > 0$  and any  $\mu \in c_0$ , we define

$$\Lambda_\epsilon(\mu) = \{\alpha \in \mathbb{N}_c^{\mathbb{N}} \mid \mu^\alpha \geq \epsilon\} \subset \mathbb{N}_c^{\mathbb{N}} \quad (4.3)$$

and

$$M_\epsilon(\mu) = \begin{cases} 0, & \text{if } \Lambda_\epsilon(\mu) = \emptyset \\ \max_{m \in \mathbb{N}} \{\mu_m \geq \epsilon\}, & \text{else} \end{cases} \quad (4.4)$$

which is finite for any  $\mu \in c_0$  and any  $\epsilon > 0$ .

By 0 we denote both, the real number zero and the sequence  $(0, 0, 0, \dots)$  and we use the convention  $0^0 = 1$ . For  $\alpha \in \mathbb{N}_c^{\mathbb{N}}$  the support of  $\alpha$  is the set where its non-zero indices are located, and  $\text{nnz}(\alpha)$  the (finite) number of nonzero entries:

$$\text{supp}(\alpha) = \{m \in \mathbb{N} \mid \alpha_m \neq 0\}, \quad \text{nnz}(\alpha) = |\text{supp}(\alpha)|. \quad (4.5)$$

In the algorithm for the location of sets  $\Lambda \subset \mathbb{N}_c^{\mathbb{N}}$  of “active” indices and the derivation of asymptotic convergence rates of the sparse tensor product approximations, the following monotonicity properties of the index sets  $\Lambda_\epsilon(\mu)$  will be crucial.

**Definition 4.2.** Let  $\mu, \bar{\mu} \in c_0$ . We write

$$\mu \preceq \bar{\mu} \iff \mu_m \leq \bar{\mu}_m \quad \forall m \in \mathbb{N}.$$

For  $\mu \in c_0$  and any pair  $\alpha, \bar{\alpha} \in \Lambda_\epsilon(\mu)$  we write

$$\alpha \leq_\mu \bar{\alpha} \iff \mu^\alpha \leq \mu^{\bar{\alpha}}.$$

*Remark 4.3.* The relation  $\leq_\mu$  is a total quasi-order on  $\Lambda_\epsilon(\mu)$ , i.e. it is transitive, reflexive and any two multiindices are comparable. The elements of  $\Lambda_\epsilon(\mu)$  can therefore be quasi-sorted according to  $\leq_\mu$ , albeit not in a unique fashion.

We now make precise the monotonicity properties of the sets  $\Lambda_\epsilon(\mu)$ .

**Lemma 4.4.** For any  $\mu, \bar{\mu} \in c_0$  and any  $\bar{\epsilon}, \epsilon > 0$  the following properties hold:

- a)  $M_\epsilon(\mu) \leq |\Lambda_\epsilon(\mu)| < \infty$
- b)  $\forall \alpha \in \Lambda_\epsilon(\mu) : \text{supp}(\alpha) \subset \{1, 2, \dots, M_\epsilon(\mu)\}$
- c)  $\bar{\epsilon} \geq \epsilon$  implies  $\Lambda_{\bar{\epsilon}}(\mu) \subseteq \Lambda_\epsilon(\mu)$
- d)  $\mu \preceq \bar{\mu}$  implies  $\Lambda_\epsilon(\mu) \subseteq \Lambda_\epsilon(\bar{\mu})$ .

*Proof.* Assume  $\bar{\epsilon} \geq \epsilon > 0$  and  $\mu, \bar{\mu} \in c_0$  with  $\mu \preceq \bar{\mu}$ . With

$$M = M_\epsilon(\mu) < \infty \quad \text{and} \quad N = \max_{n \in \mathbb{N}_0} \{\mu_1^n \geq \epsilon\} < \infty$$

we have  $M \leq |\Lambda_\epsilon(\mu)| \leq (N+1)^M < \infty$ . The second property is a direct consequence of the definition of  $M_\epsilon(\mu)$ . Further, for  $\alpha \in \Lambda_{\bar{\epsilon}}(\mu)$  it holds

$$\mu^\alpha = \prod_{m \in \mathbb{N}} \mu_m^{\alpha_m} \geq \bar{\epsilon} \geq \epsilon$$

and hence,  $\alpha \in \Lambda_\epsilon(\mu)$ . Similarly,

$$\bar{\mu}^\alpha = \prod_{m \in \mathbb{N}} \bar{\mu}_m^{\alpha_m} \geq \prod_{m \in \mathbb{N}} \mu_m^{\alpha_m} \geq \epsilon$$

and hence,  $\alpha \in \Lambda_\epsilon(\bar{\mu})$ .  $\square$

We exploit the monotonicity properties of the sets  $\Lambda_\epsilon(\mu)$  in conjunction with the following sequence  $\nu_\sigma \in c_0$  which exemplifies Karhunen-Loève eigenvalue decay (2.9) of order  $\sigma > 0$ :

$$\nu_\sigma = \{(m+1)^{-\sigma} | m = 1, 2, 3, \dots\} \quad (4.6)$$

For  $\nu_\sigma$ , sharp asymptotics on the cardinality of the index sets  $\Lambda_\epsilon(\nu_\sigma)$  as  $\epsilon \rightarrow 0$  follow from results on integer factorization in number theory:

**Proposition 4.5.** *Let  $\sigma > 0$  and  $\nu_\sigma \in c_0$  be given by (4.6). Then, as  $\epsilon \rightarrow 0$ ,*

$$|\Lambda_\epsilon(\nu_\sigma)| \sim x \frac{e^{2\sqrt{\log x}}}{2\sqrt{\pi}(\log x)^{3/4}} \quad \text{with } x = \epsilon^{-1/\sigma}. \quad (4.7)$$

*In particular, for any sequence  $\mu$  satisfying  $\mu \preceq \nu_\sigma$  the sequence  $\mu^\alpha := \mu_1^{\alpha_1} \mu_2^{\alpha_2} \dots$  satisfies  $\{\mu^\alpha\}_{\alpha \in \mathbb{N}_c^{\mathbb{N}}} \in \ell_w^\tau$  for any  $\tau < 1/\sigma$ .*

*Proof.* First we observe that

$$\forall \epsilon > 0, \sigma > 0: \quad \Lambda_\epsilon(\nu_\sigma) = \Lambda_{\epsilon^{1/\sigma}}(\nu_1). \quad (4.8)$$

Let  $f(n)$  denote the number of multiplicative partitions of  $n \in \mathbb{N}$  modulo the order of factors, e.g.,  $f(12) = 4$ , since  $12 = 1 \cdot 12 = 2 \cdot 4 = 3 \cdot 4 = 2 \cdot 2 \cdot 3$ . Sharp asymptotics of the quantity

$$F_\Sigma(x) := \sum_{n \leq x} f(n)$$

as  $x \rightarrow \infty$  were found by Oppenheim [25] and Szekeres and Turan [32] (see also [9]) who proved

$$F_\Sigma(x) = F(x) (1 + \mathcal{O}(1/\log x)) \quad \text{with } F(x) = x \frac{e^{2\sqrt{\log x}}}{2\sqrt{\pi}(\log x)^{3/4}} \quad \text{as } x \rightarrow \infty.$$

Let now  $\epsilon > 0$  be so small that  $\Lambda_\epsilon(\nu_1) \setminus \{0\} \neq \emptyset$ . Then

$$\exists 0 \neq \alpha \in \Lambda_\epsilon(\nu_1): \quad \prod_{m \in \mathbb{N}} (m+1)^{\alpha_m} \leq 1/\epsilon$$

and since the product is an integer, the multiindex  $\alpha$  represents a multiplicative partition of some integer  $x \leq 1/\epsilon$ . On the other hand, any multiplicative partition of some integer  $x \leq 1/\epsilon$  yields a multiindex  $\alpha \in \mathbb{N}_0^{\mathbb{N}}$  such that  $\prod_{m \in \mathbb{N}} (m+1)^{\alpha_m} = x$  and thus  $\nu_1^\alpha = 1/x \geq \epsilon$ . Hence,  $|\Lambda_\epsilon(\nu_1)| = F_\Sigma(1/\epsilon) \sim F(1/\epsilon)$ . By (4.8), we have

$$|\Lambda_\epsilon(\nu_\sigma)| = |\Lambda_{\epsilon^{1/\sigma}}(\nu_1)| \sim F(\epsilon^{-1/\sigma}) \quad \text{for } \epsilon \rightarrow 0$$

which implies (4.7).

Finally, to show that for  $\mu \preceq \nu_\sigma$  the sequence  $\{\mu^\alpha\}_{\alpha \in \mathbb{N}_c^{\mathbb{N}}}$  belongs to  $\ell_w^\tau$  for any  $\tau < 1/\sigma$ , we use that a sequence  $x \in \ell_w^\tau$  iff  $x_n^* \lesssim n^{-1/\tau}$  (where  $(x_n^*)_n$  denotes the decreasing rearrangement of  $(x_n)_n$ ). This is equivalent to the existence of a constant  $c > 0$  such that for all  $\epsilon > 0$  it holds that  $\#\{n : |x_n| > \epsilon\} \leq c^\tau \epsilon^{-\tau}$  (cf. e.g. [16], (4.23)). By (4.7), the assertion  $\{\nu_\sigma^\alpha\}_{\alpha \in \mathbb{N}_c^{\mathbb{N}}} \in \ell_w^\tau$  for any  $\tau < 1/\sigma$  follows. The proof is complete since  $\{\mu^\alpha\}_{\alpha \in \mathbb{N}_c^{\mathbb{N}}}$  is dominated by  $\{(\nu_\sigma)^\alpha\}_{\alpha \in \mathbb{N}_c^{\mathbb{N}}}$  (cf. Lemma 4.4, part d).  $\square$

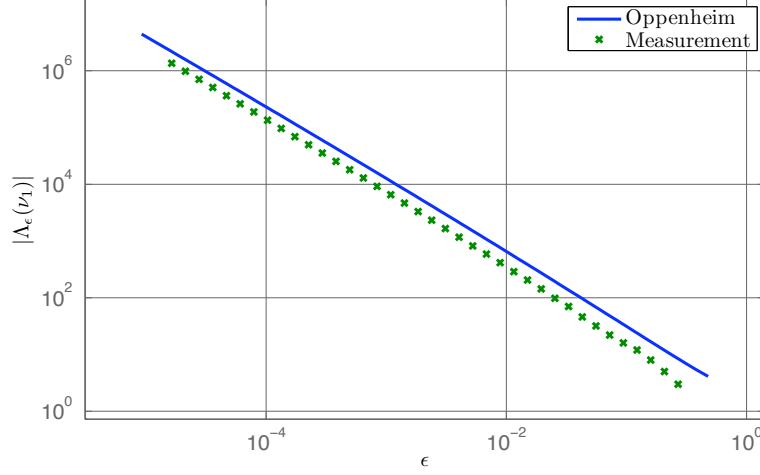


Figure 4.1: Cardinality  $|\Lambda_\epsilon(\nu_1)|$  as function of  $\epsilon$ .

The asymptotic result (4.7) is sharp already for thresholds  $\epsilon$  as large as  $10^{-2}$  as is evident from Fig. 4.1.

The following lemma will be useful to bound the number of non-zero entries in  $\alpha \in \Lambda_\epsilon(\mu)$  for particular sequences  $\mu$ .

**Lemma 4.6.** *For any fixed  $\sigma > 0$  there holds*

$$(M+1)^\sigma = o(\lceil \sigma \log M \rceil!) \quad \text{as } M \rightarrow \infty.$$

*Proof.* For  $\sigma > 0$  we have as  $n = \lceil \sigma \log M \rceil \rightarrow \infty$  by Stirling's formula

$$\lceil \sigma \log M \rceil! = n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \quad (4.9)$$

Hence

$$\lim_{M \rightarrow \infty} \frac{(M+1)^\sigma}{\lceil \sigma \log M \rceil!} \leq \lim_{M \rightarrow \infty} \frac{e^{\sigma \log(M+1)}}{e^{n \log n - n}} \leq \lim_{M \rightarrow \infty} \frac{e^{3n}}{e^{n \log n}} = 0. \quad (4.10)$$

□

Next, we introduce a class of sequences  $\mu$  with qualitatively the same algebraic decay as  $\nu_\sigma$ , which will allow us to derive complexity bounds on the set  $\Lambda_\epsilon(\mu)$ .

**Definition 4.7.** Let  $c = (c_1, c_2)$  and  $\sigma = (\sigma_1, \sigma_2) \in \mathbb{R}_+^2$ . A sequence  $\mu \in c_0$  is  $(c, \sigma)$ -algebraic, if

$$0 < c_1 m^{-\sigma_1} \leq \mu_m \leq c_2 m^{-\sigma_2} \quad \forall m \in \mathbb{N}_0.$$

In particular, the sequence  $\nu_\sigma$  in (4.6) is  $(c, (\sigma, \sigma))$ -algebraic for  $\sigma > 0$ .

**Lemma 4.8.** *Let  $\mu \in c_0$  be  $(c, \sigma)$ -algebraic. Then there exists a  $C > 0$  independent of  $\epsilon$  such that for all  $\epsilon > 0$*

$$0 \neq \alpha \in \Lambda_\epsilon(\mu) \implies \text{nnz}(\alpha) \leq C \log M_\epsilon(\mu).$$

*Proof.* First note that  $M_\epsilon(\mu) > 0$  if and only if  $\Lambda_\epsilon(\mu) \notin \{\emptyset, \{0\}\}$ . Secondly,  $M_\epsilon(\mu)$  is monotonically increasing in  $\epsilon$  given  $\mu \in c_0$  and furthermore

$$M_\epsilon(\mu) \rightarrow \infty \iff \epsilon \rightarrow 0$$

as  $\mu$  is a monotone null sequence. By assumption,  $\mu$  satisfies

$$c_1 m^{-\sigma_1} \leq \mu_m \leq c_2 m^{-\sigma_2} \quad \forall m \in \mathbb{N}_0$$

with some fixed  $c_1, c_2 > 0$  and  $\sigma_1 \geq \sigma_2 > 0$ . This implies

$$\epsilon > \mu_{M_\epsilon(\mu)+1} \geq c_1 (M_\epsilon(\mu) + 1)^{-\sigma_1}.$$

Further, as  $\mu$  is strictly decreasing with  $\mu_1 < 1$ , we may choose  $c_2 = 1$ , possibly with a decreased  $\sigma_2$ . Suppose now  $\text{nnz}(\alpha) \geq \left(1 + \frac{\sigma_1}{\sigma_2}\right) \log M_\epsilon(\mu)$  for some  $\alpha \in \mathbb{N}_0^{\mathbb{N}}$ . Then, using Lemma 4.6,

$$\begin{aligned} \mu^\alpha \leq (\lceil \text{nnz}(\alpha) \rceil!)^{-\sigma_2} &\leq \left( \left\lceil \left(1 + \frac{\sigma_1}{\sigma_2}\right) \log M_\epsilon(\mu) \right\rceil! \right)^{-\sigma_2} \\ &= o\left( \left( (M_\epsilon(\mu) + 1)^{1 + \frac{\sigma_1}{\sigma_2}} \right)^{-\sigma_2} \right) \\ &= o\left( \frac{1}{c_1} (M_\epsilon(\mu) + 1)^{-\sigma_2} (c_1 (M_\epsilon(\mu) + 1)^{-\sigma_1}) \right) \\ &= o\left( \frac{1}{c_1} (M_\epsilon(\mu) + 1)^{-\sigma_2} \epsilon \right) \quad \text{as } M \rightarrow \infty. \end{aligned}$$

With

$$\frac{1}{c_1} (M_\epsilon(\mu) + 1)^{-\sigma_2} \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0$$

we have shown that there exists  $\epsilon_0 > 0$  such that for all positive  $\epsilon \leq \epsilon_0$  it holds

$$\text{nnz}(\alpha) \geq \left(1 + \frac{\sigma_1}{\sigma_2}\right) \log M_\epsilon(\mu) \implies \mu^\alpha < \epsilon$$

and hence

$$0 \neq \alpha \in \Lambda_\epsilon(\mu) \implies \text{nnz}(\alpha) \leq \left(1 + \frac{\sigma_1}{\sigma_2}\right) \log M_\epsilon(\mu).$$

To complete the proof, define  $C$  as

$$C = 1 + \frac{\sigma_1}{\sigma_2} + \sup_{\epsilon > \epsilon_0} \max_{0 \neq \alpha \in \Lambda_\epsilon(\mu)} \frac{\text{nnz}(\alpha)}{\log M_\epsilon(\mu)}.$$

Note that the supremum is taken over a finite number of quantities, since there exist only a finite number of distinct sets  $\Lambda_\epsilon(\mu)$  with  $\epsilon > \epsilon_0 > 0$ .  $\square$

**Corollary 4.9.** *Let  $\mu \in c_0$  be  $(c, \sigma)$ -algebraic. Then there exists a constant  $C > 0$  such that*

$$\max_{\alpha \in \Lambda_\epsilon(\mu)} \text{nnz}(\alpha) \leq C \log |\Lambda_\epsilon(\mu)|$$

for any  $\epsilon > 0$  with  $\Lambda_\epsilon(\mu) \neq \emptyset$ .

*Proof.* Suppose  $\mu \in c_0$  is  $(c, \sigma)$ -algebraic. With Lemma 4.8 and Lemma 4.4 we note that for  $\epsilon > 0$  and  $0 \neq \alpha \in \Lambda_\epsilon(\mu)$  there holds

$$\text{nnz}(\alpha) \lesssim \log M_\epsilon(\mu) \leq \log |\Lambda_\epsilon(\mu)|.$$

$\square$

**Definition 4.10.** A one-parameter family  $\{\Lambda_\epsilon\}_{\epsilon > 0}$  of sets  $\Lambda_\epsilon$  of finite cardinality is of *log-linear complexity* w.r.t.  $\epsilon$  if there exist constants  $C > 0$  and  $k > 0$  independent of  $\epsilon$  such that  $\forall \epsilon > 0$  the set  $\Lambda_\epsilon$  can be constructed with the computational effort (e.g., time and memory) bounded above by

$$C |\Lambda_\epsilon| \log^k |\Lambda_\epsilon|.$$

Further analysis shall be based on

**Assumption 4.11 (On the memory model).** We assume that one memory location is able to store one real number of finite but sufficient precision. Further we assume the performance of data structures *linked list*, and *sorted set* in accordance with the C++98 standard (see [30]).

**Lemma 4.12.** *For any given  $(c, \sigma)$ -algebraic sequence  $\mu \in c_0$  the index set  $\Lambda_\epsilon(\mu)$  is of log-linear complexity w.r.t.  $\epsilon$ .*

*Proof.* The sets  $\Lambda_\epsilon(\mu)$  are indeed of log-linear complexity w.r.t.  $\epsilon$  in terms of memory: each index  $\alpha \in \Lambda_\epsilon(\mu)$  can be stored in the sparse format

$$\text{SPARSE}(\alpha, \mu^\alpha) \equiv (\{(m, \alpha_m) \mid \alpha_m \neq 0\}, \mu^\alpha) \quad (4.11)$$

as a sorted set of pairs (ordered by the first component) together with a real number. This format requires about  $(1 + 2 \cdot \text{nnz}(\alpha)) + 1$  memory locations, but essentially linear in  $\text{nnz}(\alpha)$ . By Corollary 4.9, the total number of memory locations which are required to store the set  $\Lambda_\epsilon(\mu)$  is bounded by  $C|\Lambda_\epsilon(\mu)| \log |\Lambda_\epsilon(\mu)|$  where  $C > 0$  is independent of  $\epsilon$ .

For  $\delta > 0$  consider now the set

$$\Lambda_\delta^0(\mu) = \{\alpha \in \Lambda_\delta(\mu) \mid \alpha_1 = 0\}.$$

Observe that

$$\Lambda_\epsilon(\mu) = \bigcup_{n=0}^N \left( (n, 0, 0, \dots) + \Lambda_{\epsilon/\mu_1^n}^0 \right)$$

where  $N = \max_{n \in \mathbb{N}_0} \{\mu_1^n \geq \epsilon\}$ . This suggests the following algorithm.

**Algorithm 4.13.** Given  $\epsilon > 0$ ,  $\mu \in c_0$ , set  $\alpha_0 = (0, 0, \dots)$ .

- Set  $\Lambda \leftarrow \{\text{SPARSE}(\alpha_0, 1)\} \cup \text{ENUMERATE}(\alpha_0, 0, 1)$
- Function  $\text{ENUMERATE}(\alpha, m, \epsilon_\alpha)$  with local variables  $\alpha$ ,  $m$  and  $\epsilon_\alpha$  returns a set  $\lambda \subset \Lambda_\epsilon(\mu)$ 
  - Let  $\lambda = \emptyset$
  - If  $\mu_m \epsilon_\alpha < \epsilon$  return  $\lambda$
  - For  $n = 0, 1, \dots$ 
    - \* Set  $\lambda \leftarrow \lambda \cup \text{ENUMERATE}(\alpha, m + 1, \epsilon_\alpha)$
    - \* Set  $\alpha_m \leftarrow \alpha_m + 1$  and  $\epsilon_\alpha \leftarrow \mu_m \epsilon_\alpha$
    - \* If  $\epsilon_\alpha < \epsilon$  return  $\lambda$
    - \* Set  $\lambda \leftarrow \lambda \cup \{\text{SPARSE}(\alpha, \epsilon_\alpha)\}$
- Return  $\Lambda_\epsilon(\mu) = \Lambda$

In Algorithm 4.13, the function  $\text{ENUMERATE}$  is called either right before or right after the set  $\lambda$  is increased. Hence, the number of calls is bounded by  $2|\Lambda_\epsilon(\mu)|$ . At each call

- the local variable  $\alpha$  is created by duplication. By Corollary 4.9, this step requires  $\mathcal{O}(\log |\Lambda_\epsilon(\mu)|)$  operations.
- the set  $\lambda$  is increased at most twice. By assumption 4.11 we can store  $\lambda$  as a linked list of pointers hence requiring a constant effort for this operation.

In total, the computational effort is therefore bounded by  $\mathcal{O}(|\Lambda_\epsilon(\mu)| \log |\Lambda_\epsilon(\mu)|)$ .  $\square$



For  $1 \geq \epsilon > 0$  and an integer  $\ell$ , let a sequence

$$1 \equiv \epsilon_{-1} \geq \epsilon_0 \geq \epsilon_1 \geq \dots \geq \epsilon_\ell = \epsilon > 0$$

be given. With these values we associate a sequence of multiindex sets

$$\emptyset = \Lambda_{\epsilon_0}(\mu) \subset \Lambda_{\epsilon_1}(\mu) \subset \dots \subset \Lambda_{\epsilon_\ell}(\mu)$$

where the nestedness follows from Lemma 4.4, iii). This sequence yields the following partition of  $\Lambda_\epsilon(\mu)$  into  $\ell + 1$  sets:

$$\Lambda_\epsilon(\mu) = (\Lambda_{\epsilon_\ell}(\mu) \setminus \Lambda_{\epsilon_{\ell-1}}(\mu)) \dot{\cup} (\Lambda_{\epsilon_{\ell-1}}(\mu) \setminus \Lambda_{\epsilon_{\ell-2}}(\mu)) \dot{\cup} \dots \dot{\cup} (\Lambda_{\epsilon_0}(\mu) \setminus \Lambda_{\epsilon_{-1}}(\mu)). \quad (4.12)$$

**Lemma 4.14.** *Let  $\mu \in c_0$  be  $(c, \sigma)$ -algebraic for some  $\sigma > 0$ . Given a finite sequence of threshold parameters*

$$1 \equiv \epsilon_{-1} \geq \epsilon_0 \geq \epsilon_1 \geq \dots \geq \epsilon_\ell = \epsilon > 0$$

such that

$$\Lambda_{\epsilon_k}(\mu) \setminus \Lambda_{\epsilon_{k-1}}(\mu) \neq \emptyset \quad k = 0, 1, \dots, \ell,$$

the corresponding disjoint partition (4.12) can be constructed with computational effort bounded above by

$$C |\Lambda_\epsilon(\mu)| \log |\Lambda_\epsilon(\mu)|$$

with  $C$  independent of  $\epsilon$ .

*Remark 4.15.* Note that from any sequence of threshold parameters  $\{\epsilon_k > 0\}_{k \geq 0}$  accumulating at zero we can extract a subsequence satisfying the above assumptions.

*Proof.* As shown in the previous lemma,  $\Lambda_\epsilon(\mu)$  is of log-linear complexity w.r.t.  $\epsilon$  when the indices are stored in the SPARSE format which we assume in what follows. The set  $\Lambda_\epsilon(\mu)$  can then be quasi-sorted w.r.t. the quasi-ordering  $\leq_\mu$  (cf. Remark 4.3) using at most

$$C_{\text{sort}} |\Lambda_\epsilon(\mu)| \log |\Lambda_\epsilon(\mu)|$$

comparison operations using, e.g., the merge-sort algorithm [12]. Each comparison between two multiindices  $\alpha_1, \alpha_2 \in \Lambda_\epsilon(\mu)$  involves the computation of  $\mu^{\alpha_1}$  and  $\mu^{\alpha_2}$ , which are, however, already provided by the SPARSE format in constant time. Denote now

$$\{\alpha_1 \leq_\mu \alpha_2 \leq_\mu \dots \leq_\mu \alpha_{|\Lambda_\epsilon(\mu)|}\} = \Lambda_\epsilon(\mu).$$

Note that while the ordering within each  $(\Lambda_{\epsilon_k}(\mu) \setminus \Lambda_{\epsilon_{k-1}}(\mu))$  may be ambiguous, the following holds for  $\ell \geq m > n \geq 0$ :

$$\alpha \in \Lambda_{\epsilon_m}(\mu), \beta \in \Lambda_{\epsilon_n}(\mu) \quad \Rightarrow \quad \neg(\beta \leq_\mu \alpha)$$

which is a paraphrase of (iii), Lemma 4.4. The disjoint partition (4.12) is computed by iteration over the sorted set by

**Algorithm 4.16.** Let  $k = \ell$ . For each  $\alpha = \alpha_1, \alpha_2, \dots, \alpha_{|\Lambda_\epsilon(\mu)|}$ :

- if  $\mu^\alpha \geq \epsilon_{k-1}$ , set  $k \leftarrow k - 1$
- append  $\alpha$  to the linked list  $\Lambda_{\epsilon_k}(\mu) \setminus \Lambda_{\epsilon_{k-1}}(\mu)$ .

Given Assumption 4.11, Algorithm 4.16 admits the computational complexity bound

$$C_{\text{part}}|\Lambda_\epsilon(\mu)| \log |\Lambda_\epsilon(\mu)|.$$

Thus the claim follows with essentially  $C = C_{\text{sort}} + C_{\text{part}}$ .  $\square$

*Remark 4.17.* Algorithms 4.13 and 4.16 are exemplary. Depending on further usage of the multiindex set  $\Lambda_\epsilon(\mu)$ , data structures may be adapted. E.g.:

- to each computed multiindex  $\alpha$  we may assign the level  $\ell$  such that  $\alpha \in \Lambda_{\epsilon_\ell}(\mu) \setminus \Lambda_{\epsilon_{\ell-1}}(\mu)$ .
- instead of linked lists we may use sorted sets, allowing fast access without changing the overall complexity significantly.

## 5 Implementation and numerical examples

This section summarizes the sGFEM algorithm and provides numerical examples. The sGFEM algorithm may be outlined as

1. Given the mean field  $\mathbb{E}_a(\mathbf{x})$  and the covariance  $C_a(\mathbf{x}, \mathbf{x}')$ , compute the eigenpairs  $(\lambda, \varphi)$  of the Karhunen-Loève representation of the random field  $a$  up to order  $M$ ,
2. Given a maximum level  $L$ , compute for each  $0 \leq l_2 \leq L$ , the respective deterministic stiffness matrices.
3. Choose a threshold sequence  $\{\epsilon_{l_1}\}_{l_1=0}^L$  such that  $\Lambda_{\epsilon_{l_1}}(\mu) \sim 2^{\gamma l_1}$ , where  $\mu \in c_0$  is a suitably chosen sequence and  $\gamma > 0$ .
4. Compute the index sets  $\Lambda_{\epsilon_{l_1}}(\mu)$  by Algorithm 4.13 and generate the corresponding stochastic moment matrices (see Sect. 5.4 below).
5. sGFEM: discretize (3.30) in the sparse tensor product spaces  $V_L^\Gamma \hat{\otimes} V_L^D$ . Solve the resulting system by a (preconditioned) CG algorithm.
6. Postprocessing: derive the response statistics from the computed sGFEM solution.

These issues are addressed in more detail below.

### 5.1 Karhunen-Loève eigenpairs

The Karhunen-Loève eigenpairs (see Sect. 2.3) can be computed by standard finite element methods. In this paper we only consider analytically known Karhunen-Loève expansions and their tensorized versions.

*Example 5.1.* Let  $D = (-1, 1)$ ,  $\mathbb{E}_a(x) = 5 + x$ ,  $C_a(x, x') = \frac{\min\{x, x'\} + 1}{2} \in H^{1,1}(D \times D)$ . It is straight forward to verify that  $\tilde{\lambda}_m = \frac{8}{\pi^2(2m-1)^2}$ ,  $f_m(x) = \sin((x + 1)/\sqrt{2\tilde{\lambda}_m})$  are the corresponding Karhunen-Loève eigenpairs,  $m \geq 1$ . Note the algebraic decay with rate 2.

## 5.2 Wavelet basis

The construction of wavelets has already been presented in Sect. 3.3. We will employ two different strategies to incorporate multilevel spatial discretization into the solution process. First one is to assemble the FEM stiffness matrix directly in the wavelet basis. We will do so in Example 5.9, using the wavelets from Sect. 3.3 on a uniform grid. Second one is to assemble the stiffness matrix w.r.t. the hat function basis and apply a preconditioned CG algorithm to the resulting system, where the inverse wavelet transform (and its adjoint) act as the preconditioner. This approach is described in greater detail in [3] for our case of two spatial dimensions. Both lead to a well-conditioned linear system of equations (see [7]).

## 5.3 Threshold sequence

Based on Corollary A.4 and Assumption 2.5 we propose to choose  $\mu_m \propto (1+m)^{-t}$  in Algorithm 4.13, where  $t \in (0, s-1)$ . In consequence, the multiindices  $\alpha \in \Lambda_\epsilon(\mu)$  will be a superset of the multiindices corresponding to the largest Legendre coefficients  $u_\alpha$  in the  $H_0^1(D)$ -norm by (iv), Lemma 4.4. Moreover, since (4.7) is a one to one correspondence of the size of the index set  $\#\Lambda_\epsilon(\mu)$  and  $\epsilon$  for sufficiently small  $\epsilon$ , we can efficiently choose the sequence  $\{\epsilon_n\}_{n \geq 1}$  such that  $\#\Lambda_{\epsilon_n}(\mu) \sim 2^{\gamma n}$ , see (3.17).

## 5.4 Stochastic moment matrices

The variational formulation (2.24) reduces to the linear system of equations of the form

$$(A_0 \hat{\otimes} G_0 + \sum_{m \geq 1} A_m \hat{\otimes} G_m) \mathbf{u} = \mathbf{1}.$$

For details we refer to [7]. While the matrices  $\{A_m\}_{m \geq 0}$  can be computed by standard FEM techniques, the stochastic moment matrices  $\{G_m\}_{m \geq 1}$  exhibit a non-trivial sparsity pattern. We note that  $G_0$  is diagonal and

$$(G_m)_{\alpha\beta} = \int_{\Gamma} \sqrt{\lambda_m} y_m \mathcal{L}_{M,\alpha}(\mathbf{y}) \mathcal{L}_{M,\beta}(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y}$$

where  $\alpha, \beta \in \Lambda_\gamma(L)$  are two multiindices. Hence,  $\sum_{m \geq 0} (G_m)_{\alpha\beta}$  is non-zero if and only if  $\alpha$  and  $\beta$  differ by one in exactly one location, i.e., there holds  $|\alpha_m - \beta_m| = \delta_{mn}$  for some  $n \in \mathbb{N}$ .

**Definition 5.2.** Let  $\alpha$  and  $\beta$  be two multiindices. We call  $\alpha$  and  $\beta$  *neighbors* if there exists  $n \in \mathbb{N}$  such that  $|\alpha_m - \beta_m| = \delta_{mn}$  for all  $m \in \mathbb{N}$ .

Algorithm 4.13 can be easily extended to compute such neighborhood relations in an efficient manner, with the help of the following observations.

**Lemma 5.3.** Let  $\epsilon > 0$  and  $\mu \in c_0$ . Further, let  $\alpha, \beta \in \Lambda_\epsilon(\mu)$  be two neighboring multiindices, with  $\alpha_n + 1 = \beta_n$  for some  $n \in \mathbb{N}$ . Every  $\hat{\beta} \in \Lambda$  satisfying

$$m \leq n \quad \Rightarrow \quad \hat{\beta}_m = \beta_m$$

has a neighbor  $\hat{\alpha} \in \Lambda_\epsilon(\mu)$  with  $\hat{\alpha}_n + 1 = \hat{\beta}_n$ .

*Proof.* Since  $\mu_n < 1$  by definition of  $\mu$ ,  $\mu^{\hat{\beta}} \geq \epsilon$  implies  $\mu^{\hat{\alpha}} \geq \epsilon$ , where  $\hat{\alpha}$  is as above.  $\square$

Algorithm 4.13 also suggests the following terminology.

**Definition 5.4.** Let  $\alpha$  and  $\beta$  be two multiindices.  $\alpha$  is called *parent* of  $\beta$  and  $\beta$  is called *child* of  $\alpha$  if there exists  $n \in \mathbb{N}$  such that

$$\alpha_m = \beta_m \text{ for all } m < n, \quad \alpha_n < \beta_n \quad \text{and} \quad \alpha_m = 0 \text{ for all } m > n.$$

If  $\beta$  is a child of  $\alpha$ , it can only be derived from  $\alpha$  by Algorithm 4.13 through a sequence of modifications of the type  $\alpha_m \leftarrow \alpha_m + 1$  of  $\alpha$  and its children. Conversely, each such modification performed by Algorithm 4.13 identifies the original multiindex as a parent.

**Lemma 5.5.** Let  $\epsilon > 0$ ,  $\mu \in c_0$  and  $\alpha, \beta \in \Lambda_\epsilon(\mu)$  be neighbors with  $|\alpha_m - \beta_m| = \delta_{mn}$  for some  $n \in \mathbb{N}$ . Suppose  $\alpha$  is parent of  $\beta$ . Any child  $\hat{\beta} \in \Lambda_\epsilon(\mu)$  of  $\beta$  has exactly one neighbor  $\hat{\alpha} \in \Lambda_\epsilon(\mu)$  among the children of  $\alpha$  which are not children of  $\beta$ .

*Proof.* Uniqueness follows from the restriction that  $\hat{\alpha}$  may not be a child of  $\beta$  and thus already differs from  $\hat{\beta}$  in the  $n$ -th position. Existence is a direct consequence of Lemma 5.3.  $\square$

**Lemma 5.6.** Let  $\epsilon > 0$ ,  $\mu \in c_0$  be given. If  $\hat{\alpha} \in \Lambda_\epsilon(\mu)$  and  $\hat{\beta} \in \Lambda_\epsilon(\mu)$  are neighbors (without loss of generality  $\hat{\alpha}_n \preceq \hat{\beta}_n$ ) then there exist multiindices  $\alpha \in \Lambda_\epsilon(\mu)$ , parent of  $\hat{\alpha}$  and  $\beta \in \Lambda_\epsilon(\mu)$ , parent of  $\hat{\beta}$  satisfying the assumptions of the previous lemma.

*Proof.* Set

$$\alpha_m = \begin{cases} \hat{\alpha}_m, & m \leq n, \\ 0, & \text{else} \end{cases}$$

and similarly for  $\beta$ . Note that  $\alpha_m$  is the last common parent of  $\hat{\alpha}$  and  $\hat{\beta}$  (i.e., does not have any child which is parent to  $\hat{\alpha}$  and  $\hat{\beta}$ ).  $\square$

Lemma 5.6 implies that any neighboring relation, say between  $\hat{\alpha} \in \Lambda_\epsilon(\mu)$  and  $\hat{\beta} \in \Lambda_\epsilon(\mu)$ , is easily computable. Whenever the step  $\alpha_m \leftarrow \alpha_m + 1$  of Algorithm 4.13 is executed, all the children of the new multiindex  $\alpha'$  need to be associated with some of the children of the old multiindex (which are not children of  $\alpha'$ ). This is conveniently done using a tree structure to record the parenthood relations to subsequently traverse the tree as in Lemma 5.5. The neighborhood relations of a multiindex  $\alpha$  may be stored as a list of identifiers, e.g., pointers to the neighboring multiindices. It is easy to see that the computational complexity (time and memory requirements) of the neighborhood relations is linear in their count.

*Remark 5.7.* The ease with which the neighborhood relations may be computed for  $\Lambda_\epsilon(\mu)$  is crucially a consequence of Lemma 5.3. For more general multiindex sets, e.g., in the context of adaptivity, this construction fails.

Having the neighborhood relations at hand, it is easy to traverse and if necessary build the stochastic stiffness matrices  $\{G_m\}_{m \geq 1}$  in linear time.

## 5.5 Convergence rate estimation

The extrapolated estimated order of convergence is based on the fact that the Galerkin solution is an orthogonal projection in the  $L^2(H^1)$  norm. Thus for the purpose of the estimating the order of convergence we assume

$$\|u\|_{L^2_\rho(H^1_0)}^2 - \|u_M\|_{L^2_\rho(H^1_0)}^2 \sim (\dim(V_L^\Gamma \otimes V_L^D)^{-\alpha})^2$$

for the full tensor product and

$$\|u\|_{L^2_\rho(H^1_0)}^2 - \|\hat{u}_M\|_{L^2_\rho(H^1_0)}^2 \sim (\dim(V_L^\Gamma \hat{\otimes} V_L^D)^{-\beta})^2$$

for the sparse tensor product. Assuming equality up to a multiplicative constant, the three unknown quantities, i.e. the constant, the norm  $\|u\|_{L^2(H^1)}^2$  of the (unknown) exact solution and the order of convergence, are fit using three consecutive data points.

## 5.6 Examples

*Example 5.8.* Let  $D = (0, 1)^2$  be the unit square. The sequence  $\{\varphi_m\}_{m \geq 1}$  is chosen as the tensor product of the sequence from Example 5.1 ordered by magnitude of the resulting eigenvalues  $\{\tilde{\lambda}\}_{m \geq 1}$ . The eigenvalues  $\lambda_m$  are chosen as  $\lambda_m = \tilde{\lambda}_m^\theta$  with  $\theta = 2.5$ , hence the decay of  $\{\lambda_m\}_{m \geq 1}$  and of  $\{\|\psi_m\|_{L^\infty(D)}\}_{m \geq 1}$  is algebraic with rate  $s = 5$ , see Remark 2.6.

This implies the stochastic rate  $r = 1$  and thus  $\gamma = \frac{1}{2}$ , see Sect. 3.7. The expected value  $\mathbb{E}_a(\mathbf{x})$  is assumed to equal  $\mathbb{E}_a(\mathbf{x}) = x + 5$  and we set  $a_{\min} = 4$ . The right hand side is  $f \equiv 1$ .

In order to construct the index set  $\Lambda_{\epsilon_n}(\mu)$ , based on (4.1), we choose  $\mu_m = (r_m + \sqrt{1 + r_m^2})^{-1}$  where  $r_m := a_{\min}^*/(m\sqrt{\lambda_m})$ .

The reference solution consists of  $|V_L^\Gamma \otimes V_L^D| = 6'594'624$  degrees of freedom in the full tensor case, while  $|V_L^\Gamma \hat{\otimes} V_L^D| = 155'424$  in the sparse tensor case, where  $L = 6$ , see Table 5.1.

The relative residual tolerance for the conjugate gradient methods is set to  $10^{-10}$ . In both cases the conjugate gradient algorithm requires about 125 steps to converge on the finest level.

As shown in Fig. 5.1, there appears to be no significant loss in accuracy due to the usage of the sparse tensor approximation in place of the full tensor approximation, while the total number of degrees of freedom,  $\dim(V_L^\Gamma \hat{\otimes} V_L^D)$ , is substantially smaller than  $\dim(V_L^\Gamma \otimes V_L^D)$ . For comparison purposes, we also give the rate of convergence of the Monte Carlo Method in terms of the “number of degrees of freedom” (interpreted as number of MC samples times the number of FE degrees of freedom); the higher efficiency of the sparse deterministic-stochastic tensor Galerkin FEM is evident.

Sparse tensorization is therefore an essential tool to reduce the total number of degrees of freedom in order to handle random input data with slowly convergent Karhunen-Loève expansions efficiently.

In Fig. 5.1 we also show the expected Monte Carlo error as discussed in Sect. 5.7.

$\ell$	0	1	2	3	4	5	6	Total
$ W_\ell^\Gamma $	1	1	2	3	7	15	31	64
$ W_\ell^D $	36	85	320	1240	4880	19360	77120	103041

Table 5.1: Degrees of freedom in the detail spaces used in the computation of the reference solution for Example 5.8.

*Example 5.9.* Let  $D = (-1, 1)$ . Let the Karhunen-Loève eigenpairs be given by  $\lambda_m = (m + 1)^{-\sigma}$ ,  $\sigma = 2.2$  and  $\varphi_m(x) = f_m(x)$  from Example 5.1. The load is assumed to be  $f(x) = \exp(x)$  and the mean diffusion coefficient  $\mathbb{E}_a(x) = a_{\min} + 1 + \sin(\pi x)$  with  $a_{\min} = 10$ . We set  $\Lambda(n) = \Lambda_{\epsilon_n}(\mu)$  with the underlying sequence  $\mu_m = \sqrt{\lambda_m}/a_{\min}$  with  $\epsilon_n$  such that  $|\Lambda(n)| = \lceil 2^{\gamma n} \rceil$ ,  $\gamma = 3.5$  for  $0 \leq n \leq L$ . The maximal number of levels is  $L = 6$ . In Table 5.2 the resulting number of degrees of freedom is listed. The solution for  $L = 6$  has 1838 degrees of freedom in  $\Gamma$  which are distributed among  $M = 1594$  dimensions. The relative residual tolerance for the conjugate gradient methods is set to  $10^{-11}$ , requiring less than 50 steps to converge.

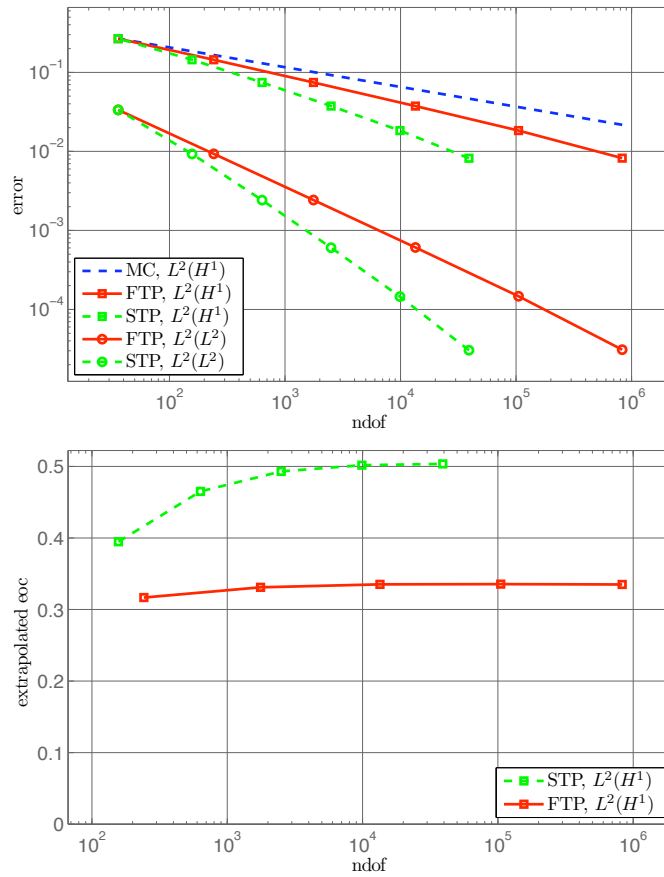


Figure 5.1: The performance of the sparse tensor approximation vs. the full tensor approximation in terms of the total number of degrees of freedom for Example 5.8.

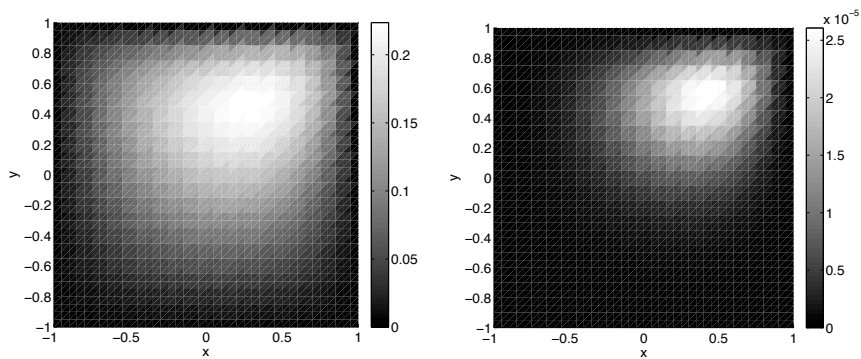


Figure 5.2: The response statistics for Example 5.8: mean and variance.

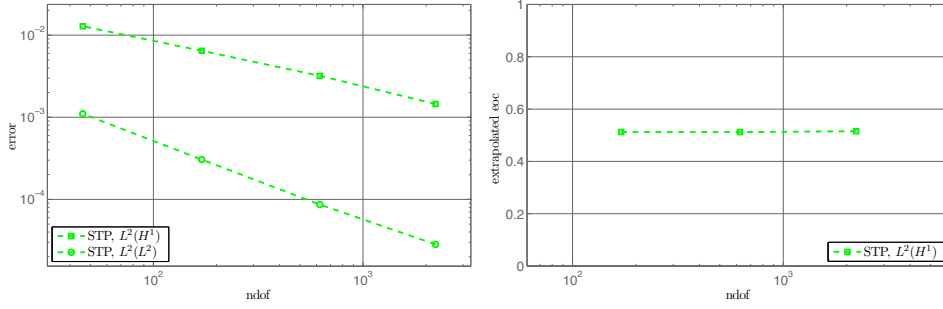


Figure 5.3: Convergence of the sparse tensor approximation in terms of the total number of degrees of freedom for Example 5.9.

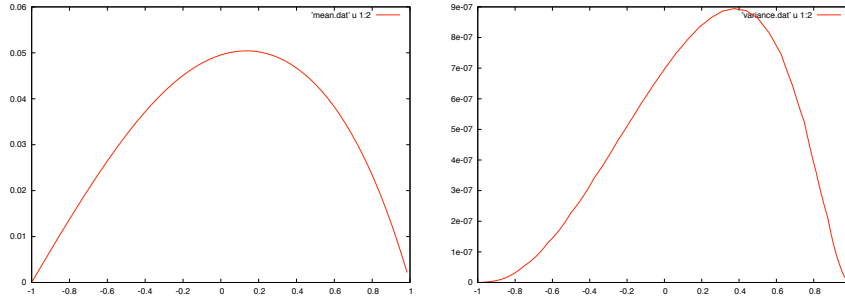


Figure 5.4: The response statistics for Example 5.9: mean and variance.

The total computation time is less than 15 min. on a laptop with 1.7GHz CPU and 2GB RAM.

*Remark 5.10.* The Karhunen-Loève eigenfunctions in this example are highly oscillatory compared to the mesh width. This requires an accurate integration routine to compute the deterministic stiffness matrices. In this case we use the adaptive 7+8 Gauss-Kronrod quadrature (see [18]) on each element with relative tolerance of  $10^{-10}$ .

*Remark 5.11.* The Karhunen-Loève eigenvalue decay rate of  $\sigma = 2.2$  with normed eigenfunctions  $\|\varphi_m\|_{L^\infty(D)} = 1$  entails  $s \leq 1.1$  in 2.28. As the assumptions of Proposition 3.1 and Proposition 3.5 do not hold in this case, no choice for  $\gamma$  is suggested by these propositions. The value  $\gamma = 3.5$  which we chose in (3.16) and in Proposition 3.1 is an empirical one, chosen such that stable convergence was observed in the numerical experiments.

The convergence plots are shown in Fig. 5.3. The computed response statistics are displayed in Fig. 5.4.

$\ell$	0	1	2	3	4	5	6	Total
$ W_\ell^I $	1	2	9	30	108	375	1313	1838
$ W_\ell^D $	3	2	4	8	16	32	64	129

Table 5.2: Degrees of freedom of the reference solution for Example 5.9.

## 5.7 Discussion

We have presented and analyzed a new class of sparse tensor discretizations for elliptic PDEs with stochastic coefficients. It is based on hierarchic Galerkin discretizations of multilevel type in the physical domain  $D$  and of “polynomial chaos” type in the probability domain  $\Gamma$ ; approximations are obtained by Galerkin projection on sparse tensor products of these discretizations. We note here that the same idea works not only in the stochastic Galerkin setting but also for stochastic collocation methods, where a hierarchic sequence of collocation operators are used in conjunction with a multilevel discretization in space, see [6]. We analyze and implement a strategy to localize hierarchic sets of data-adapted active gPC modes to be used in the sparse tensor Galerkin projections.

Our algorithm is valid for general sequences of type (4.2), i.e., our Assumption on the probability density (see Sect. 3.2) is not essential at this point. Due to the monotonicity properties of the index sets  $|\Lambda_\epsilon(\mu)|$  (see Lemma 4.4) the hierarchic moment matrices  $G_m$  can be computed efficiently, as outlined in Sect. 5.4. This issue needs to be addressed since for algebraic Karhunen-Loève decay the trivial algorithm to compute a subset of  $\{G_m\}_{m \geq 1}$  is of cubic complexity in the total number of nonzero entries, and of quadratic complexity if the multiindices are stored in a sparse format.

The presented numerical experiments illustrate the developed theory for sparse tensor approximation, see Sect. 3.7. The convergence rates as shown in Fig. 5.1 are supported by the theory. Indeed, the full tensor product approximation exhibits a rate of  $(d/p + 1/r)^{-1} = \frac{1}{3}$  where  $d = 2$  is the spatial dimension,  $p = 1$  the piecewise polynomial degree of the ansatz functions in the physical space and  $r = 1$  is the maximally allowed approximation rate, see Example 5.8, while the sparse tensor product shows the predicted rate  $\min\{r, p/d\} = \frac{1}{2}$ , see Eqn. (3.43). Both methods are superior to the straightforward Monte Carlo approach for which we expect a convergence rate of  $(d/p + 1/(1/2))^{-1} = \frac{1}{4}$  in the total number of degrees of freedom (i.e.,  $N_{\text{tot}} = N_D N_\Omega$ ) as shown in Fig. 5.1. In this analysis we have assumed that the resulting linear systems can be solved in linear time, as is the case for multiscale bases from Sect. 3.3. As is shown in Prop. 3.5, when the “stochastic regularity”  $r$  is high, the gap in efficiency between all three methods will increase even further.

In our second numerical Example 5.9 we have considered a “deep” Karhunen-Loève expansion for a algebraically decaying sequence (2.9) of Karhunen-Loève eigenvalues: in this case, our Algorithm 4.13 selects gPC “modes” which are distributed anisotropically in a possibly very high dimensional parameter space.

## A Best $N$ -term approximation of gpc expansions

To prove Proposition 3.1 we will first have to provide a bound on the Legendre coefficients  $u_\alpha(\mathbf{x})$  defined in (3.10). Eventually, we then establish some results on approximation of sequences of such coefficients and complete the Appendix by proving Proposition 3.1.

### A.1 Estimates of Legendre coefficients

We establish bounds on the coefficients  $u_\alpha(\mathbf{x})$  defined in (3.10) by complex variable techniques. To this end, it is necessary to consider the parametric, deterministic problem (2.24) also for complex parameter vectors  $\mathbf{z} = (z_1, z_2, \dots)$  with  $z_m \in \mathbb{C}$ .

To prove the bounds, we define the complex domain



$$\mathcal{U}(\mathbf{r}) = \bigotimes_{m=1}^{\bar{M}} \mathcal{U}_m(r_m) \subset \mathbb{C}^{\bar{M}}, \quad (\text{A.1})$$

where  $\mathbf{r} = (r_1, \dots, r_{\bar{M}}) \in \mathbb{R}^{\bar{M}}$  is a given vector of radii, with ‘‘component’’ domains

$$\mathcal{U}_m(r_m) = \{z_m \in \mathbb{C} : \text{dist}(z_m, \Gamma_m) \leq r_m\}. \quad (\text{A.2})$$

We introduce the following notation: given  $m \in \mathbb{N}$  and  $\mathbf{y} = (y_1, y_2, \dots) \in \Gamma$ , we define  $\mathbf{y}_m^* = (y_1, \dots, y_{m-1}, y_{m+1}, \dots)$  and write  $\mathbf{y} = (\mathbf{y}_m^*, y_m) \in \Gamma_m^* \times \Gamma_m$  if the dependence of  $\mathbf{y}$ -dependent quantities on the coordinate  $y_m$  is to be emphasized. Likewise, we denote by  $\mathbf{z} = (\mathbf{z}_m^*, z_m) \in \mathcal{U}(\mathbf{r}) = \mathcal{U}_m^*(\mathbf{r}_m^*) \times \mathcal{U}_m(r_m)$  the corresponding partitioning of the complex vector  $\mathbf{z}$ . We introduce the real-valued *parametric bilinear form*

$$B(\mathbf{y}; u_M(\mathbf{y}, \cdot), v) = \int_D a_M(\mathbf{y}, x) \nabla_x u_M \cdot \nabla_x v \, d\mathbf{x}, \quad \mathbf{y} \in \Gamma, \quad u, v \in V^D. \quad (\text{A.3})$$

With this definition, we have in (2.25)

$$b_M(u_M, v) = \int_{\mathbf{y} \in \Gamma} B(\mathbf{y}; u_M(\mathbf{y}, \cdot), v(\mathbf{y}, \cdot)) \rho(\mathbf{y}) \, d\mathbf{y}, \quad u, v \in C^0(\bar{\Gamma}, V^D).$$

Next, we extend  $B(\mathbf{y}; u, v)$  in (A.3) to complex-valued arguments: abusing notation, we denote by  $V^D = H_0^1(D)$  also the space of complex-valued functions which belong to  $H_0^1(D)$  and define, for  $\mathbf{z} \in \mathcal{U}(\mathbf{r})$  the complex extension of the (truncated) diffusion coefficient  $a_M$  by

$$a_M(\mathbf{z}, \mathbf{x}) = \mathbb{E}_a(\mathbf{x}) + \sum_{m=1}^M \psi_m(\mathbf{x}) z_m \quad (\text{A.4})$$

with  $\psi_m(\mathbf{x})$  as in (2.27). For  $u, v \in V^D$  the sesquilinear extension of  $B(\mathbf{y}; u, v)$  in (A.3) is then given by

$$B(\mathbf{z}; u, v) = \int_D a_M(\mathbf{z}, x) \nabla_x u \cdot \overline{\nabla_x v} \, d\mathbf{x}, \quad (\text{A.5})$$

The complex valued parametric problem is then given by

$$-\nabla_x \cdot (a_M(\mathbf{z}, \mathbf{x}) \nabla_x u_M(\mathbf{z}, \mathbf{x})) = f(\mathbf{x}) \quad \text{in } D, \quad u_M(\mathbf{z}, x)|_{x \in \partial D} = 0. \quad (\text{A.6})$$

and, in weak form: Find  $u_M \in L^2(\mathcal{U}(\mathbf{r}), V^D)$  s.t.

$$B(\mathbf{z}; u_M(\mathbf{z}, \cdot), v) = (f, v) \quad \forall v \in V^D, \quad \forall \mathbf{z} \in \mathcal{U}(\mathbf{r}) \quad (\text{A.7})$$

We will now prove that for certain  $\mathbf{r} = (r_1, \dots, r_M)$  problem (A.7) is uniquely solvable and its solution is ‘coordinate-wise’ analytic in  $\mathcal{U}(\mathbf{r})$ :

**Lemma A.1.** *Let  $a_M$  be given as in (A.4). Define the vector of radii  $\mathbf{r}(\delta) = (r_1(\delta), \dots, r_M(\delta))$  by*

$$r_m(\delta) := \frac{\tau_m}{C(\delta)m^{1+\delta}} \quad \text{where } 0 < \tau_m < \frac{a_{\min}^*}{\|\psi_m\|_{L^\infty(D)}} \quad (\text{A.8})$$

and

$$C(\delta) := \sum_{m \geq 1} \frac{1}{m^{1+\delta}}, \quad a_{\min}^* = \min_{\mathbf{x} \in D} \mathbb{E}_a(\mathbf{x}) - a_{\min}. \quad (\text{A.9})$$

Then, for any  $\mathbf{z} \in \mathcal{U}(\mathbf{r}(\delta))$  with  $\delta > 0$ , (A.7) admits a unique solution  $u_M \in L^2(\mathcal{U}(\mathbf{r}(\delta)), V^D)$ . Moreover,  $u_M(\mathbf{z}_m^*, z_m, \cdot) : \mathcal{U}(\mathbf{r}(\delta)) \rightarrow V^D$  is, for fixed coordinates  $\mathbf{z}_m^* \in \mathcal{U}_m^*(r_m^*(\delta))$ , a  $V^D$ -valued analytic function of  $z_m \in U_m(r_m(\delta))$

*Proof.* To prove the existence and uniqueness of a solution to (A.7) we show that the real part  $\text{Re}a(\mathbf{z}, \mathbf{x})$  of the random field, as in (A.4), with  $\mathbf{z} \in \mathcal{U}(\mathbf{r}(\delta))$  is bounded away from zero:

$$\begin{aligned} \text{Re}a_M(\mathbf{z}, \mathbf{x}) &= \mathbb{E}_a(\mathbf{x}) + \sum_{m \geq 1} \psi_m(\mathbf{x}) \text{Re}z_m \\ &\geq \mathbb{E}_a(\mathbf{x}) - \sum_{m \geq 1} b_m |z_m| \\ &\geq \mathbb{E}_a(\mathbf{x}) - \sum_{m \geq 1} b_m \frac{a_{\min}^*}{C(\delta) b_m m^{1+\delta}} \\ &\geq \min_{\mathbf{x} \in D} \mathbb{E}_a(\mathbf{x}) - a_{\min} = a_{\min}^* > 0 \end{aligned}$$

In a similar fashion we can conclude that

$$|a_M(\mathbf{z}, \mathbf{x})| \leq a_{\max}^* := \|\mathbb{E}_a\|_{L^\infty(D)} + a_{\min}^*$$

for  $\mathbf{z} \in \mathcal{U}(\mathbf{r}(\delta))$ . It follows that

$$\text{Re}B(\mathbf{z}; u, u) \geq \frac{1}{a_{\min}^*} \|u\|_{V^D}^2 \quad \text{and} \quad |B(\mathbf{z}; u, v)| \leq a_{\max}^* \|u\|_{V^D} \|v\|_{V^D} \quad (\text{A.10})$$

hence the unique solvability of (A.7) follows by Lax-Milgram. Choosing  $v = u(\mathbf{z}, \cdot)$  in (A.7), we find with (A.10) that

$$\begin{aligned} a_{\min}^* \|u(\mathbf{z}, \cdot)\|_{V^D}^2 &\leq |B(\mathbf{z}; u(\mathbf{z}, \cdot), u(\mathbf{z}, \cdot))| = |(f, u(\mathbf{z}, \cdot))| \\ &\leq \|f\|_{(V^D)^*} \|u(\mathbf{z}, \cdot)\|_{V^D} \end{aligned}$$

from where it follows that

$$\forall \mathbf{z} \in \mathcal{U}(\mathbf{r}(\delta)) : \quad \|u(\mathbf{z}, \cdot)\|_{V^D} \leq \frac{\|f\|_{(V^D)^*}}{a_{\min}^*}. \quad (\text{A.11})$$

We now establish analyticity with respect to  $z_m$  by a power series argument. To this end, we consider the parametric problem (A.6) for parameters  $(\mathbf{z}_m^*, y_m) \in \mathcal{U}_m^*(\mathbf{r}_m^*(\delta)) \times \Gamma_m$ . Differentiating (A.6) with respect to the real-valued parameter  $y_m$ , we obtain that for any  $\nu \in \mathbb{N}$  and every  $v \in V^D$ , the partial derivative  $\partial_{y_m}^\nu u_M((\mathbf{z}_m^*, y_m); \cdot) \in V^D$  solves the problem:

$$B((\mathbf{z}_m^*, y_m); \partial_{y_m}^\nu u, v) = -\nu \int_D \psi_m(x) \nabla_x (\partial_{y_m}^{\nu-1} u_M) \cdot \overline{\nabla_x v} dx \quad \forall v \in V.$$

Choosing here  $v = \partial_{y_m}^\nu u_M(\mathbf{z}_m^*, y_m, \cdot) \in V^D$  gives

$$\begin{aligned} a_{\min}^* \|\partial_{y_m}^\nu u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D}^2 &\leq |B((\mathbf{z}_m^*, y_m); \partial_{y_m}^\nu u_M, \partial_{y_m}^\nu u_M)| \\ &= \nu \left| \int_D \psi_m(\mathbf{x}) \nabla_x (\partial_{y_m}^{\nu-1} u_M(\mathbf{z}_m^*, y_m, \mathbf{x})) \cdot \overline{\nabla_x \partial_{y_m}^\nu u_M(\mathbf{z}_m^*, y_m, \mathbf{x})} dx \right| \\ &\leq \nu \|\psi_m\|_{L^\infty(D)} \|\partial_{y_m}^{\nu-1} u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D} \cdot \|\partial_{y_m}^\nu u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D} \end{aligned}$$

whence we obtain, for any  $\mathbf{z}_m^* \in \mathcal{U}_m^*(\mathbf{r}_m^*(\delta))$  and any  $m, \nu \in \mathbb{N}$ ,

$$\|\partial_{y_m}^\nu u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D} \leq \nu \frac{b_m}{a_{\min}^*} \|\partial_{y_m}^{\nu-1} u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D}.$$

Iterating this estimate gives with (A.11)

$$\begin{aligned} \|\partial_{y_m}^\nu u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D} &\leq \nu! \left(\frac{b_m}{a_{\min}^*}\right)^\nu \|u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D} \\ &\leq \frac{\nu!}{a_{\min}^*} \left(\frac{b_m}{a_{\min}^*}\right)^\nu \|f\|_{(V^D)^*}. \end{aligned} \quad (\text{A.12})$$

Now define, for arbitrary  $m \in \mathbb{N}$  and arbitrary, fixed  $y_m \in \Gamma_m$ , the formal power series  $\mathcal{U}_m(r_m(\delta)) \ni z_m \rightarrow u_M(\mathbf{z}_m^*, z_m, \cdot) \in V^D$  by

$$u(\mathbf{z}_m^*, z_m, \cdot) = \sum_{k=0}^{\infty} \frac{(z_m - y_m)^k}{k!} \partial_{y_m}^k u_M(\mathbf{z}_m^*, y_m, \cdot). \quad (\text{A.13})$$

To estimate its convergence radius, we use (A.12):

$$\begin{aligned} \|u_M(\mathbf{z}_m^*, z_m, \cdot)\|_{V^D} &\leq \sum_{k=0}^{\infty} \frac{|z_m - y_m|^k}{k!} \|\partial_{y_m}^k u_M(\mathbf{z}_m^*, y_m, \cdot)\|_{V^D} \\ &\leq \frac{\|f\|_{(V^D)^*}}{a_{\min}^*} \sum_{k=0}^{\infty} \left(b_m \frac{|z_m - y_m|}{a_{\min}^*}\right)^k. \end{aligned}$$

This infinite sum and, hence, the series (A.13) converges in  $V^D$  if

$$b_m \frac{|z_m - y_m|}{a_{\min}^*} < 1.$$

Since  $y_m \in \Gamma_m$  was arbitrary, the series (A.13) converges for any  $y_m \in \Gamma_m = [-1, 1]$  and, by a continuation argument, for any  $z_m \in \mathbb{C}$  satisfying  $\text{dist}(z_m, \Gamma_m) < a_{\min}^*/b_m$ . This completes the proof.  $\square$

Now we can prove analyticity of  $\mathcal{U}(\mathbf{r}(\delta)) \ni \mathbf{z} \rightarrow u(\mathbf{z}, \cdot) \in V^D$ .

**Theorem A.2.** *Given the assumptions of Lemma A.1, for any  $M$  the solution  $u_M(\mathbf{z}, \cdot)$  to (A.7) is analytic in the complex domain  $\mathcal{U}(\mathbf{r}(\delta))$  with  $\mathbf{r}(\delta)$  defined as in (A.8).*

*Proof.* This is a direct consequence of Hartogs' Theorem, see e.g. [19], Theorem 2.2.8  $\square$

This enables us to find bounds on the Legendre ‘‘coefficients’’  $u_\alpha \in H_0^1(D)$  due to Cauchy’s integral formula for analytic functions of several variables ([19], Theorem 2.2.1)

**Lemma A.3.** *Let the assumptions of Proposition 2.3 hold. If  $u_M$  solves the parametric deterministic problem (2.24) and is expanded in a Legendre chaos series as in (3.9), then, for any  $\delta_1 > 0$  the Legendre coefficients satisfy*

$$\|u_\alpha\|_{H_0^1(D)} \leq C \left( \prod_{m \in \text{supp}(\alpha)} 2(2\alpha_m + 1) \right) \eta(\delta_1)^{-\alpha} \quad (\text{A.14})$$

where  $C > 0$  is independent of  $M, \alpha$  with  $\eta(\delta_1) = (\eta_1(\delta_1), \dots, \eta_M(\delta_1)) \in \mathbb{R}^M$  and

$$\eta_m(\delta) := r_m(\delta) + \sqrt{1 + r_m(\delta)^2} \quad (\text{A.15})$$

and  $r_m(\delta)$  as in (A.8).

*Proof.* Note that for simplicity of notation we do not indicate in the following the dependence of  $\eta$  on  $\delta_1$ . The  $V^D$ -valued Legendre “coefficients” are given by

$$u_\alpha = \left( \prod_{m \geq 1} 2\alpha_m + 1 \right) \int_{\Gamma} u_M(\mathbf{y}) \mathcal{L}_\alpha(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y} \quad (\text{A.16})$$

where the integral is a Bochner integral w.r.t. the probability measure  $d\rho(\mathbf{y}) = \rho(\mathbf{y})d\mathbf{y}$  and the equality has to be read in  $V^D$ .

In the following we will use the abbreviations  $S = \text{supp}(\alpha) \subset \{1, \dots, M\}$  and  $\bar{S} = \{1, \dots, M\} \setminus S$  (omitting the dependence on  $\alpha$  for reasons of readability of the formulae below) and, for a set  $G \subset \{1, \dots, M\}$ , we will write  $\Gamma_G = \prod_{m \in G} \Gamma_m$  and by  $\mathbf{y}_G \in \Gamma_G$  the respective extraction from  $\mathbf{y}$ . Since, by Theorem A.2,  $u_M$  is analytic in  $\mathcal{U}(\mathbf{r}(\delta_1))$  for any  $\delta_1 > 0$  we may use Cauchy’s integral formula along  $\mathcal{E}_S = \prod_{m \in \text{supp}(\alpha)} \mathcal{E}_m$ , where  $\mathcal{E}_m$  denotes the (Bernstein-)ellipse in  $\mathcal{U}_m(r_m(\delta_1))$  with foci at  $\pm 1$  and the sum of the semiaxes

$$\eta_m := r_m(\delta_1) + \sqrt{1 + r_m(\delta_1)^2} > 1 \quad (\text{A.17})$$

to rewrite the Legendre coefficients (A.16) as

$$\begin{aligned} u_\alpha &= \left( \prod_{m \in S} \frac{2\alpha_m + 1}{2\pi i} \right) \int_{\Gamma} \mathcal{L}_\alpha(\mathbf{y}_S) \oint_{\mathcal{E}_S} \frac{u_M(\mathbf{z}_S, \mathbf{y}_{\bar{S}})}{(\mathbf{z}_S - \mathbf{y}_S)^{\mathbf{1}}} d\mathbf{z}_S \rho(\mathbf{y}) d\mathbf{y} \quad (\text{A.18}) \\ &= \left( \prod_{m \in S} \frac{2\alpha_m + 1}{2\pi i} \right) \int_{\Gamma_{\bar{S}}} \oint_{\mathcal{E}_S} u_M(\mathbf{z}_S, \mathbf{y}_{\bar{S}}) \int_{\Gamma_S} \frac{\mathcal{L}_\alpha(\mathbf{y}_S)}{(\mathbf{z}_S - \mathbf{y}_S)^{\mathbf{1}}} \rho(\mathbf{y}) d\mathbf{y}_S d\mathbf{z}_S d\mathbf{y}_{\bar{S}} \quad (\text{A.19}) \end{aligned}$$

By [31] §4.9, the innermost integral in (A.19) is a representation for the Legendre polynomials of the second kind,

$$\mathcal{Q}_\alpha(\mathbf{z}_S) = \int_{\Gamma} \frac{\mathcal{L}_\alpha(\mathbf{y}_S)}{(\mathbf{z}_S - \mathbf{y}_S)^{\mathbf{1}}} \rho_S(\mathbf{y}_S) d\mathbf{y}_S. \quad (\text{A.20})$$

Furthermore, if, for  $m \in S$ , we substitute  $z_m = \frac{1}{2}(w_m + w_m^{-1})$  (Joukowski transformation) with  $|w_m| = \eta_m$ , the Legendre polynomials of the second kind can be expanded like (cf. [14], Lemma 12.4.6)

$$Q_{\alpha_m}\left(\frac{1}{2}(w_m + w_m^{-1})\right) = \sum_{k=\alpha_m+1}^{\infty} \frac{q_{\alpha_m k}}{w_m^k}$$

with  $|q_{\alpha_m k}| \leq \pi$ . Hence,

$$|\mathcal{Q}_\alpha(\mathbf{z}_S)| \leq \prod_{m \in S} \sum_{k=\alpha_m+1}^{\infty} \frac{\pi}{\eta_m^k} = \prod_{m \in S} \pi \frac{\eta_m^{-\alpha_m-1}}{1 - \eta_m^{-1}}. \quad (\text{A.21})$$

From [8], see also [41] Chapter V.5, we learn that for any Bochner integrable function  $g \in L^2((\mathcal{M}, \mu); V)$ , with  $(\mathcal{M}, \mu)$  a measure space and  $V$  a Banach space it holds

$$\left\| \int_{\mathcal{M}} g d\mu \right\|_V \leq \int_{\mathcal{M}} \|g\|_V d\mu$$

From this, together with (A.21) and (A.19), we obtain

$$\begin{aligned}
\|u_\alpha\|_{V^D} &= \left\| \left( \prod_{m \in S} \frac{2\alpha_m + 1}{2\pi i} \right) \int_{\Gamma_{\overline{S}}} \oint_{\mathcal{E}_S} u_M(\mathbf{z}_S, \mathbf{y}_{\overline{S}}) \mathcal{Q}_\alpha(\mathbf{z}_S) d\mathbf{z}_S \rho_{\overline{S}}(\mathbf{y}_{\overline{S}}) d\mathbf{y}_{\overline{S}} \right\|_{V^D} \\
&\leq \left( \prod_{m \in S} \frac{2\alpha_m + 1}{2\pi i} \right) \int_{\Gamma_{\overline{S}}} \oint_{\mathcal{E}_S} \|u_M(\mathbf{z}_S, \mathbf{y}_{\overline{S}})\|_{V^D} \mathcal{Q}_\alpha(\mathbf{z}_S) d\mathbf{z}_S \rho_{\overline{S}}(\mathbf{y}_{\overline{S}}) d\mathbf{y}_{\overline{S}} \\
&\leq \left( \prod_{m \in S} \frac{2\alpha_m + 1}{2\pi} \text{Len}(\mathcal{E}_m) \right) \|u_M(\mathbf{z})\|_{L^\infty(\mathcal{E}_S \times \Gamma_{\overline{S}}, V^D)} \max_{\mathcal{E}_S} |\mathcal{Q}_\alpha| \\
&\leq \left( \prod_{m \in S} \frac{2\alpha_m + 1}{2\pi} \text{Len}(\mathcal{E}_m) \right) \prod_{m \in S} \pi \frac{\eta_m^{-\alpha_m - 1}}{1 - \eta_m^{-1}} \\
&\leq C \left( \prod_{m \in S} 2(2\alpha_m + 1) \right) \eta^{-\alpha} \|u_M(\mathbf{z})\|_{L^\infty(\mathcal{E}_S \times \Gamma_{\overline{S}}, V^D)} \tag{A.22}
\end{aligned}$$

where the last estimate holds due to the fact that  $\text{Len}(\mathcal{E}_m) \leq 4\eta_m$ .  $\square$

It follows from Remark 2.6 that for  $\delta_1 > 0$  sufficiently small there exists a constant  $C(s, d, \delta_1) > 0$  such that

$$\forall m \in \mathbb{N} : \quad \eta_m^{-1} \leq C_1(s, d, \delta_1) m^{-(s-1-\delta_1)} \tag{A.23}$$

where  $s$  is defined in (2.29). Since, furthermore, for any  $\delta_2 > 0$  there exists a constant  $C_2(\delta_2)$  s.t.  $2\alpha_m + 1 \leq C_2(\delta_2) \eta_m^{\delta_2 \alpha_m}$ , it follows from the definition of  $\eta_m$  together with (A.23)

**Corollary A.4.** *Given the assumptions of Lemma A.3. For any  $0 < t < s - 1$  there exists a constant  $C_3(t, d) > 0$ , independent of  $m, M, \alpha$  s.t. by defining*

$$\tilde{\eta}_m^{-\alpha_m} := \begin{cases} 1 & m = 0 \\ C_3(t, d) \eta_m^{-(1-\delta_2)\alpha_m} & m > 0 \end{cases} \tag{A.24}$$

the Legendre coefficients satisfy

$$\|u_\alpha\|_{H_0^1(D)} \leq C \prod_{m \in \text{supp}(\alpha)} \tilde{\eta}_m^{-\alpha_m} \tag{A.25}$$

where the constant  $C > 0$  does not depend on  $M, \alpha$ .

Note that it follows from (A.23), that  $\tilde{\eta}_m^{-\alpha_m} \rightarrow 0$  if either  $m$  or  $\alpha_m$  tend to infinity, hence the product in (A.25) remains finite even if  $|\text{supp}(\alpha)| \rightarrow \infty$ .

## A.2 $\tau$ -summability of the Legendre coefficients

To obtain best  $N$ -term approximation rates, suitable regularity of the function under consideration is required [16]. This means that in our case that the Legendre coefficients  $u_\alpha$  have to be  $\tau$ -summable.

**Lemma A.5.** *Let the assumptions of Proposition 2.3 hold. If  $u_M$  solves (2.24) and is expanded in a Legendre series as in (3.9), then for any  $\tau > \frac{1}{s-1}$ , where  $s$  is given by (2.29), then  $\{\|u_\alpha\|_{H_0^1(D)} : \alpha \in \mathbb{N}_c^{\mathbb{N}}\} \in \ell^\tau$  and there holds:*

$$\|\{\|u_\alpha\|_{H_0^1(D)} : \alpha \in \mathbb{N}_c^{\mathbb{N}}\}\|_{\ell^\tau}^\tau = \sum_{\alpha \in \mathbb{N}_0^M} \|u_\alpha\|_{H_0^1(D)}^\tau \leq C \exp\left(\frac{C_3 \|\eta^{-1}\|_{\ell^\tau}^\tau}{(1 - \eta_1^{-\tau})}\right) \tag{A.26}$$

with  $C = C(\tau, d) > 0$  independent of  $M, C_3$  as in (A.24) and with  $\tilde{\eta}_m$  defined as in (A.15).

*Proof.* From (A.23) it follows that  $\|(\eta_m^{-1})_m\|_{\ell_\tau} < \infty$  if  $\tau > \frac{1}{s-1}$ . Using (A.24) together with the summation formula for geometric series we obtain for any finite  $M < \infty$  that

$$\begin{aligned} \sum_{\alpha \in \mathbb{N}_0^M} \|u_\alpha\|_{H_0^1(D)}^\tau &\leq C \sum_{m=1}^M \sum_{\alpha_m=0}^{\infty} \prod_{m=1}^M \tilde{\eta}_m^{-\alpha_m \tau} \\ &= C \prod_{m=1}^M \left(1 + \frac{C_3 \eta_m^{-1}}{1 - \eta_m^{-\tau}}\right) \\ &\leq C e^{\frac{C_3}{1-\eta_1^{-\tau}} \sum_{m=1}^M \eta_m^{-\tau}} \end{aligned}$$

where we used  $1 + x \leq e^x$  if  $x \geq 0$ . Letting  $M \rightarrow \infty$  gives (A.26).  $\square$

### A.3 Proof of Proposition 3.1

Recall the definitions of  $\Pi(\Lambda; H)$ ,  $\sigma_N$  and  $\mathcal{A}_r(H)$  from (3.13)-(3.15) with  $H$  denoting a Hilbert space, e.g.  $H_0^1(D)$  in our examples. The weak  $\ell_\tau$  spaces  $\ell_\tau^w$  consist of all sequences  $(f_k)_{k \in \mathbb{N}}$  for which it holds

$$|(f_k)_k|_{\ell_\tau^w} := \sup_{k \geq 1} k^{1/\tau} f_k^* < \infty \quad (\text{A.27})$$

where  $(f_k^*)_k$  denotes the decreasing rearrangement of  $(|f_k|)_k$ . Since  $\ell_\tau \subset \ell_\tau^w$ , this enables us now to characterize the space  $\mathcal{A}_r(H)$ : In fact, a generalization of a well-known result of nonlinear approximation theory to Bochner spaces (see e.g. [16], Theorem 4) states that a function  $u \in L^2(\Gamma; H)$  is in  $\mathcal{A}_r(H)$  if and only if the  $H$ -norm of its Legendre coefficients  $u_\alpha$  form a sequence in  $\ell_\tau^w$  for  $\tau := (r + \frac{1}{2})^{-1}$ . In fact, denote by  $(u_k^*)_{k \in \mathbb{N}}$  the decreasing rearrangement of  $(\|u_\alpha\|_H)_\alpha$ . Due to

$$\begin{aligned} |u|_{\mathcal{A}_r(H)}^2 &= \sup_{N \geq 1} N^{2r} \sum_{k > N} \|u_k^*\|_H^2 \\ &\stackrel{(\text{A.27})}{\leq} |(\|u_\alpha\|_H)_\alpha|_{\ell_\tau^w}^2 \sup_{N \geq 1} N^{2r} \sum_{k > N} k^{-2r-1} \\ &\leq |(\|u_\alpha\|_H)_\alpha|_{\ell_\tau^w}^2 \sup_{N \geq 1} \int_N^\infty \frac{N^{2r}}{x^{2r+1}} dx \\ &\leq |(\|u_\alpha\|_H)_\alpha|_{\ell_\tau^w}^2 \frac{1}{2r+1} \end{aligned}$$

and

$$(u_{2k-1}^*)^2 \leq (u_{2k}^*)^2 \leq \frac{1}{k} \sum_{l=k+1}^{2k} (u_l^*)^2 \leq \frac{1}{k} \sigma_k(u)^2 \leq |u|_{\mathcal{A}_r}^2 k^{-2r-1} \quad (\text{A.28})$$

we have that

$$|u|_{\mathcal{A}_r(H)} \lesssim 2\tau |(\|u_\alpha\|_H)_\alpha|_{\ell_\tau^w} \lesssim \frac{2}{2r+1} |u|_{\mathcal{A}_r(H)}$$

Since Lemma A.5 ensures that the Legendre coefficients of the solution  $u_M$  belong to  $\ell_\tau^w$  for  $\tau > (s-1)^{-1}$  it follows that the solution  $u_M$  is in  $\mathcal{A}_r(H)$  for  $\tau = (r+1/2)^{-1} > (s-1)^{-1}$  and hence for  $0 < r < s - 3/2$ . Proposition 3.1 then follows from the definition of  $\mathcal{A}_r(H)$  (3.15).  $\square$

## References

- [1] Robert A. Adams and John J.F. Fournier. *Sobolev Spaces*. Pure and applied mathematics. Academic Press, second edition, 2003.
- [2] Robert J. Adler. *The geometry of random fields*. Wiley series in probability and mathematical statistics. John Wiley, Chinchester, 1981.
- [3] R. Andreev. Sparse wavelet-galerkin methods for stochastic diffusion problems. Bachelor thesis, ETHZ, 2007.
- [4] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Num. Anal.*, 45(3):p. 1005–1034, 2007.
- [5] I. Babuška, R. Tempone, and G.E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Num. Anal.*, 42(2):p. 800–825, 2002.
- [6] M. Bieri. A sparse composite collocation finite element method for elliptic sPDEs. submitted.
- [7] M. Bieri and Ch. Schwab. Sparse high order FEM for elliptic sPDEs. Technical Report 2008-22, Seminar for Applied Mathematics, ETH Zürich, 2008. *Comp. Meth. Appl. Mech. Engrg.*, in press.
- [8] Salomon Bochner. Integration von Funktionen deren Werte die Elemente eines Vektorraumes sind. *Fundamenta Mathematicae*, 20:p. 262–276, 1935.
- [9] E. Rodney Canfield, Paul Erdős, and Carl Pomerance. On a problem of Oppenheim concerning “factorisatio numerorum”. *Journal of Number Theory*, 17:1–28, 1983.
- [10] A. Cohen, R. DeVore, and Ch. Schwab. Convergence rates of best N-term stochastic Galerkin FE-approximations for a class of elliptic sPDEs. Technical Report 2009-02, Seminar for Applied Mathematics, ETH Zürich, 2009.
- [11] Albert Cohen. *Numerical analysis of wavelet methods*, volume 32 of *Studies in Mathematics and its Applications*. Elsevier, Amsterdam, 2003.
- [12] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, 2nd edition, 2001.
- [13] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series In Applied Mathematics*. SIAM, Philadelphia, 1992.
- [14] Philipp J. Davis. *Interpolation and approximation*. Introductions to higher mathematics. Blaisdell Publishing Company, 1963.
- [15] M.-K. Deb, I. Babuška, and J.T. Oden. Solution of stochastic partial differential equations using galerkin finite element techniques. *Comput. Meth. Appl. Mech. Engrg.*, 190:p. 6359–6372, 2001.
- [16] R.A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:p. 51–150, 1998.
- [17] P. Frauenfelder, Ch. Schwab, and R.-A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Comp. Methods Appl. Mech. Engrg.*, 194:p. 205–228, 2005.

- [18] Walter Gautschi. *Orthogonal polynomials. Computation and Approximation*. Numerical Mathematics and Scientific Computation. Oxford University Press Inc., 2004.
- [19] Lars Hörmander. *An Introduction to Complex Analysis in Several Variables*. North Holland Mathematical Library. North Holland, 3rd edition, 1990.
- [20] G.E. Karniadakis, C.-H. Su, D. Xiu, D. Lucor, Ch. Schwab, and R.-A. Todor. Generalized polynomial chaos solution for differential equations with random inputs. Technical Report 2005-01, Seminar for Applied Mathematics, ETH Zürich, 2005.
- [21] H.G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comp. Meth. Appl. Mech. Engrg.*, 194:p. 1295–1331, 2005.
- [22] Hoang Nguyen. *Finite element wavelets for solving partial differential equations*. PhD thesis, Universiteit Utrecht, 2005.
- [23] F. Nobile, R. Tempone, and C.G. Webster. An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Num. Anal.*, 46(5):p. 2411–2442, 2008.
- [24] F. Nobile, R. Tempone, and C.G. Webster. A sparse grid stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Num. Anal.*, 46(5):p. 2309–2345, 2008.
- [25] A. Oppenheim. On an arithmetic function. *J. London Math. Soc.*, s1-2(2):123–130, 1926.
- [26] Martin Schlather. Introduction to positive definite functions and to unconditional simulation of random fields. Technical Report ST-99-10, Department of Mathematics and Statistics, Lancaster University UK, 1999.
- [27] Ch. Schwab and R.-A. Todor. Sparse finite elements for stochastic elliptic problems. *Numerische Mathematik*, 95:p. 707–734, 2003.
- [28] Ch. Schwab and R.-A. Todor. Sparse finite elements for stochastic elliptic problems - higher order moments. *Computing*, 71:p. 43–63, 2003.
- [29] Ch. Schwab and R.-A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):p. 100–122, 2006.
- [30] Bjarne Stroustrup. *The C++ Programming Language (Special 3rd Edition)*. Addison-Wesley Professional, February 2000.
- [31] Gabor Szegő. *Orthogonal polynomials*. AMS Colloquium Publications, 1939.
- [32] G. Szekeres and P. Turan. Ueber das zweite hauptproblem der "factorisatio numerorum". *Acta Litt. Sci. Szeged*, 6:143–154, 1933.
- [33] R.-A. Todor and Ch. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA Journal of Numerical Analysis*, 27:p. 232–261, 2007.
- [34] Hans Triebel. *Interpolation theory, Function spaces, Differential operators*. Johann Ambrosius Barth Verlag, second edition, 1995.



- [35] Erik Vanmarcke. *Random fields*. The MIT press, Cambridge-Mass. & London, 1983.
- [36] T. von Petersdorff and Ch. Schwab. Sparse finite element methods for operator equations with stochastic data. *Applications of Mathematics*, 52:p. 145–180, 2006.
- [37] X. Wan and G.E. Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM J. Sci. Comput.*, 28(3):p. 901–928, 2006.
- [38] X. Wan and G.E. Karniadakis. Error control in multi-element polynomial chaos method for elliptic problems with random coefficients. *Communications in Computational Physics*, (5/2–4):pp. 793–820, 2009.
- [39] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Scientific Computing*, 27(3):p. 1118–1139, 2005.
- [40] D. Xiu and G.E. Karniadakis. The wiener-asky polynomial chaos for stochastic differential equations. *SIAM J. Scientific Computing*, 24(2):p. 619–644, 2002.
- [41] Kôzaku Yosida. *Functional Analysis*. Springer Classics in Mathematics. Springer, reprint of the 1980 edition, 1995.

# Research Reports

No.	Authors/Title
09-07	<i>M. Bieri, R. Andreev, C. Schwab</i> Sparse tensor discretization of elliptic sPDEs
09-06	<i>A. Moiola</i> Approximation properties of plane wave spaces and application to the analysis of the plane wave discontinuous Galerkin method
09-05	<i>D. Kressner</i> A block Newton method for nonlinear eigenvalue problems
09-04	<i>R. Hiptmair, J. Li, J. Zou</i> Convergence analysis of Finite Element Methods for $H(\text{curl};\Omega)$ -elliptic interface problems
09-03	<i>A. Chernov, T. von Petersdorff, C. Schwab</i> Exponential convergence of $hp$ quadrature for integral operators with Gevrey kernels
09-02	<i>A. Cohen, R. DeVore, C. Schwab</i> Convergence rates of best $N$ -term Galerkin approximations for a class of elliptic sPDEs
09-01	<i>B. Adhikari, R. Alam, D. Kressner</i> Structured eigenvalue condition numbers and linearizations for matrix polynomials
08-32	<i>R. Sperb</i> Optimal bounds in reaction diffusion problems with variable diffusion coefficient
08-31	<i>R. Hiptmair</i> Discrete compactness for $p$ -version of tetrahedral edge elements
08-30	<i>H. Heumann, R. Hiptmair</i> Extrusion contraction upwind schemes for convection-diffusion problems
08-29	<i>N. Hilber, C. Schwab, C. Winter</i> Variational sensitivity analysis of parametric Markovian market models
08-28	<i>K. Schmidt, S. Tordeux</i> Asymptotic modelling of conductive thin sheets
08-27	<i>R. Hiptmair, P.R. Kotiuga, S. Tordeux</i> Self-adjoint curl operator
08-26	<i>N. Reich</i> Wavelet compression of anisotropic integrodifferential operators on sparse tensor product spaces