Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

A posteriori error control for stochastic Galerkin FEM with high-dimensional random parametric PDEs

Martin Eigel, Christian Merdon

submitted: February 12, 2025

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
E-Mail: martin.eigel@wias-berlin.de
christian.merdon@wias-berlin.de

No. 3174
Berlin 2025

2020 *Mathematics Subject Classification.* 35J25, 35R60, 41A10, 41A63, 42C10, 65N30, 65N50.

Key words and phrases. Parametric partial differential equations, stochastic Galerkin methods, finite elements, a posteriori error estimation, adaptive refinement.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

A posteriori error control for stochastic Galerkin FEM with high-dimensional random parametric PDEs

Martin Eigel, Christian Merdon

Abstract

PDEs with random data are investigated and simulated in the field of Uncertainty Quantification (UQ), where uncertainties or (planned) variations of coefficients, forces, domains and boundary conditions in differential equations formally depend on random events with respect to a pre-determined probability distribution. The discretization of these PDEs typically leads to high-dimensional (deterministic) systems, where in addition to the physical space also the (often much larger) parameter space has to be considered. A proven technique for this task is the Stochastic Galerkin Finite Element Method (SGFEM), for which a review of the state of the art is provided. Moreover, important concepts and results are summarized. A special focus lies on the a posteriori error estimation and the derivation of an adaptive algorithm that controls all discretization parameters. In addition to an explicit residual based error estimator, also an equilibration estimator with guaranteed bounds is discussed. Under certain mild assumptions it can be shown that the successive refinement produced by such an adaptive algorithm leads to a sequence of approximations with guaranteed convergence to the true solution. Numerical examples illustrate the practical behavior for some common benchmark problems. Additionally, an adaptive algorithm for a problem with a non-affine coefficient is shown. By transforming the original PDE a convection-diffusion problem is obtained, which can be treated similarly to the standard affine case.

1 Introduction

High-dimensional parametric PDEs have become an important applied mathematical research area in the last decade mainly because of the popularity the field of Uncertainty Quantification (UQ) has experienced. On the one hand, this is due to the importance of incorporating uncertainties in the simulation of real-world problems, e.g. in the engineering and natural sciences. On the other hand, introducing dependence on random parameters is a natural extension of differential equations, opening up a broad analytical and methodological research areas by combining concepts from numerical, functional and stochastic analysis.

Classically, two conceptual approaches co-exist peacefully: statistical sampling methods based on the universal workhorse of Monte Carlo (MC) simulations, and functional approximations such as Stochastic Collocation (SC) and Stochastic Galerkin FEM (SGFEM), which are rooted in a function space perspective. To mention just a few aspects and provide initial references for the interested reader, in the last decade progress with sampling methods has for instance be achieved with multilevel MC (MLMC), multi-index MC (MIMC), and quasi-MC (QMC) methods [TSGU13, HS17, HANT16]. Stochastic Collocation is one of the most popular functional representations, based on a sparse grid polynomial interpolation in parameter space. Seminal works are [BTZ05, BNT10] with later extensions e.g. in [EST18] and related multilevel quadrature methods for quantities of interest in [HPS16]. Another common functional method is an extension of classical finite element methods to the parametric setting, called Stochastic (Galerkin) FEM (SGFEM), initially presented in[GS91, GK96] with further developments in [KM03, Mat08]. The tensor structure of the solution space

DOI 10.20347/WIAS.PREPRINT.3174 Berlin 2025

of parametric PDEs lends itself to higher-order tensor compression for which [MZ12, EHL$^+$13] are early examples.

To obtain a better numerical efficacy, the discretization has to be adjusted iteratively. The challenge consists of the derivation of reliable error estimators for all discretization parameters and a balanced refinement procedure, taking all error components into account. As with deterministic FEM, also stochastic FEM provide a solid basis for this task due to the projection property of the discrete solution. This can be exploited to derive reliable residual based error estimators for the parametric truncation error, see [Git13] and in combination with the physical approximation error see [EGSZ14, EGSZ15, EM16a], all of which lead to sparse generalized (Legendre) polynomial chaos (GPC) representations. Similar ASGFEM approaches with low-rank tensor representations were shown in [EPS17] and (for the first time) for the lognormal (Hermite chaos) setting in [EMPS20]. Generalization to approximate Galerkin projections by sample-based nonlinear least-squares tensor reconstructions can be found in [EFHT23a, EST22]. A strand of work using hierarchical error estimators with sparse GPC SGFEM was started in [BPS12, BPS14, BS16]. It was later extended in [BPRR19b, BPR22] to goal-oriented error estimation and multilevel adaptive grids, which are also used in [EGSZ14]. Convergence results for these ASGFEM are derived in [EGSZ15] for the residual based error estimator and in [BPRR19a, BPR22] for the hierarchical one. An early work on a posteriori error estimation with small uncertainties is [GNP16] and recently the standard residual based error estimator was transferred to SC methods [GN18], which before only were based on a priori information or heuristic indicators during runtime [NTTT16]. With a reliable error estimator readily available, convergence results for adaptive SC became feasible as shown in [FS21, EST22].

In this review chapter for ASGFEM with sparse GPC expansion, we recall a standard linear benchmark model[1] with affine parameter dependence of the coefficient on the parameters. To make it concrete, consider some domain $D \subset \mathbb{R}^d$ and let (Ω, Σ, π) be a σ-finite probability space. The coefficient $a(\omega, x)$ is assumed to be isotropic for the sake of simplicity. Moreover, we require $a \in L^\infty(\Omega \times D)$ to be bounded, strictly positive[2] and to be represented in terms of independent and identically distributed (i.i.d.) random variables $Y_m \sim U[-1, 1]$ in an expansion of the form

$$a(\omega, x) = a_0 + \sum_{m=1}^{\infty} a_m(x) Y_m(\omega),$$

which is affine in the “random coordinates” Y_m. The model problem with random operator is given by

$$-\operatorname{div}(a(\omega, x)\nabla u(\omega, x)) = f(x) \quad \text{for } (\omega, x) \in \Omega \times D. \tag{1}$$

Here, the coefficient $a(\omega, x)$ and consequently also the solution $u(\omega, x)$ are measurable functions $\Omega \times D \to \mathbb{R}$. The representation of $a(\omega, x)$ can be based on artificial or factual (“expert opinion”) assumptions regarding the characterization of the considered random field, or it can be based on real-world measurements from which empirical statistical data of the actual field is determined. Given an artificial or empirical covariance operator, a random field representation for use in the PDE model can then be computed as a Karhunen-Loève expansion (KLE) as discussed in Section 2.1. Although the numerical experiments in the later sections use an artificial KLE, we nevertheless cite some theoretical results regarding the decay behavior with respect to regularity properties of some prescribed covariance, which should be of general interest. Instead of working with random variables, the image of the (countable infinite) random vector $(Y_m)_{m\geq 1}$ is

[1] usually called the “affine Darcy problem”

[2] An exception to this setting is presented with the log-transformed problem in Section 7.

given by $\Gamma = [-1,1]^{\mathbb{N}}$ with an associated product probability measure π. A basis for Γ is given by the so-called generalized polynomial chaos, i.e. polynomials orthogonal with respect to π, which in this setting is determined by the joint (uniform) distribution of the model parameters. The density of the respective tensorized Legendre polynomials is a result of the famous Cameron-Martin theorem [EMSU12, LPS14]. We recall properties required for constructing the parameter space approximations in Section 2.2. Section 3 begins with a definition of the affine linear model problem used in this work and its weak formulation in Section 3.1. With this at hand, Section 3.2 introduces the SGFEM for that problem, in particular the notation for the discrete spaces that are used in the Galerkin projection. Moreover, the algebraic structure of the discretization, its tensor structure and low-rank compression as well as some central theoretical (convergence) results are recalled in Section 3.3, Section 3.4 and Section 3.5, respectively. Section 4 reviews an error splitting into subresiduals in Section 4.1 and defines an explicit residual based error estimator in Section 4.2, which is known to be efficient and reliable, albeit (as usual with FEM) with unknown multiplicative constants. We then also introduce a guaranteed and constant-free flux equilibration error estimator, which is state-of-the-art in deterministic FEM in Section 4.3 and discuss the handling of the potentially infinite dimensional stochastic boundary in Section 4.4. Section 5 describes the adaptive algorithm, which uses the error estimators for the two error components (approximation and parametric truncation) of the preceding section to successively refine all discretization parameters. For this adaptive algorithm, convergence results are known, which we briefly recall. Numerical experiments are depicted in Section 6, where the performance of the adaptive algorithm is illustrated on the unit square and L-shaped domain for different decay rates of the KLE modes. In Section 7, we briefly touch upon the lognormal problem, consisting of the same linear PDE as before but with the affine coefficient replaced by a lognormal field. We present the numerical convergence of an adaptive algorithm akin the one developed for the affine case, which becomes possible because of a suitable transformation, resulting again in an equation with an affine coefficient.

2 Random field expansion and polynomial chaos

For the discretization of the random PDE (1), a parameter dependent representation of the random data is required, which we discuss in this section. The most common approach is the Karhunen-Loève expansion (KLE, also known as proper orthogonal decomposition), which is an affine expansion in terms of independent random variables with respect to the eigenfunctions of the covariance integral operator. It separates the deterministic and stochastic variables optimally in a mean square sense. In the same reference, the stochastic Galerkin FEM (SGFEM) with polynomial chaos, i.e. univariate polynomials that are orthonormal with respect to the joint distribution of the data random variables, was initially presented. For the random field representation, different other approaches can be used that might exhibit favorable properties from a theoretical or practical point of view, see e.g. [BC24, BV22].

2.1 Representation of random fields

Random fields admissible as data in the framework we consider here have to adhere to certain (in many cases non-restricting) properties. Inevitably, a parametric representation such as the KLE has to be available if they are to be used in a SGFEM discretization. A vital notion (not unique to the KLE) is the separation of spatial and random variables and by this of a random coordinate system spanned by independent random variables with known product distribution. We recapitulate central results of the KLE in the following, examine the crucial decay properties and its numerical approximation. For further details, we refer to [ST06b, LPS14, FST05, Loè77].

2.1.1 The Karhunen-Loève expansion

For the coefficient $a(\omega, x)$ in (1) we assume its mean field and two-point correlation to be specified, i.e.,

$$E_a(x) := \int_\Omega a(\omega, x)\mathrm{d}\pi(\omega), \qquad C_a(x, x') := \int_\Omega a(\omega, x)a(\omega, x')\mathrm{d}\pi(\omega) \quad \text{for } x, x' \in D \subset \mathbb{R}^d. \tag{2}$$

Note that this implies the covariance of the field

$$V_a(x, x') = C_a(x, x') - E_a(x)E_a(x'), \tag{3}$$

the smoothness of which directly determines how many terms are required in a parametric expansion to reach a certain accuracy in a certain norm. The KLE is optimal with respect to the L^2-norm.

To derive the KLE of $a(\omega, x)$, assume that it has bounded variance $a \in L^2(\Omega \times D)$. It follows that $V_a \in L^2(D \times D)$. For $u \in L^2(D)$ its covariance operator

$$V_a : L^2(D) \to L^2(D), \qquad (V_a u)(x) := \int_D V_a(x, x')u(x')\,\mathrm{d}x' \tag{4}$$

is symmetric, non-negative and compact, resulting in a countable sequence of eigenpairs $(\lambda_m, a_m)_{m\geq 1}$ with λ_m converging to 0 from above for $m \to \infty$ and such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. The KLE of the random field $a(\omega, x)$ is then given by

$$a(\omega, x) = E_a(x) + \sum_{m=1}^{\infty} \sqrt{\lambda_m} a_m(x) X_m(\omega). \tag{5}$$

Here, the sequence $(X_m)_{m\geq 1}$ of centered independent random variables

$$\int_\Omega X_m(\omega)\,\mathrm{d}\pi(\omega) = 0, \qquad \int_\Omega X_m(\omega)X_n(\omega)\,\mathrm{d}\pi(\omega) = \delta_{mn} \quad \text{for all } m, n \geq 1, \tag{6}$$

defines a coordinate system of the random space that is used in the reformulation of the stochastic into a parametric problem in Section 3.1. The KLE (5) converges in $L^2(\Omega \times D)$ since

$$\sum_{m=1}^{\infty} \lambda_m = \int_\Omega \int_D (a(\omega, x) - E_a(x))^2\,\mathrm{d}\pi(\omega)\,\mathrm{d}x < \infty. \tag{7}$$

A stronger uniform convergence is obtained if $(a_m)_{m\geq 1}$ and $(X_m)_{m\geq 1}$ are uniformly bounded in their respective spaces and if $\sum_{m\geq 1} \lambda_m < \infty$.

In any numerical method, the number of terms of the expansion (5) has to be finite. It hence is crucial to understand the decay properties of the sequence of eigenvalues. A common assumption is piecewise analyticity of V_a, meaning that there exists a finite decomposition of hypercubes $\overline{D} \subseteq \bigcup_{j=1}^{J} \overline{D}_j$ and $V|_{D_j \times D_{j'}}$ admits an analytic continuation in a neighborhood of $D_j \times D_{j'}$ for any $j \neq j'$. For the eigenvalues, it then holds that for some $c > 0$

$$0 \leq \lambda_m \lesssim e^{-cm^{1/d}} \qquad \text{for } m \geq 1.$$

For practical purposes, i.e. when the covariance is defined explicitly or based on empirical estimates from measurement data, the following results are of interest.

Proposition 2.1 ([FST05] Propositions 2.4 & 2.5)**.**

1 *Assume the Gaussian covariance kernel*

$$V_a(x,x') = \sigma^2 \exp(|x-x'|^2/(\gamma|D|)^2)$$

with standard deviation $\sigma > 0$ and correlation length $\gamma > 0$. It admits an analytic continuation in the whole complex space $\mathbb{C}^d$ with eigenvalue decay given by

$$0 < \lambda_m \leq \sigma^2 \frac{\gamma^{-m^{1/d}-2}}{\Gamma(\frac{1}{2}m^{1/d})}.$$

2 *For a less (piecewise Sobolev $H^{p,0}$, $p \geq 0$) regular covariance, e.g., with $\delta \in [0,1)$,*

$$V_a(x,x') = \sigma^2 \exp(-\frac{|x-x'|^{1+\delta}}{\gamma^{1+\delta}|D|^{1+\delta}}),$$

algebraic decay rates are obtained,

$$0 \geq \lambda_m \lesssim m^{-p/q} \qquad \text{for } m \geq 1.$$

The next result provides decay rates for the pointwise error.

Proposition 2.2 ([FST05] Proposition 2.6)**.** *Assume V_a to be piecewise smooth on a decomposition of D as above. For the ordered sequence of eigenpairs (λ_m, ϕ_m) with normalized eigenfunction $||\phi_m||_{L^2(D)} = 1$, for any $s > 0$ and any multi-index $\alpha \in \mathbb{N}^d$ it holds that*

$$||\partial^\alpha \phi_m||_{L^\infty(D_j)} \leq C(s,\alpha,V_a)|\lambda_m|^{-s}, \qquad \text{for all } j = 1,\dots,J,\ m \geq 1,$$

for some $C(s,\alpha,V_a) > 0$.

With respect to the results cited above, a theorem of Widom relates the asymptotic decay of the radial spectral density of an isotropic covariance function (such as the common Whittle-Matèrn covariance) and the decay rate of the respective eigenvalues. The interested reader is advised to consult [LPS14, Loè77] and references therein to learn more about spectral properties of covariance operators and the associated KLE.

Remark 2.3. Typically, the eigenpairs $(\lambda_m, a_m)_{m\geq 1}$ of the covariance (Hilbert-Schmidt) operator V_a (4) satisfying the Fredholm integral equation

$$V_a a_m = \lambda_m a_m, \qquad m \in \mathbb{N} \tag{8}$$

do not exhibit a known closed form and hence have to be computed numerically. Assuming a finite element basis $V_n = \operatorname{span}\{\varphi_i : i = 1,\dots,N_h\}$ as in Section 3.2 and setting $\mathbf{\Phi}_h(x) = (\varphi_1(x),\dots,\varphi_{N_h}(x))$, a projection of (8) onto this discrete space yields the generalized eigenvalue problem (EVP)

$$\mathbf{W a_{m,h}} = \mathbf{\lambda_{m,h} M a_{m,h}} \tag{9}$$

with $a_{m,h}(x) = \mathbf{\Phi}_h(x)\mathbf{a}_{m,h}$ and

$$W_{i,j} := (\varphi_i, V_a\varphi_j)_{L^2(D)} = \iint_{D\times D} \varphi_i(x)V_a(x,x')\varphi_j(x')\,\mathrm{d}x\,\mathrm{d}x',$$
$$M_{i,j} := (\varphi_i, \varphi_j)_{L^2(D)} = \int_D \varphi_i(x)\varphi_j(x)\,\mathrm{d}x \qquad \text{for } i,j = 1,\dots,N_h.$$

Note that $\mathbf{W}$ is symmetric positive semi-definite and the Gram (mass) matrix $\mathbf{M}$ is symmetric positive definite. When an explicit discretization of V_a in the FE space exists, i.e.,

$$V_a(x,x') \approx V_{a,h}(x,x') := \sum_{i,j=1}^{N_h} \varphi_i(x)V_{i,j}\varphi(x') = \mathbf{\Phi}(\mathbf{x})\mathbf{V}\mathbf{\Phi}(\mathbf{x}')^\intercal$$

then $\mathbf{W} \approx \mathbf{MVM}$ and instead of (9), the following EVP can be solved

$$\mathbf{MVMa_{m,h}} = \lambda_{m,h}\mathbf{Ma_{m,h}},$$

which (due to regularity of $\mathbf{M} = \mathbf{LL}^\intercal$) is equivalent to the possibly advantageous reformulation as standard EVP

$$\mathbf{L}^\intercal\mathbf{VL}\tilde{\mathbf{a}}_{\mathbf{m,h}} = \lambda_{\mathbf{m}}\tilde{\mathbf{a}}_{\mathbf{m,h}}$$

with $\mathbf{a_{m,h}} = \mathbf{L}^{-\intercal}\tilde{\mathbf{a}}_{\mathbf{m,h}}$. Further details can be found in [Kee03].

When used in SGFEM discretizations, the truncation error (due to restriction to M terms) and the FE approximation error in principle have to be taken into account, which can classically be achieved by some Strang lemma as shown in [Mat08, FST05]. For other efficient approaches to solve the KLE numerically, we refer the interested reader to [HPS15, ST06a, BC24].

2.2 Polynomial chaos expansion

The term *polynomial chaos* (also called Wiener-Hermite expansion) was originally introduced by Wiener in [Wie38] in the context of statistical mechanics. The ideas were extended subsequently by Cameron and Martin in [CM47], showing that any square-integrable functional on the set of continuous functions on the unit interval can be expanded in an L^2-convergent series of Hermite polynomials in a countable sequence of Gaussian random variables. A modern exposition of the subject can be found in [Jan97]. We follow the presentation in [EMSU12], which analyses the convergence of generalized polynomial chaos expansions and also provides references for the historical context.

To recall the setting, assume a probability space (Ω,Σ,π) admitting the definition of nontrivial normally distributed random variables $\xi \sim \mathcal{N}(0,\sigma^2)$. By $L^2(\Gamma,\Sigma,\pi)$ we denote the Hilbert space of equivalence classes of random variables defined on the probability space with values in $\mathbb{R}$. The respective inner product $(\cdot,\cdot)_\pi$ induces the norm $\|\cdot\|_\pi$ and the convergence with respect to this norm is called *mean-square convergence*. A complete linear subspace of $L^2(\Gamma,\Sigma,\pi)$ is called a *Gaussian Hilbert space* $\mathcal{H}$ if it consists of centered Gaussian random variables. For any Gaussian linear space $\mathcal{H}$ and $n \in \mathbb{N}_0$, the set

$$\mathcal{P}_n(\mathcal{H}) := \{P(\xi_1,\dots,\xi_m) : \deg(p) \le n,\ \xi_i \in \mathcal{H},\ i = 1,\dots,m,\ m \in \mathbb{N}\}$$

is a linear subspace of $L^2(\Gamma,\Sigma,\pi)$ spanned by polynomials up to degree n in an arbitrary number of random variables and $\{\mathcal{P}_n(\mathcal{H})\}_{n\ge0}$ forms a strictly increasing sequence of subspaces. Note that $\mathcal{P}_0(\mathcal{H})$

and $\mathcal{P}_1(\mathcal{H})$ consist of (degenerate a.s.) constant and normally distributed random variables, respectively. For $n > 1$, $\mathcal{P}_n(\mathcal{H})$ contains also random variables that are non-Gaussian. By an orthogonal decomposition given by the spaces

$$\mathcal{H}_n := \overline{\mathcal{P}_n}(\mathcal{H}) \cap \mathcal{P}_{n-1}(\mathcal{H})^{\perp}, \qquad n \in \mathbb{N},$$

it follows that

$$\overline{\mathcal{P}_n}(\mathcal{H}) = \bigoplus_{i=1}^{n} \mathcal{H}_i.$$

With this, the famous Cameron-Martin density result for polynomials of Gaussian random variables can be summarized as follows: $\{\mathcal{H}_n\}_{n\geq 0}$ forms a sequence of closed, pairwise orthogonal linear subspaces of $L^2(\Gamma, \Sigma, \pi)$ such that

$$\bigoplus_{n\geq 0} \mathcal{H}_n = L^2(\Gamma, \Sigma, \pi).$$

Moreover, if for the sigma algebra it holds that $\sigma(\mathcal{H}) = \Sigma$ then the following orthogonal decomposition is given

$$L^2(\Gamma, \Sigma, \pi) = \bigoplus_{n\geq 0} \mathcal{H}_n.$$

For further details we refer to [EMSU12]. Since many problems involve random variables that are non-Gaussian, generalizations of the Wiener-Hermite chaos were proposed in particular in [XK02] and other works of these authors. A central idea is to construct orthogonal polynomials with respect to a measure that is close to the actual measure of the problem parameters. This in principle is possible for any probability distribution. The mentioned reference proposed to use hypergeometric orthogonal polynomials of the Askey scheme and by this introduced the *generalized polynomial chaos (GPC) expansion* that is used in this work.

The preceding considerations motivate the use of polynomials in random variables as a means to discretize the probability space of the random variables Y_m that determine the randomness on the coefficient $a(\omega, x)$ of (1). Since it would be inconvenient to operate on a probability space, any method resulting in a functional approximation of the random solution $u(\omega, x)$ considers the images of the random variables Y_m, denoted by Γ and associated with the probability measure of $Y = (Y_m)_{m\geq 1}$. This leads to a “change of variables” and a deterministic problem in a parameter vector $y = (y_1, \ldots)$ in the parameter space $L^2_\pi(\Gamma)$. For details, we refer to [SG11, LPS14].

Stochastic Galerkin methods rely on a discretization of the parameter space $L^2_\pi(\Gamma)$ by an orthogonal basis for a given parameter domain $\Gamma := \bigotimes_{m=1}^{\infty} \Gamma_m$, which is associated to the image of the random variables parametrizing the random coefficient in (1). Since these are i.i.d. and in our setting uniformly distributed in $[-1, 1]$, the parameter space is endowed with the product probability measure $\pi := \bigotimes_{m=1}^{\infty} \pi_m$. This renders the construction of an orthogonal basis of $L^2_\pi(\Gamma)$ simple, since it can be obtained by a tensorization of a univariate basis of $L^2_{\pi_1}$. In case of uniform distributions, i.e., $\Gamma_m = [-1, 1]^{\mathbb{N}}$ and $\pi_m(y_m) := \frac{1}{2}\,\mathrm{d}\pi(y_m)$, the appropriate orthogonal polynomials are Legendre polynomials. In general, orthogonal polynomials satisfy a recurrence relation that allows a recursive computation. For the Legendre polynomials, the recurrence relation reads

$$(n+1)L_{n+1}(y) = (2n+1)yL_n(y) - nL_{n-1} \quad \text{for } n \geq 2 \quad \text{and } L_0(y) = 1,\ L_1(y) = y.$$

The norm of the Legendre polynomials is given by

$$\|L_n\|^2_{L^2(\Gamma)} = 2/(2n+1).$$

Hence, the L^2_π-normalized Legendre polynomials $P_n := L_n/||L_n||_{L^2_\pi(\Gamma)}$ satisfy the recurrence relation

$$(n+1)P_{n+1}(y) = (2n+1)y\frac{\|L_n\|_{L^2_\pi(\Gamma)}}{\|L_{n+1}\|_{L^2_\pi(\Gamma)}}P_n(y) - n\frac{\|L_{n-1}\|_{L^2_\pi(\Gamma)}}{\|L_{n+1}\|_{L^2_\pi(\Gamma)}}P_{n-1}$$

with

$$P_0(y) = 1 \qquad \text{and} \qquad P_1(y) = \frac{L_1}{||L_1||_{L^2_\pi(\Gamma)}}.$$

An orthogonal basis of $L^2_\pi(\Gamma)$ is obtained by tensorization of the univariate polynomials above. For this, we introduce the set of finitely supported multi-indices

$$\mathcal{F} := \{\mu \in \mathbb{N}_0^\infty : |\mathrm{supp}(\mu)| < \infty\},$$

where $\mathrm{supp}(\mu) := \{m : \mu_m \neq 0\}$. Then, for $\mu \in \mathcal{F}$ the associated multi-variate polynomial reads

$$P_\mu(y) = \bigotimes_{m=1}^{\infty} P_{\mu_m}(y_m) = \prod_{m\in\mathrm{supp}\,\mu} P_{\mu_m}(y_m) \tag{10}$$

and all $\{P_\mu : \mu \in \mathcal{F}\}$ form an orthogonal basis of $L^2_\pi(\Gamma)$. Note that the recurrence relations above still hold for $P_{\mu+\epsilon_m}$, P_μ and $P_{\mu-\epsilon_m}$ if only the m-th position in μ is changed by $\epsilon_m := (\delta_{mn})_{n=1^\infty}$, i.e.,

$$\alpha_m P_{\mu+\epsilon_m}(y) = \beta_m y_m P_\mu(y) + \gamma_m P_{\mu-\epsilon_m} \tag{11}$$

for some coefficients $\alpha_m, \beta_m, \gamma_m$.

3 SGFEM for the parametric Poisson model problem

This section introduces the model problem in more detail and studies its discretization via a Galerkin orthogonal projection onto a finite dimensional subspace as well as some important analytical results.

3.1 The model problem and its weak formulation

As a common benchmark model, we consider the parametric Poisson problem on some Lipschitz domain D. With this, we seek a solution u such that

$$-\mathrm{div}(a(y,x)\nabla u(y,x)) = f(x) \quad \text{for } (y,x) \in \Gamma \times D, \tag{12}$$

where a is a stochastic coefficient depending on a countable infinite set of parameters $y = (y_m)_{m\in\mathbb{N}}$. These parameters can be understood as the image of i.i.d. random variables parametrizing the model data. We assume an affine dependence of a on y in the sense that

$$a(y,x) = a_0(x) + \sum_{m=1}^{\infty} y_m a_m(x). \tag{13}$$

Moreover, assume $y \in \Gamma := [-1,1]^\infty$, $a_m \in W^{1,\infty}(D)$ and

$$\operatorname*{ess\,inf}_{x\in D} a_0(x) > 0 \quad \text{and} \quad \sum_{m=1}^{\infty} \left\| \frac{a_m}{a_0} \right\|_{L^\infty(D)} \le \gamma < 1. \tag{14}$$

These conditions imply in particular that $a(y,x)$ is uniformly bounded away from zero and from above and the series in (13) converges in the L^2 sense. The coefficient functions a_m may, e.g., stem from a Karhunen-Loève expansion as in Section 2.1.1. It consists in this case of plane wave Fourier modes and can be understood as eigenfunctions of a smooth covariance on the square domain, see [LPS14] for details.

Unique solvability of $u \in \mathcal{V} := L^2_\pi(\Gamma; V) \simeq H^1_0(D) \otimes L^2_\pi(\Gamma)$ follows from the uniform ellipticity of the bilinear form $A : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ defined by

$$\begin{aligned} A(u,v) &:= \int_\Gamma \int_D a(y,x) \nabla u(y,x) \cdot \nabla v(y,x) \,\mathrm{d}x \,\mathrm{d}\pi(y) \\ &= \int_\Gamma \int_D \left(a_0 + \sum_{m=1}^{\infty} y_m a_m \right) \nabla u(y,x) \cdot \nabla v(y,x) \,\mathrm{d}x \,\mathrm{d}\pi(y). \end{aligned} \tag{15}$$

This also induces the energy norm and mean energy norm

$$\|u\|_A^2 := A(u,u) \quad \text{and} \quad \|u\|_{A_0}^2 := \int_\Gamma \int_D a_0 \nabla u \cdot \nabla u \,\mathrm{d}x \,\mathrm{d}\pi(y). \tag{16}$$

All assumptions lead to the well-posedness of the weak formulation of (12): seek $u \in \mathcal{V}$ such that, for all $v \in \mathcal{V}$,

$$A(u,v) = F(v) := \int_\Gamma \int_D f(x) v(y,x) \,\mathrm{d}x \,\mathrm{d}\pi(y) \tag{17}$$

where existence and uniqueness follow from the Riesz representation theorem as usual.

Since the spatial and parametric spaces V_h and $\mathcal{P}_n$ are dense in V and $\mathcal{H}$ for $h \to 0$ and $n \to \infty$, it can be shown [CDS10, CDS11] that the solution u has an $L^2_\pi(\Gamma; V)$ convergent expansion of the form

$$u(y,x) = \sum_{\mu \in \mathcal{F}} u_\mu(x) P_\mu(y) \tag{18}$$

with coefficients $u_\mu \in V$. Then, (11) allows to identify the u_μ in (18) as the solution of the variational equations

$$\int_D a_0 \nabla u_\mu, \nabla v_\mu \,\mathrm{d}x + \sum_{m=1}^{\infty} \int_D a_m \left(\frac{\alpha_m}{\beta_m} \nabla u_{\mu+\epsilon_m} + \frac{\gamma_m}{\beta_m} \nabla u_{\mu-\epsilon_m} \right) \cdot \nabla v_\mu \,\mathrm{d}x = \int_D f_\mu v_\mu \,\mathrm{d}x \quad \text{for all } v_\mu \in V. \tag{19}$$

Here, $f_\mu(x) := \int_\Gamma f(x) P_\mu(y) \,\mathrm{d}\pi(y) = f \delta_{\mu 0}$ for a deterministic right-hand side f.

3.2 Discretization

The tensorized representation (18) is the point of departure for the SGFEM. The method involves an approximation of the stochastic parameter dimension $\mathcal{F}$ by selection of a finite subset $\Lambda \subset \mathcal{F}$ (ideally the most important ones) and a spatial approximation of the coefficients u_μ for $\mu \in \Lambda$ by FEM for some discrete space $V_{\mu,h}$ based on some regular triangulation of D, i.e., by some

$$u_N(y,x) = \sum_{\mu\in\Lambda} u_{N,\mu}(x) P_\mu(y). \tag{20}$$

We assume[3] that all $u_{N,\mu}$ live in the same discrete ansatz space, i.e., $u_{N,\mu} \in V_h$ for all μ.

Remark 3.1. It is in principle possible to seek each $u_{N,\mu}$ in a different approximation space $V_{h,\mu}$, e.g., based on a different triangulation or a different polynomial order. In fact, one would expect mode-dependent adaptive methods to lead to a discretization that is optimally tailored to the problem at hand in terms of complexity. However, due to the coupling of the spatial coefficients of neighboring stochastic modes, this requires costly interpolations between the different spaces, at least when realized naively. A first result for this was shown in [EGSZ14] (see also the comparison with single-mesh approximations in [EGSZ15]) and later in [BPR22]. Moreover, an approach based on hierarchical frames representations was demonstrated in [BEEV24]. For Stochastic Collocation, this was realized in [BS23]. Note that the suboptimality of a single-mesh representation can at least be compensated partially by higher-order methods as shown in [Git14].

Remark 3.2. A polynomial chaos representation exhibits noteworthy advantages not only when evaluating specific realizations of random fields such as the solution but also enables a fast computation of statistical quantities, in particular moments. It is easy to see [EPS17, LMK10] that the expectation is just

$$\mathbb{E}[u_N(\cdot,x)] = \int_\Gamma u(y,x)\,\mathrm{d}\pi(y) = u(0,x).$$

For the variance it holds that

$$\mathbb{V}(u_N(\cdot,x)) = \mathbb{E}[u_N(\cdot,x)^2] - \mathbb{E}[u_N(\cdot,x)]^2$$

where we can use that, due to orthogonality,

$$\begin{aligned}\mathbb{E}[u_N(\cdot,x)^2] &= \int_\Gamma u_N(y,x)^2\,\mathrm{d}\pi(y)\\ &= \sum_{\nu,\nu'\in\Lambda} u_{N,\nu}(x)u_{N,\nu'}(x)\int_\Gamma P_\nu(y)P_{\nu'}(y)\,\mathrm{d}\pi(y)\\ &= \sum_{\nu\in\Lambda} u_{N,\nu}(x)^2.\end{aligned}$$

While the ansatz space for the stochastic part P_μ is discussed above, the spatial discretization needs some further introduction. Here, classical conforming H^1-conforming Lagrange FEM on a regular triangulation $\mathcal{T}$ of D into simplices is employed. The vertices and edges (or faces in three dimensions) of the triangulation are denoted by $\mathcal{N}$ and $\mathcal{E}$, respectively. The diameter of a simplex $T \in \mathcal{T} \cup \mathcal{E}$ is denoted by h_T.

Any discrete function $V_h := P_k(\mathcal{T}) \cap V$ can be written as a continuous piecewise polynomial from the space

$$P_k(\mathcal{T}) := \{q_h \in L^2(D) : q_h|_T \in P_k(T) \quad \text{for all } T \in \mathcal{T}\},$$

[3] for simplicity, see the following remark

where $P_k(T)$ are the polynomials of maximal degree k on a simplex $T \in \mathcal{T}$.

Given a finite set of multi-indices $\Lambda \subset \mathcal{F}$ and the finite element space $V_h \subset H_0^1(D)$, the discrete product space that approximates $\mathcal{V}$ reads

$$\mathcal{V}_N(\Lambda, \mathcal{T}) := \left\{ \sum_{\mu \in \Lambda} v_{N,\mu} P_\mu : \mu \in \Lambda,\ v_{N,\mu} \in V_h \right\} \subset \mathcal{V}.$$

The discrete problem seeks the Galerkin projection $u_N \in \mathcal{V}_N$ with

$$A(u_N, v_N) = F(v_N) \quad \text{for all } v_N \in \mathcal{V}_N. \tag{21}$$

Note that for implementing (21), the sum in the evaluation of the operator A in (15) can be truncated at the maximal involved stochastic mode $M := \max_{\mu \in \Lambda} \operatorname{len}(\mu)$.

Moreover, testing (21) with $v_h = P_\mu w_h$ for $w_h \in V_h$ yields the mode-wise Galerkin orthogonality

$$A(u - u_N, P_\mu w_h) = F(P_\mu w_h) - A(u_N, P_\mu w_h) = 0 \quad \text{for all } w_h \in V_h. \tag{22}$$

This is crucial for the a posteriori error control discussed in Section 4 below.

3.3 Algebraic structure and linear solver

The system matrix of problem (21) can be written in the form

$$A = G_0 \otimes A_0 + \sum_{m=1}^{M} G_m \otimes A_m, \tag{23}$$

where A_m are the spatial stiffness matrices connected to the coefficients a_m. Given some enumeration $\tau : [\dim(\Lambda)] \subset \mathbb{N} \to \Lambda$ of the multi-indices in Λ, the parametric matrices G_m can be expressed by

$$\begin{aligned} G_0 &= \left(\int_\Gamma P_{\tau(j)}(y) P_{\tau(k)}(y) \, \mathrm{d}\pi(y) \right)_{j,k=1,\dots,\dim(\Lambda)} = \mathbb{I}_{\dim(\Lambda)}, \\ G_m &= \left(\int_\Gamma y_m P_{\tau(j)}(y) P_{\tau(k)}(y) \, \mathrm{d}\pi(y) \right)_{j,k=1,\dots,\dim(\Lambda)} \quad \text{for m = 1,...,M.} \end{aligned}$$

Here, the matrix G_0 simply is the identity matrix $\mathbb{I}_{\dim(\Lambda)} \in \mathbb{R}^{\dim(\Lambda) \times \dim(\Lambda)}$ due to the orthonormality of the stochastic ansatz functions, and the (sparse) matrices G_m can be computed by the recurrence relation (11). Analogously to (19), $u_{h,\mu}$ is the solution of the subproblem

$$\int_D a_0 \nabla u_{h,\mu}, \nabla v_{h,\mu} \, \mathrm{d}x + \sum_{m=1}^{\infty} \int_D a_m \left(\frac{\alpha_m}{\beta_m} \nabla u_{h,\mu+\epsilon_m} + \frac{\gamma_m}{\beta_m} \nabla u_{h,\mu-\epsilon_m} \right) \cdot \nabla v_{h,\mu} \, \mathrm{d}x = \int_D f_\mu v_{h,\mu} \, \mathrm{d}x$$
$$\text{for all } v_{h,\mu} \in V_h. \tag{24}$$

Hence, the linear system with the system matrix from (23) can be written explicitly in the form

$$\begin{bmatrix} \ddots & & & \vdots & & & \iddots \\ & A_0 & \cdots & \frac{\gamma_m}{\beta_m} A_m & \cdots & 0 & \\ & \vdots & \ddots & \vdots & \iddots & \vdots & \\ \cdots & \frac{\gamma_m}{\beta_m} A_m & \cdots & A_0 & \cdots & \frac{\alpha_m}{\beta_m} A_m & \cdots \\ & \vdots & \iddots & \vdots & \ddots & \vdots & \\ & 0 & \cdots & \frac{\alpha_m}{\beta_m} A_m & \cdots & A_0 & \\ \iddots & & & \vdots & & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \mathbf{u}_{\mu-\epsilon_m} \\ \vdots \\ \mathbf{u}_{\mu} \\ \vdots \\ \mathbf{u}_{\mu+\epsilon_m} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{f}_{\mu-\epsilon_m} \\ \vdots \\ \mathbf{f}_{\mu} \\ \vdots \\ \mathbf{f}_{\mu+\epsilon_m} \\ \vdots \end{bmatrix}$$

where $\mathbf{u}_\mu$ denotes the coefficients of $u_{h,\mu}$ with respect to the basis of V_h and $\mathbf{f}_\mu$ denotes the discrete right-hand side with entries $(\mathbf{f}_\mu)_j := \int_D f_\mu \varphi_j \,\mathrm{d}x$ for the j-th spatial basis function φ_j from V_h.

Due to the large size of the system matrix, direct solvers become prohibitively expensive for larger numbers of stochastic modes. To alleviate this computational burden, it is advised to use an iterative solver that exploits the sparse block structure of the problem. The simplest but already quite powerful approach is a mean-based construction relying on a factorization of A_0, e.g., using $P_0 := G_0 \otimes A_0$ and its inverse $P_0^{-1} = G_0 \otimes A_0^{-1}$ in a conjugated gradients algorithm. Further discussions and more sophisticated approaches can be found in [EPSU09, Ull10]. We also point out that a full construction of the involved matrices is not required and block-wise solution process is much more efficient memory-wise. For details on how to approach this with a preconditioned conjugate gradient (PCG) iterative solver, we refer to [Git13, EGSZ14].

3.4 Tensor structure and low-rank compression

According to the Bochner tensor space $\mathcal{V} = H_0^1(D) \otimes \left(\bigotimes_{m=1}^{\infty} L^2_{\pi_m}(\Gamma_m)\right)$ of the model problem (12), the representation of the stochastic Galerkin FEM discretization (23) exhibits a natural tensor structure, which lends itself to modern low-rank tensor formats [BSU16, Nou17]. An in principle direct translation of the adaptive algorithm described in Section 5 and likewise [EGSZ14] can be found in [EPS17]. Concretely, the algebraic Galerkin system has the form

$$\mathbf{A}(U) := \left(\sum_{m=0}^{M} \mathbf{A}_m\right)(U) = F \tag{25}$$

with

$$\mathbf{A}_m := K_m \otimes I \otimes \cdots \otimes B_m \otimes I \otimes \cdots \otimes I, \tag{26}$$

$$F := \mathbf{f} \otimes e_1 \otimes \cdots \otimes e_1. \tag{27}$$

Here, e_1 denotes the first unit vector and for $m = 1, \ldots, M$,

$$K_m(i,j) := \int_D a_m(x) \nabla\varphi_i(x) \cdot \nabla\varphi_j(x) \,\mathrm{d}x, \qquad i,j = 1,\ldots,N_h, \tag{28}$$

$$B_m(\mu,\nu) := \int_{\Gamma_m} y_m P_{\mu_m} P_{\nu_m} \,\mathrm{d}\pi_m(y_m), \qquad \mu,\nu \in \mathcal{F}, \tag{29}$$

$$\mathbf{f}(j) := \int_D f(x)\varphi_j(x) \,\mathrm{d}x, \qquad j = 1,\ldots,N_h. \tag{30}$$

By this, the coefficient tensor $U \in \mathbb{R}^{\dim V_h \times d_1 \times \cdots \times d_M}$ for the tensorized finite dimensional approximation space is determined by the tensor set

$$\Lambda := \{(\mu_1, \ldots, \mu_M, 0, \ldots) : \mu_m = 1, \ldots, d_m;\ m = 1, \ldots, M\}$$

such that

$$u_N(y, x) = \sum_{i=1}^{N} \sum_{\mu \in \Lambda} U(i, \mu) \varphi_i(x) P_\mu(y). \tag{31}$$

Here, in contrast to the sparse discretization Section 3.2 with respect to some selection of active modes in the set $\Lambda \subset \mathcal{F}$, the space $\mathcal{V}_N(\Lambda, \mathcal{T})$ contains all polynomials up to a certain degree d_m in mode m. Since this discrete space and hence the coefficient tensor in (31) scales like $\mathcal{O}(d^M)$ in the parameter dimensions with $d = \max\{d_1, \ldots, d_M\}$, the problem cannot be solved without some form of model reduction or compression. In [EPS17], the popular tensor train (TT) format (also known as Matrix Product States (MPS) in quantum physics) is used, which exhibits a storage complexity of $\mathcal{O}(M d r_{\max}^2)$, i.e. linear in the dimension M and quadratic in the rank $r_{\max} = \max \mathbf{r}$. Solving (25) with an alternating linear scheme (ALS) then leads to the representation

$$U(i, \mu_1, \ldots, \mu_M) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M+1}^{r_{M+1}}} U_0(i, k_1) \prod_{m=1}^{M} U_m(k_m, \mu_m, k_{m+1})$$

with ranks $\mathbf{r} = (r_1, \ldots, r_{M+1})$, $r_{M+1} = 1$ and component tensors $U_m \in \mathbb{R}^{r_{m-1} \times d_m \times r_m}$. This format can significantly reduce the representation complexity of the Galerkin approximation.

We note in passing that an explicit tensor construction for the substantially more involved lognormal Darcy problem is shown in [EMPS20] and a tensor reconstruction of the lognormal representation with different means is derived in [EFHT23b].

3.5 Convergence analysis

A priori convergence results for sparse polynomial representations were derived in [CDS10, CDS11]. These were refined later in [BCM17] with the main novelty of compactly and locally supported expansions of the data, resulting in a larger class of admissible coefficient expansions. We recall the main statements as given in [SG11], which also includes the overall convergence when combining the physical FE with the parameter space discretization. For further details, the interested reader is referred to [CD15].

Theorem 3.3 ([SG11] Theorem 3.7)**.** *Under the boundedness assumptions on $a(y, x)$ and if the summability condition $(||a_m||_{L^\infty(D)})_{m \geq 1} \in \ell^p(\mathbb{N})$ holds true for some $p < 1$, then the coefficient sequences $(||u_\nu||_V)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})$ and the Legendre GPC expansion converges in $L^2_\pi(\Gamma; V)$ like*

$$\left\| u - \sum_{\nu \in \Lambda_N} u_\nu P_\nu \right\|_{L^2_\pi(\Gamma; V)} \leq \|(||v_\nu||_V)\|_{\ell(\mathcal{F})} N^{-s}, \quad s = \frac{1}{p} - \frac{1}{2},$$

where $\Lambda_N \subset \mathcal{F}$ is the set of indices corresponding to the N largest $||u_\nu||_V$ of (18).

For the a priori convergence rate of the SGFEM, approximation properties of the physical discretization have to be considered. Assuming sufficient regularity of the right-hand side $f \in L^2(D) \subset V'$, $u \in W :=$

$H^2 \cap V$ and finite element degrees of freedom $M_h = \dim(V_h) \sim h^{-d}$ with standard conforming Lagrange elements on a regular triangulation with mesh width h of the convex polyhedral domain $D \subset \mathbb{R}^d$, for some $0 < t \leq 1/d$ we can expect that

$$\inf_{v_h \in V_h} \|u - v_h\|_V \lesssim M_h^{-t} |u|_W. \tag{32}$$

This leads to the following regularity result of the solution.

Theorem 3.4 ([SG11] Theorem 3.8)**.** *Under the boundedness assumptions on $a(y,x)$, with $f \in L^2(D)$, $||\nabla a_0||_{L^\infty(D)} < \infty$ and if the summability conditions $(||a_m||_{L^\infty(D)})_{m\geq 1} \in \ell^p(\mathbb{N})$ and $(||\nabla a_m||_{L^\infty(D)})_{m\geq 1} \in \ell^p(\mathbb{N})$ hold true for some $p < 1$, then the coefficient sequence $(||u_\nu||_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$.*

Finally, we recall the overall discretization convergence rate.

Theorem 3.5 ([SG11] Theorem 3.9)**.** *For the set $\Lambda_N \subset \mathcal{F}$ consisting of the multi-indices of the N largest $||u_\nu||_W$ of (18), for $\nu \in \Lambda_N$ there exist finite element spaces $V_\nu \subset V$ of dimension M_ν such that*

$$\left\| u - \sum_{\nu\in\Lambda_N} u_\nu P_\nu \right\|_{L^2_\pi(\Gamma;V)} \lesssim \left(\sum_{\nu\in\Lambda_N} M_\nu \right)^{-\min\{s,t\}} \qquad \text{with } s = \frac{1}{p} - \frac{1}{2}.$$

Remark 3.6. Note that the optimal convergence stated above cannot be achieved with the single mesh approach used for the numerical examples in Section 6. Instead, different meshes adjusted to the required accuracy of each stochastic mode have to be used as in [BEEV24, BPR22, EGSZ14].

It is known that the stochastic convergence, i.e. the convergence in the generalized polynomial chaos representation of the parameter space, is exponential in the polynomial degrees if the model randomness is finite dimensional. This is in particular the case when certain pre-determined parameters in the model are varied (in contrast to representing a random field as with the KLE) a common problem for which is the so-called "cookie problem" with varying values on a fixed number of inclusions in the physical domain. We refer to the convergence result in [BNT07, Theorem 4.1] for details.

4 A posteriori error control

While theoretical properties of the affine model problem are well understood as sketched in Section 3.5 and [SG11, CD15, BCM17], obtaining optimal discretizations numerically is challenging due to the possibly large dimensions of the parameter space and the required balancing of approximation and truncation errors. Adaptive Stochastic Galerkin FEM (ASGFEM) have proved to be an efficient method to obtain sparse GPC representations with reliable error estimates, in particular exploiting the orthogonality property of the Galerkin projection not given in most other methods.

In this section we discuss residual and equilibration based a posteriori error control for the mean energy error of the approximation, i.e., for the norm

$$\|u\|_{A_0}^2 = \int_\Gamma \int_D a_0(x) |\nabla u(y,x)|^2 \,\mathrm{d}x \,\mathrm{d}\pi(y) = \sum_{\mu\in\mathcal{F}} \int_D a_0 |\nabla u_\mu|^2 \,\mathrm{d}x =: \sum_{\mu\in\mathcal{F}} \|u_\mu\|_{a_0}^2.$$

Due to (14), this mean energy norm is equivalent to the energy norm [EGSZ14, BPS14] via

$$(1-\gamma)\|u\|_{A_0}^2 \leq \|u\|_A^2 \leq (1+\gamma)\|u\|_{A_0}^2.$$

Since $A(\cdot,\cdot)$ is a scalar product on $\mathcal{V}$, the Riesz representation theorem yields

$$\|u-u_N\|_A = \sup_{v\in\mathcal{V}} \frac{A(u-u_N,v)}{\|v\|_A} \leq (1-\gamma)^{-1/2}\|R\|_{A_0^\star} \tag{33}$$

for the residual $R \in \mathcal{V}^\star$ defined by

$$R(v) := A(u-u_N,v) \quad \text{for all } v\in\mathcal{V}$$

measured in the dual norm

$$\|R\|_{A_0^\star} := \sup_{v\in\mathcal{V}} \frac{|R(v)|}{\|v\|_{A_0}}.$$

4.1 Error expansion

The control of the error between u and u_N is based on the decomposition of the residual into components with respect to the PCE of the parametric space, namely

$$R(v) = A(u-u_N,v) = A\Big(u-u_N, \sum_{\mu\in\mathcal{F}} v_\mu P_\mu\Big) = \sum_{\mu\in\mathcal{F}} r_\mu(v_\mu) \quad \text{for } v\in\mathcal{V},$$

where the subresiduals are defined for all $v \in V$ (now a deterministic function like v_μ above) by

$$r_\mu(v) := F(P_\mu v) - A(u_N, P_\mu v) = \int_D f_\mu v \,\mathrm{d}x - \int_D \sigma_\mu \nabla v \,\mathrm{d}x \quad \text{with} \quad f_\mu := \begin{cases} f & \text{for } \mu = 0 \\ 0 & \text{else} \end{cases} \tag{34}$$

and some effective numerical subresidual stress σ_μ defined by

$$\sigma_\mu := a_0\nabla u_{N,\mu} + \sum_{m=1}^{\infty} a_m \cdot \nabla\left(\frac{\alpha_m}{\beta_m}u_{N,\mu+\epsilon_m} + \frac{\gamma_m}{\beta_m}u_{N,\mu-\epsilon_m}\right).$$

This effective subresidual stress σ_h mainly reflects the coupling of different solution modes to the mode μ by the recurrence relation (11), mirroring the operator coupling structure (24).

Since $u_{N,\mu} = 0$ for $\mu \in \mathcal{F}\setminus\Lambda$, it follows that σ_μ vanishes for those μ where $\mu \pm \epsilon_m \notin \Lambda$ for all m. This leads to the definition of the boundary set of modes

$$\partial\Lambda := \Big\{\nu\in\mathcal{F} : \exists\mu\in\Lambda, m\in\mathbb{N} \text{ such that } \nu = \mu+\epsilon_m \in \mathcal{F}\setminus\Lambda\Big\}.$$

Consequently, the subresiduals r_μ beyond the boundary $\partial\Lambda$ vanish, i.e.,

$$r_\mu \equiv 0 \quad \text{for all } \mu\in\mathcal{F}\setminus(\Lambda\cup\partial\Lambda). \tag{35}$$

Because of the orthonormality of the PCE, the Parseval identity (see e.g. [EM16a, Theorem 1] for details) and (35) yield

$$\|R\|^2_{A_0^\star} = \sum_{\mu\in\mathcal{F}} \|r_\mu\|^2_{a_0^\star} = \sum_{\mu\in\Lambda} \|r_\mu\|^2_{a_0^\star} + \sum_{\mu\in\partial\Lambda} \|r_\mu\|^2_{a_0^\star}, \tag{36}$$

where the subresiduals are measured in the dual norm

$$\|r\|_{a_0^\star} := \sup_{v\in V} \frac{r(v)}{\|v\|_{a_0}}. \tag{37}$$

Hence, to bound the dual norm of the full residual one can bound the dual norm of all subresiduals. Recall from (22), that the subresiduals for $\mu \in \Lambda$ enjoy Galerkin orthogonality, i.e. it holds that

$$r_\mu(v_h) = 0 \quad \text{for all } v_h \in V_h.$$

This allows to apply standard techniques known for deterministic problems. The subresiduals and their local contributions are subsequently used to steer the adaptive refinement in the spatial dimension.

Note that the remaining subresiduals $\mu \in \partial\Lambda$ lack Galerkin orthogonality but are crucial since they estimate the error incurred by neglecting the modes that are not in Λ (yet). Therefore, the estimators for these subresiduals are essential to steer the stochastic refinement.

Below, two approaches to derive reliable error estimators for the subresiduals are discussed, namely the *explicit residual-based error estimator* and the *equilibration error estimator*.

4.2 Explicit residual-based error estimator

The classical explicit residual-based error estimator involves a volume contribution consisting of the piecewise divergence div_h and normal jumps $[[\sigma_\mu \cdot \mathbf{n}_F]]$ over faces $F \in \mathcal{F}$ of the discrete stress σ_μ.

The resulting local error estimator for the dual norm (37) of the subresidual (34) on some element $T \in \mathcal{T}$ with edges $\mathcal{E}(T) := \{E \in \mathcal{E} : E \subset \partial T\}$ reads

$$\eta_{\mu,T}^2 := \frac{h_T^2}{a_{0,T}} \|f_\mu + \operatorname{div}_h \sigma_\mu\|_{L^2(T)}^2 + \sum_{E\in\mathcal{E}(T)} \frac{h_E}{a_{0,E}} \|[[\sigma_\mu \cdot \mathbf{n}_E]]\|_{L^2(E)}^2 \quad \text{for } \mu \in \Lambda, \text{ and} \tag{38}$$

$$\eta_{\mu,T}^2 := \frac{1}{a_{0,D}} \|f_\mu + \operatorname{div}_h \sigma_\mu\|_{L^2(T)}^2 + \sum_{E\in\mathcal{E}(T)} \frac{h_E^{-1}}{a_{0,D}} \|[[\sigma_\mu \cdot \mathbf{n}_E]]\|_{L^2(E)}^2 \quad \text{for } \mu \in \partial\Lambda, \tag{39}$$

where $a_{0,T} := \operatorname{essinf}_{x\in T} a_0(x)$ and $a_{0,E} := \operatorname{essinf}_{x\in\omega_E} a_0(x)$ for the edge patch ω_E of $E \in \mathcal{E}$. Note that the jump $[[\cdot]]$ is defined as zero for boundary edges $E \in \mathcal{E}, E \subset \partial D$. The estimator (38) is derived in the same way as in the classical deterministic context by using the Galerkin orthogonality and a quasi-interpolation operator $J : V \to V_h$ (e.g. the Scott–Zhang interpolator [SZ90]) such that

$$r_\mu(v) = r_\mu(v - Jv) = \int_D (f + \operatorname{div}\sigma_\mu)(v - Jv)\,\mathrm{d}x + \sum_{E\in\mathcal{E}} \int_E [[\sigma_\mu \cdot \mathbf{n}_F]](v - Jv)\,\mathrm{d}s.$$

Subsequently, the first order approximation properties of J are used, namely

$$\|(1-J)v\|_{L^2(T)} \lesssim h_T \|\nabla v\|_{L^2(\Omega_T)} \quad \text{and} \quad \|(1-J)v\|_{L^2(E)} \lesssim h_E^{1/2} \|\nabla v\|_{L^2(\Omega_E)},$$

for local neighborhoods Ω_T and Ω_E of $T \in \mathcal{T}$ and $E \in \mathcal{E}$, respectively, and some overlap argument to get the bound

$$\|r_\mu\|_{a_0^\star} \leq C_{\mathsf{rel}} \eta_\mu \quad \text{for} \quad \eta_\mu^2 := \sum_{T\in\mathcal{T}]} \eta_{\mu,T}^2$$

with some reliability constant C_{rel} that only depends on the shape constants of $\mathcal{T}$. For (39) Galerkin orthogonality cannot be used. Instead, the residual is estimated directly and the trace inequality

$$\|v\|^2_{L^2(E)} \lesssim h_E^{-1/2}\|v\|_{L^2(T_E)} + h_E^{1/2}\|\nabla v\|_{L^2(T_E)}$$

for some neighboring simplex T_E of $E \in \mathcal{E}$ is employed. With these estimates it follows that

$$\begin{aligned}\sum_{E\in\mathcal{E}} \int_E [[\sigma_\mu \cdot \mathbf{n}_E]] v \,\mathrm{d}s &\leq \sum_{E\in\mathcal{E}} \frac{h_F^{-1/2}}{a_{0,D}^{1/2}} \|[[\sigma_\mu \cdot \mathbf{n}_E]]\|_{L^2(E)} a_{0,D}^{1/2} \left(\|v\|_{L^2(T_E)} + h_E\|\nabla v\|_{L^2(T_E)}\right)\\ &\lesssim \left(\sum_{E\in\mathcal{E}} \frac{h_E^{-1}}{a_{0,D}} \|[[\sigma_\mu \cdot \mathbf{n}_E]]\|^2_{L^2(E)}\right)^{1/2} (1+h)\,\|\nabla v\|_{a_0}.\end{aligned}$$

Here, $a_{0,D} := \operatorname{essinf}_{x\in D} a_0(x)$ denotes the global infimum of a_0. Efficiency can be derived in the spirit of [Ver13] by testing the residual with appropriate bubble functions. The combination of this shows that equivalence holds in the sense that

$$C_{\mathsf{eff}}(\eta_\mu + \operatorname{osc}(f_\mu)) \leq \|r_\mu\|_{a_0^\star} \leq C_{\mathsf{rel}}\eta_\mu$$

with the usual data oscillations $\operatorname{osc}^2(f) := \sum_{T\in\mathcal{T}} h_T^2 \|f - |T|^{-1}\int_T f\,dx\|^2_{L^2(T)}$.

4.3 Flux equilibration-based error estimator

Flux equilibration in the spirit of [BS08, Voh11, Mer13, EM16a] is based on the idea that by an integration by parts, any $q_\mu \in H(\operatorname{div}, D)$ exhibits the property that

$$r(v) = \int_D (f_\mu + \operatorname{div} q_\mu) v \,\mathrm{d}x + \int_D (q_\mu - \sigma_\mu)\cdot \nabla v \,\mathrm{d}x.$$

If q_μ additionally satisfies the equilibration constraint $\int_T f_\mu + \operatorname{div} q_\mu \,\mathrm{d}x = 0$ for all $T \in \mathcal{T}$, v can be replaced by $v - v_{\mathcal{T}}$ in the first term, where $v_{\mathcal{T}}$ is the $P_0(\mathcal{T})$ best-approximation of v. This allows to apply a piecewise Poincaré inequality, i.e.,

$$\int_D (f_\mu + \operatorname{div} q_\mu) v \,\mathrm{d}x \leq \frac{h_T}{\pi a_{0,T}^{1/2}} \|f_\mu + \operatorname{div} q_\mu\|_{L^2(T)} \|a_0^{1/2}\nabla v\|_{L^2(T)} \quad \text{for all } T\in\mathcal{T}.$$

Altogether, this leads to the element-wise flux equilibration error estimator

$$\eta_{\mu,T} := \frac{h_T}{a_{0,T}^{1/2}} \|f_\mu + \operatorname{div} q_\mu\|_{L^2(T)} + \|a_0^{-1/2}(\sigma_\mu - q_\mu)\|_{L^2(T)} \quad \text{for all } T\in\mathcal{T}.$$

For $\mu \in \Lambda$, a suitable q_μ can be designed via the local patch problems

$$\begin{aligned}q_{\mu,z} = \operatorname*{argmin}_{\tau_z \in \mathrm{BDM}_k(\mathcal{T}(z))} \Big\{ \|a_0^{-1/2}(\varphi_z\sigma_\mu - \tau_z)\|_{L^2(\omega_z)} : \tau_z \cdot \vec{n} = 0 \text{ along } \partial\omega_z \setminus \partial D,& \qquad (40)\\ \operatorname{div}(\tau_z) + \pi_{k-1}(f_\mu\varphi_z + \sigma_\mu \cdot \nabla\varphi_z)\} = 0 \Big\}&\end{aligned}$$

and their superposition into the global flux

$$q_\mu = \sum_{z\in\mathcal{N}} q_{\mu,z} \in \mathrm{BDM}_k(\mathcal{T}).$$

We point out that the local space in (40) is not empty since Galerkin orthogonality shows that the compatibility condition (essentially the Gauss theorem) between the divergence and the boundary conditions on the node patch ω_z holds true, i.e. $\int_{\omega_z} \operatorname{div}\tau_z \,\mathrm{d}x = \int_{\partial\omega_z} \tau_z \cdot \vec{n} \,\mathrm{d}s = 0$. As derived above, the estimator yields a guaranteed upper bound in the sense that, for all $\mu \in \Lambda$,

$$\|r_\mu\|_{a_0^\star} \leq \eta_\mu \quad \text{for} \quad \eta_\mu^2 := \sum_{T\in\mathcal{T}} \eta_{\mu,T}^2.$$

For $\mu \in \partial\Lambda_h$, Galerkin orthogonality cannot be used and one can simply set $q_\mu = 0$ to arrive at

$$\|r_\mu\|_{a_0^\star} \leq \eta_\mu := \frac{1}{a_{0,D}^{1/2}} \|f_\mu\|_{L^2(D)} + \|a_0^{-1/2}\sigma_\mu\|_{L^2(D)} \quad \text{for} \quad \mu \in \partial\Lambda. \tag{41}$$

4.4 Tail error estimator

In the two previous subsections, two possible estimates for $\|r_\mu\|_{a_0^\star}$ for $\mu \in \partial\Lambda$ were suggested, namely (39) and (41). However, a remaining problem persists in that $\partial\Lambda$ contains countable infinitely many terms. There are (at least) two ways to proceed from here, which we discuss in the following.

The strategy in [EGSZ14, EM16b] relies on the assumption that $f_\mu = 0$ for $\mu \in \partial\Lambda$, which is the case for Poisson model problem with deterministic right-hand side f, and then further estimates the last term in (41) by

$$\|a_0^{-1/2}\sigma_\mu\| \lesssim \|a_0^{1/2}\nabla u_{N,\mu}\| + \sum_{m=1}^{\infty} \Big\|\frac{a_m}{a_0^{1/2}}\Big\|_{L^\infty(D)} \Big\|\nabla\left(\frac{\alpha_m}{\beta_m} u_{N,\mu+\epsilon_m} + \frac{\gamma_m}{\beta_m} u_{N,\mu-\epsilon_m}\right)\Big\|_{L^2(D)}.$$

For $\mu \in \partial\Lambda$, we have that $u_{N,\mu} = 0$ and usually also $u_{N,\mu+\epsilon_m} = 0$. Hence, it remains

$$\|a_0^{-1/2}\sigma_\mu\| \lesssim \sum_{m=1}^{\infty} \Big\|\frac{a_m}{a_0^{1/2}}\Big\|_{L^\infty(D)} \frac{\gamma_m}{\beta_m} \|u_{N,\mu-\epsilon_m}\|_{L^2(D)}.$$

Here, contributions to the sum are only nonzero if $\mu - \epsilon_m \in \Lambda$, which is a finite set. Consequently, a reordering that collects all contributions to the modes in Λ is possible by defining

$$\zeta_\mu := \|u_{N,\mu}\| \sum_{\substack{m=1\\ \mu+\epsilon_m\in\partial\Lambda}}^{\infty} \Big\|\frac{a_m}{a_0^{1/2}}\Big\|_{L^\infty(D)} \frac{\gamma_m}{\beta_m} \qquad \text{for } \mu \in \Lambda.$$

This quantity measures the tail error of all modes that are connected to the mode $\mu \in \Lambda$ via the recurrence relation (11). A refinement procedure based on this strategy first computes all $\zeta_\mu \in \Lambda$ and then selects those m with $\mu + \epsilon_m \in \partial\Lambda_h$ with the largest factors $\|a_m a_0^{-1/2}\| \gamma_m \beta_m^{-1}$ for refinement until the marking criterion is satisfied.

However, in this paper a more pragmatic and general strategy is favored: given some desired "tail scanning length" assignment $S_\Lambda : \Lambda \to \mathbb{N}$, a finite dimensional set can be constructed by

$$\partial_h\Lambda := \Big\{\nu \in \mathcal{F} : \exists \mu \in \Lambda, m \in \operatorname{len}(\mu) + \{1, \dots, S_\Lambda(\mu)\} \text{ such that } \nu = \mu + \epsilon_m \in \mathcal{F} \setminus \Lambda\Big\}. \tag{42}$$

Here $\operatorname{len}(\mu)$ is the length of μ, i.e., the index of the last nonzero entry of μ.

By replacing $\partial\Lambda$ with the truncated finite subset $\partial\Lambda_h$ from (42), only the η_μ for $\mu \in \partial_h\Lambda$ are computed. This strategy can be generalized straight-forwardly to problems with stochastic right-hand sides f (as required in the numerical example of Section 7) and also allows for a convenient "all-at-once" implementation approach, treating all subresiduals in the same way. The decay property of the coefficient function a_m suggests that the error incurred from neglecting $\partial\Lambda \setminus \partial_h\Lambda$ becomes relatively small quite quickly for larger m, i.e. increasingly many considered parameter dimensions. This can be further accelerated by increasing the values of S_Λ in the definition of (42).

5 Adaptive refinement algorithm

This section explains how to adaptively refine the spatial and stochastic degrees of freedom based on the a posteriori error estimators $(\eta_\mu)_{\mu\in\Lambda\cup\partial_h\Lambda}$ from the previous section and collects some known results on convergence of similar algorithms.

5.1 The algorithm

The pseudo-code for one iteration of the

$$\texttt{Solve} \quad\rightarrow\quad \texttt{Estimate} \quad\rightarrow\quad \texttt{Mark} \quad\rightarrow\quad \texttt{Refine}$$

loop is given in Algorithm 1 and explained step by step below.

```
$\mathcal{T}, \Lambda \leftarrow \texttt{ASGFEM}[\mathcal{T}, \Lambda, \theta_x, \theta_y]$
$u_N \leftarrow \texttt{Solve}[\Lambda, \mathcal{T}]$
$(\eta_T)_{T\in\mathcal{T}}, (\eta_\mu)_{\mu\in\Lambda}, \eta(\Lambda), \eta(\partial_h\Lambda) \leftarrow \texttt{Estimate}[u_N, \mathcal{T}, \Lambda, \partial_h\Lambda]$
if $\eta(\Lambda) \geq \eta(\partial_h\Lambda)$ then
    $\mathcal{T}_{\text{refine}} \leftarrow \texttt{Mark}[\theta_x, (\eta_T)_{T\in\mathcal{T}}]$
    $\mathcal{T} \leftarrow \texttt{Refine}[\mathcal{T}, \mathcal{T}_{\text{refine}}]$
else
    $\Lambda_{\text{new}} \leftarrow \texttt{Mark}[\theta_y, (\eta_\mu)_{\mu\in\partial_h\Lambda}]$
    $\Lambda \leftarrow \Lambda \cup \Lambda_{\text{new}}$
```

After computing the solution u_N of (21) with the current triangulation $\mathcal{T}$ and the current set of stochastic modes Λ with the method $\texttt{Solve}$, the error estimators are computed in the method $\texttt{Estimate}$. This involves the computation of all η_μ for $\mu \in \Lambda \cup \partial_h\Lambda$ as well as their combination to the total spatial and stochastic error

$$\eta(\Lambda) := (1-\gamma)^{-1/2}\left(\sum_{\mu\in\Lambda}\eta_\mu^2\right)^{1/2} \qquad\text{and}\qquad \eta(\partial_h\Lambda) := (1-\gamma)^{-1/2}\left(\sum_{\mu\in\partial_h\Lambda}\eta_\mu^2\right)^{1/2}$$

to approximate the two sums in (36). The factor $(1-\gamma)^{-1/2}$ is due to (33) such that

$$\|u-u_N\|_A^2 \leq \eta(\Lambda)^2 + \eta(\partial_h\Lambda)^2$$

in case of the equilibration error estimator (up to an estimate for the missing modes in $\partial\Lambda\setminus\partial_h\Lambda$). Whichever quantity of these two is larger determines whether a spatial or a stochastic refinement is performed.

If spatial refinement is selected ($\eta(\Lambda) > \eta(\partial_h\Lambda)$), cell-wise refinement indicators $(\eta_T)_{T\in\mathcal{T}}$ are computed by accumulating the cell-wise contributions of each η_μ for $\mu \in \Lambda$, i.e.,

$$\eta_T^2 := \sum_{\mu\in\Lambda} \eta_{\mu,T}^2.$$

Dörfler marking in the spirit of [Dör96] is used to select the elements with the largest indicators for refinement. More precisely, for a given parameter $\theta_x \in (0,1]$, the smallest set $\mathcal{T}_{\text{refine}} \subseteq \mathcal{T}$ is selected such that

$$\sum_{T\in\mathcal{T}_{\text{refine}}} \eta_T^2 \geq \theta_x \eta(\Lambda)^2.$$

An actual implementation of this strategy computes cumulative sums of the error contributions $(\eta_T)_{T\in\mathcal{T}}$ sorted in decreasing order. After marking, the triangles in $\mathcal{T}_{\text{refine}}$ are refined via classical procedures, e.g., red-green-blue refinement or newest vertex bisection.

If stochastic refinement is selected ($\eta(\Lambda) \leq \eta(\partial_h\Lambda)$), a similar marking procedure selects the smallest subset $\Lambda_{\text{new}} \subseteq \partial_h\Lambda$ such that

$$\sum_{\mu\in\Lambda_{\text{new}}} \eta_\mu^2 \geq \theta_y \eta(\partial_h\Lambda)^2.$$

The selected modes Λ_{new} are added to the active set Λ for the next iteration.

Remark 5.1. Algorithm 1 essentially is the same as [BPRR19a, Algorithm 4+Criterion A] where the set $\partial_h\Lambda$ is called “detail set” and also other marking criterias are discussed. Therein, also a weighting parameter θ in the comparison of $\eta(\Lambda)$ and $\eta(\partial_h\Lambda)$ is considered.

5.2 Convergence of adaptive SGFEM

In the analysis of deterministic adaptive FEM, results on the numerical convergence and optimality of the iteratively constructed approximations are by now common knowledge, cf. [CKNS08, CFPP14, BDD04]. However, for ASGFEM, there are only few results available as yet, which me mention briefly in this section. For the linear model problem discussed in this work with affine coefficient dependence, the approach of [CKNS08] was extended to also accommodate the stochastic tail truncation error in [EGSZ15]. It can then be shown that for the quasi error consisting of the energy error and the weighted approximation and truncation error estimators, the adaptive algorithm is a contraction in the sense that for some $\omega_\eta,\ \omega_\zeta > 0$ and $\delta \in (0,1)$,

$$\|u_{N,j+1} - u\|_A^2 + \omega_\eta \eta_{j+1}^2 + \omega_\zeta \zeta_{j+1}^2 \leq \delta \left(\|u_{N,j} - u\|_A^2 + \omega_\eta \eta_j^2 + \omega_\zeta \zeta_j^2\right).$$

Remark 5.2. A generalization of the convergence proof in [EGSZ15] is provided in [EH23]. There, also the setting with lognormal coefficient $\exp(a(y,x))$ was scrutinized. In this more challenging setting with unbounded parameter domain and hence a lacking uniform ellipticity property of the operator, the contraction rate depends on the iteration and cannot be stated uniformly[4].

[4] at least not with the used techniques, it seems

By using a hierarchical error estimator for the physical and the parametric error components of the stochastic FEM discretization of the same model problem, in [BPRR19a] convergence of an adaptive algorithm was shown under a (for this type of error estimator common) saturation assumption. Moreover, under somewhat stronger assumptions, linear convergence could be shown.

An alternative approach was developed in [BEEV24], which relies on a multilevel expansions of the coefficient field and achieves an error reduction with a uniform rate. Notably, the (sometimes only assumed) saturation property of the refinements is ensured by the adaptive refinement of mode-dependent finite element meshes. The error estimator is based on an appropriate residual approximation with adaptive operator compression in the parametric variables.

For adaptive stochastic collocation, two similar convergence results (again only for the model problem discussed in this work) were derived in [FS21, EEST22].

6 Numerical experiments

This section presents some numerical examples to illustrate the performance of the stochastic Galerkin finite element approximation and the presented adaptive refinement algorithm.

6.1 Experimental setup

To test the capabilities of the adaptive algorithm, two domains are studied. First, the convex unit square domain $D_\square = (0,1)^2$, where one expects a rather uniform mesh refinement. Second, an L-shaped domain $D_L := (-1,1)^2 \setminus ([-1,0] \times [0,1])$ where one expects a strong refinement at the reentrant corner. On both domains, the deterministic right-hand side is $f \equiv 1$ and the coefficient κ is chosen as

$$\kappa(x,y) := a_0 + \left(\sum_{m=1}^{\infty} a_m(x) y_m \right) \qquad \text{where}$$

$$a_m(x) := \gamma m^{-\sigma} \cos(2\pi \varrho_1(m) x_1) \cos(2\pi \varrho_2(m) x_2)$$

with $\gamma = 0.9$. The coefficients ϱ_1 and ϱ_2 are computed as in [EGSZ14, EGSZ15] by

$$\varrho_1(m) = m - k(m)(k(m)+1)/2 \quad \text{and} \quad \varrho_2(m) = k(m) - \varrho_1(m),$$

where $k(m) = \lfloor -1/2 + \sqrt{1/4 + 2m} \rfloor$, i.e., the coefficient functions a_m enumerate all planar Fourier cosine modes in increasing total order. The coefficient a_0 is the mean value of $a(y,x)$ and is set to $a_0 = 1$.

The adaptive algorithm (Algorithm 1) is performed with refinement parameters $\vartheta_x = \vartheta_y = 0.5$ with a target number of degrees of freedom of $5 \cdot 10^5$. The total number of degrees of freedom is calculated by $\dim(\Lambda) \cdot \dim(V_h)$. For the tail error estimation, the truncated boundary $\partial\Lambda_h$ from (42) is used with $S_\Lambda([0]) = 10$ and $S_\Lambda(\mu) = 2$ otherwise.

The real energy error $\|u - u_N\|_A$ is unknown and therefore approximated by a Monte Carlo estimator with $M = 150$ samples $y(i), i = 1, \dots, M$ in the sense that

$$\begin{aligned} \|u - u_N\|_A^2 &= \int_\Gamma \|a(y, \bullet)^{1/2} \nabla (u(y) - u_N(y))\|_V^2 \,\mathrm{d}\gamma(y) \\ &\approx \left(\sum_{i=1}^{M} \gamma(y^{(i)}) \|a(y^{(i)}, \bullet)^{1/2} \nabla (u(y^{(i)}) - u_N(y^{(i)}))\|_V^2 \right) \left(\sum_{i=1}^{M} \gamma(y^{(i)}) \right)^{-1} . \end{aligned}$$

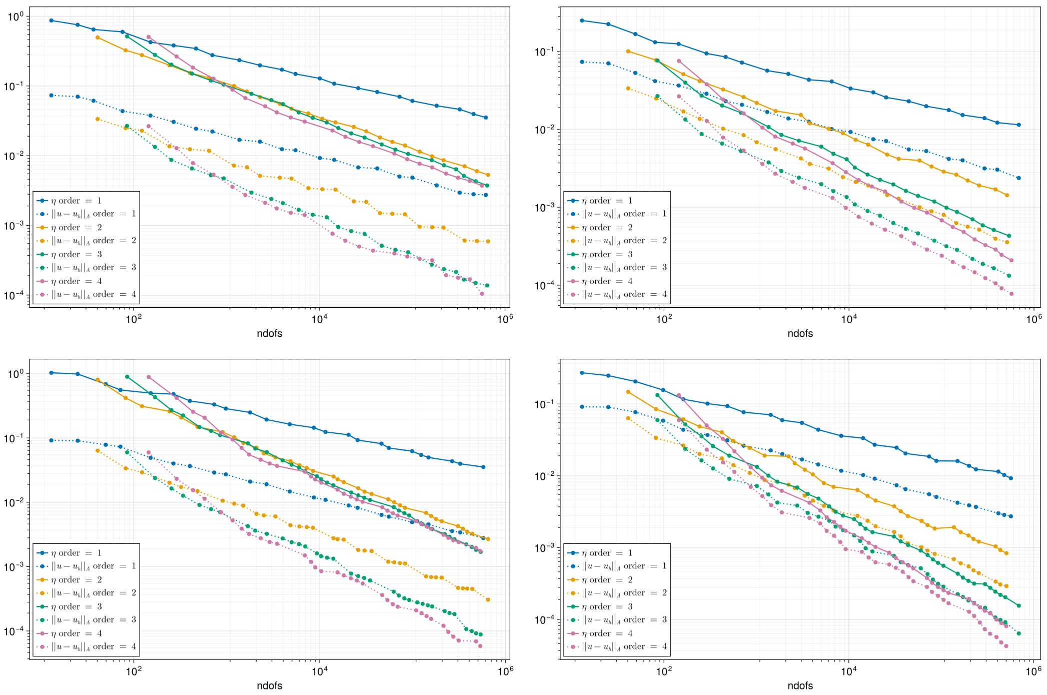

Figure 1: Square domain: Convergence history of the error and the explicit residual-based error estimator (left column) or equilibration error estimator (right column) for order $k \in \{1, 2, 3, 4\}$ and decay $\sigma = 2$ (top row) and decay $\sigma = 4$ (bottom row).

Here, the exact solutions $u(y^{(i)})$ are approximated by solving problem (12) with fixed $y = y^{(i)}$ on the same triangulation with one polynomial degree higher. The coefficient expansion in (13) is truncated after 150 terms, i.e., the samples have length $y^{(i)} \in \mathbb{R}^{150}$.

6.2 Unit square domain

Figure 1 depicts the convergence histories of the H^1 and L^2 errors and the two error estimators η from Sections 4.2 and 4.3. It can be seen that the equilibration error estimator is much closer to the real H^1 error than the explicit residual error estimator. Otherwise, the convergence of the errors is comparable. Despite the truncated boundary $\partial_h\Lambda$, no underestimation is observed, indicating that the neglected stochastic dimensions do not contribute significantly to the error. Figure 2 depicts the convergence history of the two error contributions $\eta(\Lambda)$ and $\eta(\partial_h\Lambda)$. After the initial gap between spatial and stochastic error is resolved, both components are reduced similarly while the number of degrees of freedom in space $\dim(V_h)$ and the number of stochastic modes $\dim(\Lambda)$ are increased by alternating refinements.

In the bar plots in Figure 3, the maximal polynomial order of the stochastic ansatz polynomials can be seen for each stochastic dimension y_m. Moreover, the qualitative behavior is the same for both error estimators and roughly follows the rule that a larger decay leads to a more concentrated refinement (i.e. higher order polynomials) of the lower stochastic dimensions. Moreover, a higher polynomial order k in the spatial ansatz spaces also allows for more concentration on the overall stochastic refinement. Both error estimators show a similar behavior, but the equilibration error seems to concentrate a little bit more pronounced on the stochastic refinements. This is probably due to the more accurate spatial error estimation, which might flip

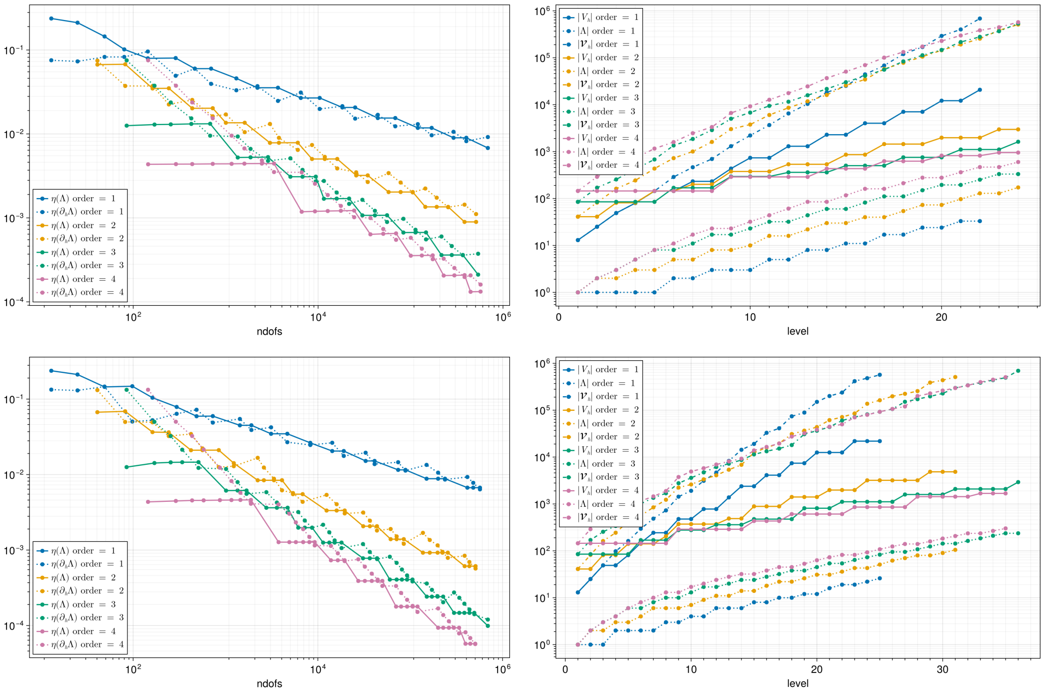

Figure 2: Square domain: Convergence history of the error estimator components $\eta(\Lambda)$ and $\eta(\partial_h\Lambda)$ (left) and the number of degrees of freedom (right) versus the level for decay $\sigma = 2$ (top) and $\sigma = 4$ (bottom).

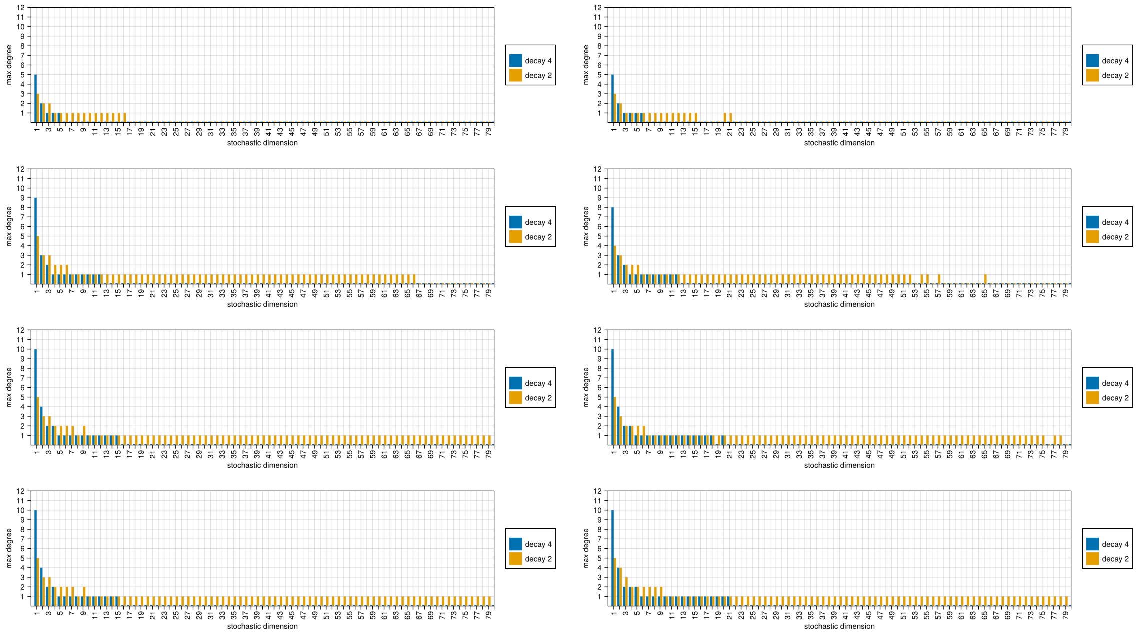

Figure 3: Square domain: Maximal polynomial degree in each stochastic dimensions y_m for order $k \in \{1, 2, 3, 4\}$ (from top to bottom) when using the explicit residual-based error estimator (left) or the equilibration error estimator (right).

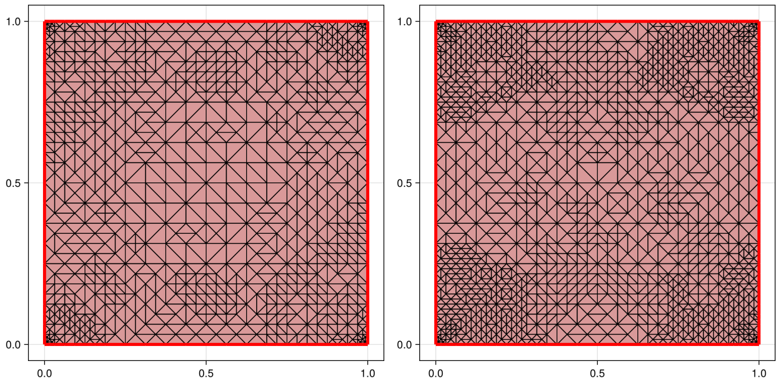

Figure 4: Square domain: refined grids after about 500.000 total degrees of freedom for order $k = 2$ with decay $\sigma = 2$ (left) and $\sigma = 4$ (right).

the condition $\eta(\Lambda) < \eta(\partial_h \Lambda)$ more often or earlier into the direction of stochastic refinement[5].

Figure 4 shows adaptively refined grids for different values of spatial polynomial order k and decay factor σ where the equilibration error estimator is used. As expected, the spatial refinement is less pronounced for larger k. At least for order $k = 2$, also some more spatial refinement can be seen when the decay σ is larger since less degrees of freedom need to be spent on the stochastic refinement.

6.3 L-shaped domain

Figure 5 presents the convergence histories for the L-shaped domain. The overall assessment is similar to the unit square case. The main difference is the larger spatial error due to the singularity at the reentrant corner of the domain. Figure 6 confirms that the spatial error is much larger in the beginning compared to the square case and that the spatial refinement begins much earlier in particular for larger polynomial degrees k. This is also confirmed by Figure 7 that compares the stochastic refinements between the two error estimators and the different decay factors. It can be observed that the stochastic refinement after about 500.000 degrees of freedom is not as strong as in the square domain case.

Figure 8 shows adaptively refined grids for different values of spatial polynomial order k and decay factor σ, where the equilibration error estimator is used. As expected, the grid refinement is concentrated at the spatial singularity, Moreover. A larger decay σ (and hence less stochastic influence) leads to slightly higher concentration on spatial refinement. Qualitatively similar results are obtained for the explicit residual-based error-estimator.

Finally, Table 1 compares the stochastic refinement history for the two domains. It shows which multi-indices are added to Λ at the iteration of the adaptive refinement loop of Algorithm 1 for the two domains under consideration and a fixed set of parameters $k = 2$ and $\sigma = 2$ using the equilibration error estimator. The two main observations are the following. First, the stochastic refinement is stronger for the square domain, which confirms the earlier observations that the spatial error is less dominant here. Second, the ordering of the multi-indices is more or less the same, they are just added in a later refinement level in case of the L-shaped domain.

[5] Experiments not presented here with larger sets $\partial_h \Lambda$ do not show a noticeable improvement compared to the shown results.

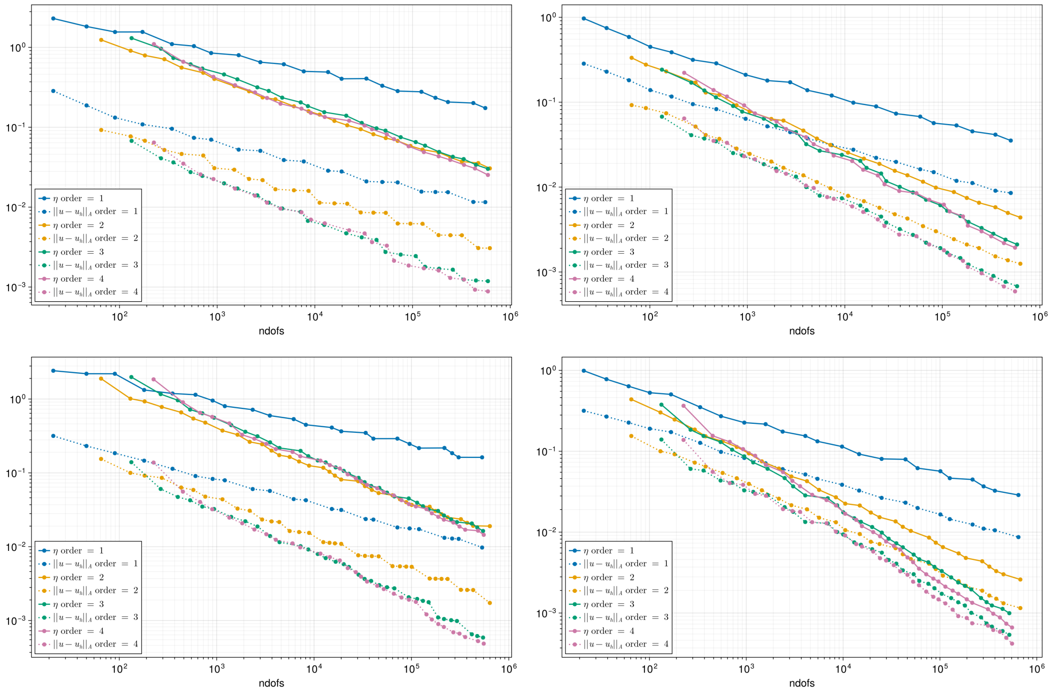

Figure 5: L-shaped domain: Convergence history of the error and the explicit residual-based error estimator (left column) or equilibration error estimator (right column) for order $k \in \{1, 2, 3, 4\}$ and decay $\sigma = 2$ (top row) and decay $\sigma = 4$ (bottom row).

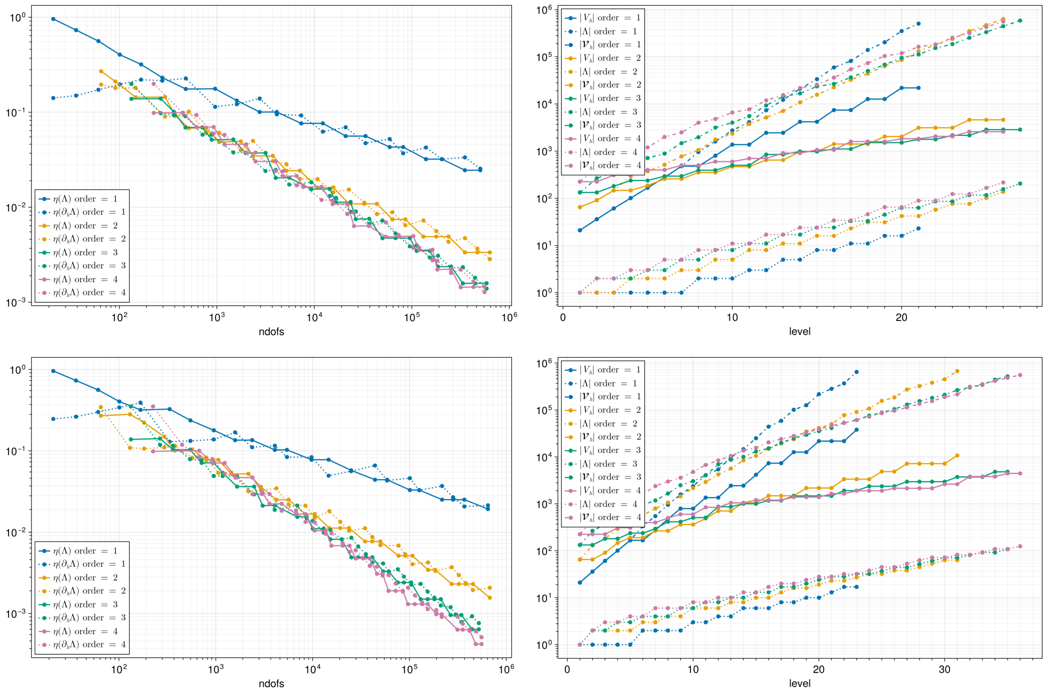

Figure 6: L-shaped domain: Convergence history of the error estimator components $\eta(\Lambda)$ and $\eta(\partial_h\Lambda)$ (left) and the number of degrees of freedom (right) versus the level for decay $\sigma = 2$ (top) and $\sigma = 4$ (bottom).

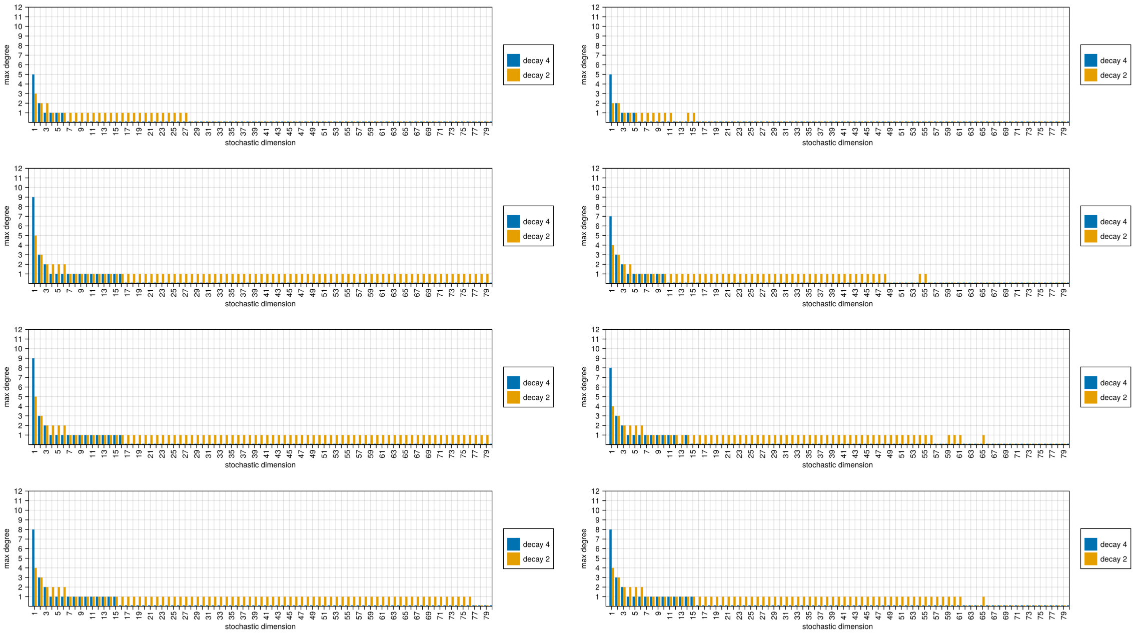

Figure 7: L-shaped domain: Maximal polynomial degree in each stochastic dimension y_m for order $k \in \{1,2,3,4\}$ (from top to bottom) when using the explicit residual-based error estimator (left) or the equilibration error estimator (right).

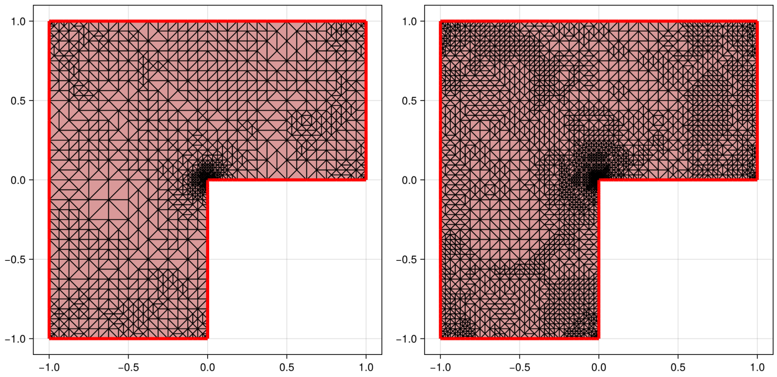

Figure 8: L-shaped domain: refined grids after about 500.000 total degrees of freedom for order $k = 2$ with decay $\sigma = 2$ (left) and $\sigma = 4$ (right).

Table 1: Stochastic refinement history for the square domain up to level 14 (left) and the L-shaped domain up to level 17 (right) with polynomial order $k = 2$ and decay $\sigma = 2$. Refinement levels without new multi-indices indicate that spatial refinement was performed.

level	added multi-indice(s) ($D = D_{\square}$)
1	[0]
2	[1]
4	[0, 1]
6	[0, 0, 1]
	[2, 0, 0]
8	[1, 1, 0, 0, 0]
	[0, 0, 0, 1, 0]
	[0, 0, 0, 0, 1]
10	[0, 0, 0, 0, 0, 1]
	[1, 0, 1, 0, 0, 0]
11	[0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
	[0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
	[0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
	[0, 2, 0, 0, 0, 0, 0, 0, 0, 0]
	[1, 0, 0, 1, 0, 0, 0, 0, 0, 0]
13	[0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
	[3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]
	[2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
14	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
	[2, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]
	[0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
	[0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
	[1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

level	added multi-indice(s) ($D = D_L$)
1	[0]
4	[1]
7	[0, 1]
9	[0, 0, 1]
	[2, 0, 0]
11	[0, 0, 0, 1, 0]
	[1, 1, 0, 0, 0]
	[0, 0, 0, 0, 1]
13	[0, 0, 0, 0, 0, 1, 0]
	[1, 0, 1, 0, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 1]
15	[0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
	[0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
	[1, 0, 0, 1, 0, 0, 0, 0, 0, 0]
	[0, 2, 0, 0, 0, 0, 0, 0, 0, 0]
17	[1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
	[1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0]

7 Extension to log-normal coefficients

This section gives a brief outlook onto a non-affine parametric problem, namely the random Poisson problem with log-normal coefficient $\kappa(y,x) = \exp(a(y,x))$. As with the affine parametric model problem, theoretical results are well established [CD13, HS14, BCDM17, Git10, GS09]. However, the unboundedness of the parameter space and the much significantly more complex coupling structure of the operator in principle require different numerical techniques. Only few results exist regarding adaptive methods. In [EMPS20, EH23], low-rank tensor formats are used to compress the algebraic equations, which either is constructed explicitly or by solving a least squares problem in the nonlinear tensor manifold. However, sparse GPC discretizations as discussed in this work are not directly possible due to an prohibitively high computational complexity. An interesting way to obtain an equivalent equation equivalent to the lognormal problem can be obtained by a log-transformation as shown in [UEE12]. By a multiplication of the model by $\exp(-a)$, it is transformed to a parametric advection-diffusion problem that allows to apply the strategies from the affine case, cf. [UEE12, UP15]. Further details of the ASGFEM presented below can be found in the upcoming work [EGM25].

7.1 The model problem and its transformation

Consider once more a model problem of the form

$$-\operatorname{div}(\kappa(y,x)\nabla u(y,x)) = f(x) \quad \text{for } (y,x) \in \Gamma \times D. \tag{43}$$

This time, we assume an isotropic Gaussian random field $\log(\kappa(x,y))$ and a Karhunen–Loève type expansion for $a(y,x) = \log(a(x,y))$, namely

$$a(y,x) = \sum_{m=1}^{\infty} y_m a_m(x), \tag{44}$$

where each $y_m \in \mathbb{R}$ is associated with an independent Gaussian random variable. To ensure summability in (44), we assume that $a_m \in L^\infty(D)$ for all $m \in \mathbb{N}$ such that

$$\sum_{m=1}^{\infty} \|a_m\|_{L^\infty(D)} < \infty. \tag{45}$$

This condition – although not preventing the existence of sequences $(y_m)_{m\in\mathbb{N}}$ leading to divergence of the sum in (44) – guarantees path-wise uniform boundedness of $a(y,x)$ on the set $\Gamma := \{y \in R^{\mathbb{N}} : \sum_{m=1}^{\infty} a_m|y_m| < \infty\}$, which is a set of measure 1 with respect to the product measure π of all Gaussians. This allows us to restrict the parameter domain from $R^{\mathbb{N}}$ to Γ. We refer to [HS14] for further discussion on the well-posedness.

From the computational side of view, a direct discretization with the nonlinear (with respect to y_m) lognormal coefficient κ would lead to a strongly coupled infinite system of equations, which easily becomes computationally intractable. In order to circumvent this issue with the numerical solution of the system, one can transform the model problem (12) to assume again a formulation with only affine parameter dependence as in the standard problem (12). In fact, a multiplication with $\exp(-a(x,y))$ and an application of the product rule yields the equivalent convection diffusion problem

$$-\Delta u(y,x) - \nabla a(y,x) \cdot \nabla u(y,x) = e^{-a(y,x)} f(y,x) \qquad \text{for } (y,x) \in \Gamma \times D. \tag{46}$$

One can show easily that any weak solution of the original problem (43) is also a weak solution of (46) and vice versa and it holds stability in the sense

$$\|u(y)\|_V \leq c(y)\|f\|_{V^*} \quad \text{for all } y \in \Gamma$$

with some y-dependent constant $c(y)$ that bounds $\kappa(y,\cdot)$ from below. For practical evaluations, the reformulation (46) has significant advantages, since the coefficient $a(x,y)$ appears linearly (affine dependence on y) in the stochastic part of the differential operator. Consequently, the coupling structure is much sparser than in the lognormal case and akin the affine model problem, see [UEE12] for further discussions.

The weak formulation of (46) involves the operator

$$\begin{aligned} A(u,v) &:= \int_\Gamma \int_D \nabla u(y,x) \cdot \nabla v(y,x) - \nabla a(x) \cdot \nabla u(y,x) v(y,x) \,\mathrm{d}x \,\mathrm{d}\pi(y) \\ &:= \int_\Gamma \int_D \nabla u(y,x) \cdot \nabla v(y,x) - \sum_{m=1}^{\infty} y_m \nabla a_m(x) \cdot \nabla u(y,x) v(y,x) \,\mathrm{d}x \,\mathrm{d}\pi(y). \end{aligned}$$

Moreover, the right-hand side is given by

$$F(v) := \int_\Gamma \int_D e^{-a} f(x) v(y,x) \,\mathrm{d}x \,\mathrm{d}\pi(y). \tag{47}$$

With this, the weak solution $u \in \mathcal{V} := L^2_\pi(\Gamma; V)$ is determined by

$$A(u,v) = F(v) \quad \text{for all } v \in \mathcal{V}.$$

7.2 Discretization

The main steps to derive a discretization by the SGFEM are the same as in the affine case. Based on a triangulation $\mathcal{T}$ and a discrete set of active stochastic modes Λ, we seek the coefficient functions $u_{N,\mu} \in V_h$ for $\mu \in \Lambda$ in the expansion

$$u_N(y,x) = \sum_{\mu \in \Lambda} u_{N,\mu}(x) P_\mu(y).$$

One important difference however is that this time orthogonal polynomials with respect to $\Gamma_m = [-\infty, \infty]$ and the probability measure with Gaussian density $\pi_m(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ are needed. This leads to Hermite polynomials and their multivariate tensorizations $\{P_\mu : \mu \in \mathcal{F}\}$ that again satisfy a recurrence relation of the form (11).

For the discretization of the right-hand side functional F in (47), we now have to use a representation of $e^{-a} f$ of the form

$$e^{-a} f = \left(\sum_\mu \lambda_\mu P_\mu \right) f = \sum_\mu f_\mu P_\mu \qquad \text{where} \qquad f_\mu := \lambda_\mu f. \tag{48}$$

The following lemma provides a formula to compute the coefficient functions $\lambda_\mu := \int_\Gamma e^{-a} P_\mu \,\mathrm{d}\pi(y)$ analytically[6].

[6] If f is stochastic (i.e. depends on parameters), the coefficient $f_\mu := \int_\Gamma e^{-a} f P_\mu \,\mathrm{d}\pi(y)$ has to be determined by quadrature.

Lemma 7.1. *It holds that*

$$\exp(-a(x,y)) = \exp\left(\frac{1}{2}\sum_{m=1}^{\infty} a_m(x)^2\right) \sum_{\mu\in\mathcal{F}} (-1)^{|\mu|} \frac{a(x)^\mu}{\mu!} P_\mu(y).$$

Proof. The generating function relation

$$\exp\Big(- a_m(x)y_m - \frac{a_m(x)^2}{2}\Big) = \sum_{k=1}^{\infty} \frac{a_m(x)^k}{k!} P_k(y_m) \tag{49}$$

yields the expansion

$$\begin{aligned}\exp(-a(x,y)) &= \exp\left(-\sum_{m=1}^{\infty} a_m(x)y_m\right) = \prod_{m=1}^{\infty} \exp(-a_m(x)y_m)\\ &= \prod_{m=1}^{\infty}\left(\frac{1}{2}\exp\left(a_m(x)^2\right)\sum_{k=1}^{\infty}(-1)^k \frac{a_m(x)^k}{k!}P_k(y_m)\right).\end{aligned} \tag{50}$$

Expanding the products completes the proof. □

With this, the SGFEM seeks a discrete $u_N \in \mathcal{V}_h := \left\{\sum_{\mu\in\Lambda} v_{h,\mu}P_\mu : v_{h,\mu} \in V_h\right\}$ with

$$A(u_N, v_h) = F(v_h) \quad \text{for all } v_h \in \mathcal{V}_h.$$

Thanks to the representation (48) and the orthogonality of P_μ with respect to the Gaussian measure π, the sum in the evaluation of the right-hand side is also finite, i.e., for any $v_h = \sum_\mu v_{h,\mu}P_\mu \in \mathcal{V}_h$ it holds that

$$F(v_h) = \sum_{\nu\in\mathcal{F},\mu\in\Lambda} \int_\Gamma P_\nu P_\mu \,\mathrm{d}\pi(y) \int_D f_\nu v_{h,\mu}\,\mathrm{d}x = \sum_{\mu\in\Lambda}\int_D f_\nu v_{h,\mu}\,\mathrm{d}x.$$

Remark 7.2. The system matrix can (again) be written in tensor operator form

$$A = G_0 \otimes A_0 + \sum_{m=1}^{M} G_m \otimes A_m, \tag{51}$$

where A_0 is the classical deterministic discrete Laplacian, i.e., the representation of $(\nabla u_h, \nabla v_h)_{L^2(D)}$, and A_m are representations of the convection operators $(\nabla a_m \cdot \nabla u_h, v_h)_{L^2(D)}$ for $u_h, v_h \in V_h$. The matrices G_0 and G_m are defined as in the affine case, see Section 3.3. Efficient preconditioners for the log-transformed problem are discussed, e.g., in [UEE12].

7.3 Error estimator

To devise an (residual based) error estimator, the main idea is to treat the discrete solution component $u_{N,\mu}$ as an approximation to the perturbed Poisson problem

$$-\Delta u_\mu(x) = f_\mu(x) + \int_\Gamma P_\mu(y)\nabla a(y,x)\cdot\nabla u(x)\,\mathrm{d}\pi(y),$$

which is motivated by the Galerkin orthogonality $r_\mu(w_h) := A(u - u_N, P_\mu w_h) = 0$ for all $w_h \in V_h$ and $\mu \in \Lambda$ for the subresidual

$$r_\mu(w_h) = \int_D (f_\mu + \lambda_\mu) w_h \,\mathrm{d}x - \int_D \nabla u_{N,\mu} \cdot \nabla w_h \,\mathrm{d}x$$

$$\text{where} \qquad \lambda_\mu := \int_\Gamma \nabla a \cdot \nabla u_N P_\mu \,\mathrm{d}\pi = \sum_{m=1}^{\infty} \nabla a_m \cdot \nabla \left(\frac{\alpha_m}{\beta_m} u_{N,\mu+\epsilon_m} + \frac{\gamma_m}{\beta_m} u_{N,\mu-\epsilon_m} \right).$$

To bound its dual norm

$$\|r_\mu\|_{V^\star} := \sup_{v_h \in V_h \setminus \{0\}} \frac{|r_\mu(v)|}{\|\nabla v_h\|_{L^2(D)}},$$

classical deterministic theory leads to the error estimator

$$\eta_\mu^2 = \sum_{T \in \mathcal{T}} \eta_{\mu,T}^2 \quad \text{with} \quad \eta_{\mu,T}^2 := h_T^2 \|f_\mu + \lambda_\mu + \Delta_h u_{N,\mu}\|_{L^2(T)}^2 + \sum_{F \in \mathcal{E}(T)} h_E \|[[\nabla u_{N,\mu} \cdot \mathbf{n}_E]]\|_{L^2(E)}^2. \tag{52}$$

Similarly, for $\mu \in \mathcal{F} \setminus \Lambda$ (without Galerkin orthogonality) one obtains

$$\eta_\mu^2 := \|f_\mu + \lambda_\mu\|_{L^2(D)}^2. \tag{53}$$

As in the affine coefficient case, a split of the dual norm of the total error residual $R(v) = A(u - u_N, v)$ into subresiduals is possible, namely

$$\|R\|_{L^2_\pi(\Gamma;V)^\star}^2 = \sum_{\mu \in \mathcal{F}} \|r_\mu\|_{V^\star}^2 = \sum_{\mu \in \Lambda} \|r_\mu\|_{V^\star}^2 + \sum_{\mu \in \mathcal{F} \setminus \Lambda} \|r_\mu\|_{V^\star}^2.$$

This motivates the decomposition into the spatial and stochastic errors

$$\eta^2(\Lambda) := \sum_{\mu \in \Lambda} \eta_\mu^2 \quad \text{and} \quad \eta^2(\partial_h \Lambda) := \sum_{\mu \in \partial_h \Lambda} \eta_\mu^2.$$

As before, $\partial_h \Lambda$ denotes a finite-dimensional approximation of the infinite-dimensional remainder $\mathcal{F} \setminus \Lambda$.

Remark 7.3*.* Opposite to the affine case with deterministic right-hand side, this time one cannot assume $r_\nu \equiv 0$ for $\nu \in \mathcal{F} \setminus (\Lambda \cup \partial\Lambda)$. In some situations it probably makes sense to modify the definition of the discrete boundary (42) to include multi-indices that are more than an ϵ_m away from multi-indices in Λ, i.e., multi-indices from $\mathcal{F} \setminus (\Lambda \cup \partial\Lambda)$. However, this is not further examined here.

7.4 A numerical example

This contribution concludes with a brief numerical example to illustrate that the suggested error estimator is reliable and efficient and leads to reasonable mesh refinements. Further details and proofs are the content of a future publication [EGM25].

The experiment is conducted on the L-shaped domain D_L with data $f \equiv 1$ and the coefficient κ chosen as

$$\kappa(x, y) := \exp\left(\sum_{m=1}^{\infty} a_m(x) y_m \right) \qquad \text{where}$$

$$a_m(x) := \gamma m^{-\sigma} \cos(2\pi \varrho_1(m) x_1) \cos(2\pi \varrho_2(m) x_2)$$

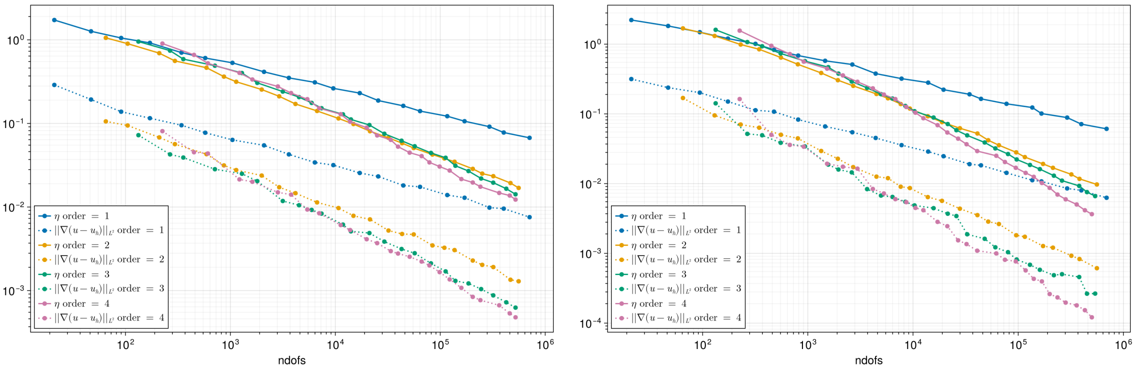

Figure 9: L-shaped domain logtransformed case: Convergence history of the error and the explicit residual-based error estimator for order $k \in \{1, \dots, 4\}$ with decay $\sigma = 2$ (left) and $\sigma = 4$ (right).

with $\gamma = 1$. The coefficients ϱ_1 and ϱ_2 are computed as in the affine case, see Section 6.1.

Algorithm 1 is performed with the error estimator from the previous section and the same parameters for $\partial_h \Lambda$ as in the affine case until the total number of degrees exceeds $5 \cdot 10^5 < \dim(\Lambda) \cdot \dim(V_h)$. Again the exact error $\|\nabla(u - u_N)\|_{L^2} = \int_\Gamma \|\nabla(u(y) - u_N(y))\|_V^2 \,\mathrm{d}\pi(y)$ is approximated by Monte Carlo sampling as described in Section 6 for the affine case.

Figure 9 shows the convergence history of the error estimator suggested above and the exact error for polynomial order $k = 1, \dots, 4$. As in the affine case, the error estimator seems reliable and efficient and leads to reasonable refinement in V_h and Λ. Qualitatively similar (despite the mesh refinement) but undocumented results are obtained for the square domain. This illustrates that the affine reformulation of the otherwise numerically quite challenging lognormal Darcy problem performs quite similarly to the well-known affine case with a sparse ASGFEM discretization.

8 Summary and Outlook

The stochastic Galerkin method is a reliable and efficient method to tackle high-dimensional parametric PDEs that has been one of the standard numerical approaches in Uncertainty Quantification since the scientific field became popular. As in the classical deterministic case, it computes a Galerkin projection of the parameter-to-solution map onto a finite dimensional product space of a spatial finite element space and a parameter space spanned by a linear combination of orthogonal polynomials with respect to stochastic dimensions encoded in a set of multi-indices. This property renders Galerkin methods particularly well suited for adaptive methods, which aim to achieve (quasi-)optimality with respect to the convergence rate as well as the overall complexity with a problem-adapted refinement strategy. In our parametric setting, the purpose of adaptive error control comprises not only the identification of areas of lower regularity in the physical domain, resulting in appropriate refinements of the finite element space, but also the balanced selection of the most influential stochastic dimensions. An adaptive algorithm that incorporates these two tasks is presented together with an overview over known results regarding its convergence. Modern equilibration error estimators, built from known recipes for deterministic problems, even allow for guaranteed error bounds. Interestingly, inspired by deterministic adaptive FEM, convergence of the adaptive algorithm can be shown at least in the affine setting. A generalization was developed in [EH23], where it is pointed out that a uniform error reduction cannot be expected with the usual analytical techniques. Moreover, the question of optimality is still an open active research topic and possibly requires involved implementations as in [BEEV24] or additional (as yet hard to verify) assumptions as in [BPR22].

While this contribution mainly summarizes well-established results for the stationary linear case with affine coefficients, much less is known for nonlinear problems or cases with non-affine coefficients. In particular the possible lack of uniform boundedness of the operator poses significant challenges for the theoretical and numerical treatment and likely requires the development of new mathematical tools. As an example, the well-known log-normal case was discussed, where it is possible to transform the problem to a structurally more convenient affine case via an appropriate transformation. Nevertheless, the involved (energy) norms have to be handled carefully to derive a posteriori error estimates and more advanced results such as convergence of an adaptive numerical scheme do not exist yet. A new approach towards a more general analysis of parametric PDEs, which in particular does not rely on the common holomorphic parameter dependence, was recently presented in [ADF$^+$23].

Another promising direction is to consider approximate Galerkin approximations, e.g. in a statistical learning framework. As representation scheme, neural networks have become ubiquitous in many application areas and there exist way too many approaches to attempt to provide an overview here. However, proper error control and refinement of the approximation quality remain open research topics despite the respectable success that were already obtained also with complicated problems. This is somewhat different for low-rank tensor formats such as the popular tensor trains. Using a tensor reconstruction (in terms of a least squares optimization) as in [EFHT23a], reliable error control also for non-affine and in principle non-linear problems can be obtained, at least with high probability. Moreover, the implementation is much less involved than specific (sparse) Galerkin discretizations and can be used for a wider range of problems.

Generally speaking, this (representation independent) direction of *statistical operator learning* is a promising research area, which will likely play an important role in the upcoming improvements of numerical methods for high-dimensional PDEs, especially when it comes to the solution of non-linear real-world problems.

References

[ADF$^+$23] Xin An, Josef Dick, Michael Feischl, Andrea Scaglioni, and Thanh Tran. Sparse grid approximation of the stochastic Landau-Lifshitz-Gilbert equation. *arXiv preprint arXiv:2310.11225*, 2023.

[BC24] Markus Bachmayr and Albert Cohen. Multilevel representations of random fields and sparse approximations of solutions to random PDEs. In *Multiscale, Nonlinear and Adaptive Approximation II*, pages 25–54. Springer, 2024.

[BCDM17] Markus Bachmayr, Albert Cohen, Ronald DeVore, and Giovanni Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part II: lognormal coefficients. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(1):341–363, 2017.

[BCM17] Markus Bachmayr, Albert Cohen, and Giovanni Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(1):321–339, 2017.

[BDD04] Peter Binev, Wolfgang Dahmen, and Ron DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004.

[BEEV24] Markus Bachmayr, Martin Eigel, Henrik Eisenmann, and Igor Voulis. A convergent adaptive finite element stochastic Galerkin method based on multilevel expansions of random fields. *arXiv preprint arXiv:2403.13770*, 2024.

[BNT07] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3):1005–1034, 2007.

[BNT10] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM review*, 52(2):317–355, 2010.

[BPR22] Alex Bespalov, Dirk Praetorius, and Michele Ruggeri. Convergence and rate optimality of adaptive multilevel stochastic Galerkin FEM. *IMA Journal of Numerical Analysis*, 42(3):2190–2213, 2022.

[BPRR19a] Alex Bespalov, Dirk Praetorius, Leonardo Rocchi, and Michele Ruggeri. Convergence of adaptive stochastic Galerkin FEM. *SIAM Journal on Numerical Analysis*, 57(5):2359–2382, 2019.

[BPRR19b] Alex Bespalov, Dirk Praetorius, Leonardo Rocchi, and Michele Ruggeri. Goal-oriented error estimation and adaptivity for elliptic PDEs with parametric or uncertain inputs. *Computer Methods in Applied Mechanics and Engineering*, 345:951–982, 2019.

[BPS12] Alexei Bespalov, Catherine E Powell, and David Silvester. A priori error analysis of stochastic Galerkin mixed approximations of elliptic PDEs with random data. *SIAM Journal on Numerical Analysis*, 50(4):2039–2063, 2012.

[BPS14] Alex Bespalov, Catherine E Powell, and David Silvester. Energy norm a posteriori error estimation for parametric operator equations. *SIAM Journal on Scientific Computing*, 36(2):A339–A363, 2014.

[BS08] Dietrich Braess and Joachim Schöberl. Equilibrated residual error estimator for edge elements. *Mathematics of Computation*, 77(262):651–672, 2008.

[BS16] Alex Bespalov and David Silvester. Efficient adaptive stochastic Galerkin methods for parametric operator equations. *SIAM Journal on Scientific Computing*, 38(4):A2118–A2140, 2016.

[BS23] Alex Bespalov and David Silvester. Error estimation and adaptivity for stochastic collocation finite elements part ii: multilevel approximation. *SIAM Journal on Scientific Computing*, 45(2):A781–A797, 2023.

[BSU16] Markus Bachmayr, Reinhold Schneider, and André Uschmajew. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Found. Comput. Math.*, 16(6):1423–1472, 2016.

[BTZ05] Ivo Babuška, Raúl Tempone, and Georgios E Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Computer methods in applied mechanics and engineering*, 194(12-16):1251–1294, 2005.

[BV22] Markus Bachmayr and Igor Voulis. An adaptive stochastic Galerkin method based on multilevel expansions of random fields: Convergence and optimality. *ESAIM: Mathematical Modelling and Numerical Analysis*, 56(6):1955–1992, 2022.

[CD13] Julia Charrier and Arnaud Debussche. Weak truncation error estimates for elliptic PDEs with lognormal coefficients. *Stochastic partial differential equations: analysis and computations*, 1(1):63–93, 2013.

[CD15] Albert Cohen and R DeVore. High dimensional approximation of parametric PDEs. *Acta Numerica*, 2015.

[CDS10] Albert Cohen, Ronald DeVore, and Christoph Schwab. Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.

[CDS11] Albert Cohen, Ronald Devore, and Christoph Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.

[CFPP14] Carsten Carstensen, Michael Feischl, Marcus Page, and Dirk Praetorius. Axioms of adaptivity. *Computers & Mathematics with Applications*, 67(6):1195–1253, 2014.

[CKNS08] J. Manuel Cascon, Christian Kreuzer, Ricardo H. Nochetto, and Kunibert G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008.

[CM47] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann. of Math. (2)*, 48:385–392, 1947.

[Dör96] Willy Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM Journal on Numerical Analysis*, 33(3):1106–1124, 1996.

[EEST22] Martin Eigel, Oliver G Ernst, Björn Sprungk, and Lorenzo Tamellini. On the convergence of adaptive stochastic collocation for elliptic partial differential equations with affine diffusion. *SIAM Journal on Numerical Analysis*, 60(2):659–687, 2022.

[EFHT23a] Martin Eigel, Nando Farchmin, Sebastian Heidenreich, and Philipp Trunschke. Adaptive non-intrusive reconstruction of solutions to high-dimensional parametric pdes. *SIAM Journal on Scientific Computing*, 45(2):A457–A479, 2023.

[EFHT23b] Martin Eigel, Nando Farchmin, Sebastian Heidenreich, and Philipp Trunschke. Efficient approximation of high-dimensional exponentials by tensor networks. *International Journal for Uncertainty Quantification*, 13(1), 2023.

[EGM25] Martin Eigel, Claude-J. Gittelson, and Christian Merdon. Adaptive stochastic Galerkin FEM for a log-transformed PDE with non-affine unbounded random coefficients. *in preparation*, 2025^{+}.

[EGSZ14] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. Adaptive stochastic Galerkin FEM. *Comput. Methods Appl. Mech. Engrg.*, 270:247–269, 2014.

[EGSZ15] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(5):1367–1398, 2015.

[EH23] Martin Eigel and Nando Hegemann. Guaranteed quasi-error reduction of adaptive Galerkin FEM for parametric PDEs with lognormal coefficients. *arXiv preprint arXiv:2302.02839*, 2023.

[EHL$^+$13] Mike Espig, Wolfgang Hackbusch, Alexander Litvinenko, Hermann G Matthies, and Elmar Zander. Efficient analysis of high dimensional data in tensor formats. In *Sparse Grids and Applications*, pages 31–56. Springer, 2013.

[EM16a] Martin Eigel and Christian Merdon. Equilibration a posteriori error estimation for convection–diffusion–reaction problems. *Journal of Scientific Computing*, 67(2):747–768, 2016.

[EM16b] Martin Eigel and Christian Merdon. Local equilibration error estimators for guaranteed error control in adaptive stochastic higher-order Galerkin finite element methods. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1372–1397, 2016.

[EMPS20] Martin Eigel, Manuel Marschall, Max Pfeffer, and Reinhold Schneider. Adaptive stochastic Galerkin FEM for lognormal coefficients in hierarchical tensor representations. *Numerische Mathematik*, 145(3):655–692, 2020.

[EMSU12] Oliver G. Ernst, Antje Mugler, Hans-Jörg Starkloff, and Elisabeth Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM Math. Model. Numer. Anal.*, 46(2):317–339, 2012.

[EPS17] Martin Eigel, Max Pfeffer, and Reinhold Schneider. Adaptive stochastic Galerkin FEM with hierarchical tensor representations. *Numerische Mathematik*, 136(3):765–803, 2017.

[EPSU09] O.G. Ernst, C.E. Powell, D.J. Silvester, and E. Ullmann. Efficient solvers for a linear stochastic Galerkin mixed formulation of diffusion problems with random data. *SIAM Journal on Scientific Computing*, 31(2):1424–1447, 2009.

[EST18] Oliver G Ernst, Björn Sprungk, and Lorenzo Tamellini. Convergence of sparse collocation for functions of countably many gaussian random variables (with application to elliptic PDEs). *SIAM Journal on Numerical Analysis*, 56(2):877–905, 2018.

[EST22] Martin Eigel, Reinhold Schneider, and Philipp Trunschke. Convergence bounds for empirical nonlinear least-squares. *ESAIM: Mathematical Modelling and Numerical Analysis*, 56(1):79–104, 2022.

[FS21] Michael Feischl and Andrea Scaglioni. Convergence of adaptive stochastic collocation with finite elements. *Computers & Mathematics with Applications*, 98:139–156, 2021.

[FST05] Philipp Frauenfelder, Christoph Schwab, and Radu Alexandru Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer methods in applied mechanics and engineering*, 194(2-5):205–228, 2005.

[Git10] C. J. Gittelson. Stochastic Galerkin discretization of the log-normal isotropic diffusion problem. *Math. Models Methods Appl. Sci.*, 20(2):237–263, 2010.

[Git13] Claude Gittelson. An adaptive stochastic Galerkin method for random elliptic operators. *Mathematics of Computation*, 82(283):1515–1541, 2013.

[Git14] Claude Jeffrey Gittelson. High-order methods as an alternative to using sparse tensor products for stochastic Galerkin FEM. *Computers & Mathematics with Applications*, 67(4):888–898, 2014.

[GK96] Roger G Ghanem and Robert M Kruger. Numerical solution of spectral stochastic finite element systems. *Computer Methods in Applied Mechanics and Engineering*, 129(3):289–303, 1996.

[GN18] Diane Guignard and Fabio Nobile. A posteriori error estimation for the stochastic collocation finite element method. *SIAM Journal on Numerical Analysis*, 56(5):3121–3143, 2018.

[GNP16] Diane Guignard, Fabio Nobile, and Marco Picasso. A posteriori error estimation for elliptic partial differential equations with small uncertainties. *Numer. Methods Partial Differ. Equations*, 32(1):175–212, 2016.

[GS91] Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[GS09] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity. *SIAM J. Numer. Anal.*, 47(5):3624–3651, 2009.

[HANT16] Abdul-Lateef Haji-Ali, Fabio Nobile, and Raúl Tempone. Multi-index monte carlo: when sparsity meets sampling. *Numerische Mathematik*, 132:767–806, 2016.

[HPS15] Helmut Harbrecht, Michael Peters, and Markus Siebenmorgen. Efficient approximation of random fields for numerical applications. *Numerical Linear Algebra with Applications*, 22(4):596–617, 2015.

[HPS16] Helmut Harbrecht, Michael Peters, and Markus Siebenmorgen. Multilevel accelerated quadrature for PDEs with log-normally distributed diffusion coefficient. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):520–551, 2016.

[HS14] Viet Ha Hoang and Christoph Schwab. N-term Wiener chaos approximation rate for elliptic PDEs with lognormal Gaussian random inputs. *Math. Models Methods Appl. Sci.*, 24(4):797–826, 2014.

[HS17] L Herrmann and Ch Schwab. Qmc algorithms with product weights for lognormal-parametric, elliptic PDEs. 2017.

[Jan97] Svante Janson. *Gaussian hilbert spaces*. Number 129. Cambridge university press, 1997.

[Kee03] Andreas Keese. *A review of recent developments in the numerical solution of stochastic partial differential equations (stochastic finite elements)*. Univ.-Bibl., 2003.

[KM03] Andreas Keese and Hermann G Matthies. Hierarchical parallel solution of stochastic systems. In *Computational Fluid and Solid Mechanics 2003*, pages 2023–2025. Elsevier, 2003.

[LMK10] Olivier Le Maître and Omar M Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media, 2010.

[Loè77] Michel Loève. *Probability Theory. I*. Springer-Verlag, New York-Heidelberg, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[LPS14] Gabriel J. Lord, Catherine E. Powell, and Tony Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[Mat08] Hermann G Matthies. Stochastic finite elements: Computational approaches to stochastic partial differential equations. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik: Applied Mathematics and Mechanics*, 88(11):849–873, 2008.

[Mer13] Christian Merdon. *Aspects of guaranteed error control in computations for partial differential equations*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2013.

[MZ12] Hermann G Matthies and Elmar Zander. Solving stochastic systems with low-rank tensor compression. *Linear Algebra and its Applications*, 436(10):3819–3838, 2012.

[Nou17] Anthony Nouy. Low-rank methods for high-dimensional approximation and model order reduction. *Model reduction and approximation, P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, eds., SIAM, Philadelphia, PA*, pages 171–226, 2017.

[NTTT16] Fabio Nobile, Lorenzo Tamellini, Francesco Tesei, and Raúl Tempone. An adaptive sparse grid algorithm for elliptic PDEs with lognormal diffusion coefficient. In *Sparse Grids and Applications-Stuttgart 2014*, pages 191–220. Springer, 2016.

[SG11] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[ST06a] Christoph Schwab and Radu Alexandru Todor. Karhunen–loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.

[ST06b] Christoph Schwab and Radu Alexandru Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[SZ90] L. Ridgway Scott and Shangyou Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Mathematics of Computation*, 54(190):483–493, 1990.

[TSGU13] Aretha L Teckentrup, Robert Scheichl, Michael B Giles, and Elisabeth Ullmann. Further analysis of multilevel monte carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.

[UEE12] Elisabeth Ullmann, Howard C. Elman, and Oliver G. Ernst. Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems. *SIAM J. Sci. Comput.*, 34(2):A659–A682, 2012.

[Ull10] Elisabeth Ullmann. A Kronecker product preconditioner for stochastic Galerkin finite element discretizations. *SIAM Journal on Scientific Computing*, 32(2):923–946, 2010.

[UP15] Elisabeth Ullmann and Catherine E Powell. Solving log-transformed random diffusion problems by stochastic Galerkin mixed finite element methods. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):509–534, 2015.

[Ver13] Rüdiger Verfürth. *A Posteriori Error Estimation Techniques for Finite Element Methods*. Numerical Methods and Scientific Computation. Oxford University Press, 2013.

[Voh11] Martin Vohralík. Guaranteed and fully robust a posteriori error estimates for conforming discretizations of diffusion problems with discontinuous coefficients. *Journal of Scientific Computing*, 46(3):397–438, 2011.

[Wie38] Norbert Wiener. The Homogeneous Chaos. *Amer. J. Math.*, 60(4):897–936, 1938.

[XK02] Dongbin Xiu and George Em Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.