MOX–Report No. 04/2007

An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data

Fabio Nobile, Raul Tempone, Clayton G. Webster

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 29 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data

F. Nobile[1], R. Tempone[2,3], C. G. Webster[2]

[1]MOX, Dipartimento di Matematica, Politecnico di Milano, Italy
[2]SCS and Dep. of Mathematics, Florida State University, Tallahassee, FL, USA
[3]Dalhquist Fellow, Numerical Analysis, KTH, S-100 44 Stockholm, Sweden

Abstract

This work proposes and analyzes an anisotropic sparse grid stochastic collocation method for solving elliptic partial differential equations with random coefficients and forcing terms (input data of the model). The method consists of a Galerkin approximation in the space variables and a collocation, in probability space, on sparse tensor product grids utilizing either Clenshaw-Curtis or Gaussian knots. Even in the presence of nonlinearities, the collocation approach leads to the solution of uncoupled deterministic problems, just as in the Monte Carlo method. This work includes *a priori* and *a posteriori* procedures to adapt the anisotropy of the sparse grids to each given problem. These procedures seem to be very effective for the problems under study. The proposed method combines the advantages of isotropic sparse collocation with those of anisotropic full tensor product collocation: the first approach is effective for problems depending on random variables which weigh approximately equally in the solution, while the benefits of the latter approach become apparent when solving highly anisotropic problems depending on a relatively small number of random variables, as in the case where input random variables are Karhunen-Loève truncations of "smooth" random fields. This work also provides a rigorous convergence analysis of the fully discrete problem and demonstrates: (sub)-exponential convergence in the asymptotic regime and algebraic convergence in the pre-asymptotic regime, with respect to the total number of collocation points. Numerical examples illustrate the theoretical results and are used to compare this approach with several others, including the standard Monte Carlo. In particular, for moderately large dimensional problems, the sparse grid approach with a properly chosen anisotropy seems to be very efficient and superior to all examined methods.

Key words: Collocation techniques, PDEs with random data, differential equations, finite elements, uncertainty quantification, anisotropic sparse grids, Smolyak algorithm, multivariate polynomial interpolation.

Introduction

Stochastic formulations provide a natural way to include uncertainty quantification in practical applications. Uncertainty usually appears through in the input data of the problem, such as model coefficients, forcing terms, boundary conditions, geometry, etc. See for instance [1, 25] and the references therein.

In this work we focus on the solution of elliptic partial differential equations (PDEs) with random coefficients and forcing terms (input data of the model), and especially address the situation where the input data are assumed to depend on a moderately large number of random variables.

The method proposed here, namely an anisotropic Sparse Grid Stochatic Collocation, extends the isotropic method analyzed in [25], and consists of a Galerkin approximation in the space variables and a collocation, in probability space, on sparse tensor product grids utilizing either Clenshaw-Curtis or Gaussian knots. As a consequence of the collocation approach our technique naturally lead to the solution of uncoupled deterministic problems, just as in the Monte Carlo method.

The work [25] revealed that the isotropic sparse collocation algorithm is very effective for problems whose input data depend on a moderate number of random variables, which "weigh equally" in the solution. For such isotropic situations the displayed convergence is faster than standard collocation techniques built upon full tensor product spaces. On the other hand, the convergence rate of the isotropic sparse collocation algorithm [25] deteriorates for highly anisotropic problems, such as those appearing when the input random variables come e.g. from Karhunen-Loève -type truncations of "smooth" random fields. In such cases, a full anisotropic tensor product approximation may still be more effective for a small or modest number of random variables. However, if the number of random variables is large, the construction of the full tensor product spaces becomes infeasible, since the dimension of the approximating space grows exponentially fast with respect to the number of random variables in the problem.

The main contribution of this work is to propose and analyze the use of anisotropic sparse tensor product spaces constructed from a weighted Smolyak interpolant with suitable abscissas. This approach is particularly attractive in the case of truncated expansions of random fields, since the anisotropy can be tuned to the decay properties of the expansion. We will present *a priori* and *a posteriori* procedures for choosing the anisotropy of the sparse grids which are extremely effective for the problems under study.

This work provides a rigorous convergence analysis of the fully discrete problem and demonstrates: (sub)-exponential convergence in the asymptotic regime and algebraic convergence in the pre-asymptotic regime, with respect to the total number of collocation points. Numerical examples illustrate the theoretical results and are used to compare this approach with several others, including the standard Monte Carlo. In particular, for moderately large dimensional problems, the sparse grid approach with properly chosen anisotropy seems to be very

efficient and superior to all examined methods.

The Monte Carlo method described in [11] is the classical and most used technique for approximating expected values of quantities of interest depending on the solution of a partial differential equations with random inputs. The algorithm approximates the desired expectation by a sample average of independent identically distributed (iid) realizations. When solving partial differential equations with random inputs, this method implies the solution on one deterministic differential equation for each realization of the input parameters. This makes the method simple to implement, allows for maximal code reusability and it is straightforward to parallelize. Its numerical error is approximately $\mathcal{O}(1/\sqrt{M})$, where M is the number of realizations. The advantage of utilizing this approach is that the rate does not deteriorate with respect to the number of random variables in the problem, making the method very attractive for problems with large dimensional random inputs. On the other hand, when solving large-scale applications, the exponent $1/2$ in the rate of convergence generates a tremendous amount of computational work required to achieve accurate solutions. Other ensemble based methods like Quasi Monte Carlo, Latin Hypercube Sampling, etc. (see e.g. [24, 19] and references therein), have been devised to produce faster convergence, $\mathcal{O}(\log(M)^r/M)$, where the coefficient $r > 0$ becomes larger with the dimension of the random input. We explore alternative methods that obtain faster convergence rates, exploiting the high regularity that the solution of elliptic PDEs may have with respect to the random input, while preserving the implementation advantages of ensemble-based methods.

Another possible substitute for the Monte Carlo method is the so called Spectral Galerkin method, see e.g. [18]. It employs standard approximations in space (finite elements, finite volumes, spectral or h-p finite elements, etc. and polynomial approximation in the probability domain, either by full polynomial spaces [33, 23, 17], tensor product polynomial spaces [2, 12, 28] or piecewise polynomial spaces [2, 20]. This family of methods exploit the regularity of the solution to acquire faster convergence rates. However, in general, this technique requires solving a system of equations that couples all degrees of freedom in the approximation to the stochastic solution.

A new numerical technique, which has gained much attention recently by the computational community, see [1, 13, 25, 22, 32], is Stochastic Collocation, which can be based either on full or sparse tensor product approximation spaces. As shown in [1], Stochastic Collocation essentially preserves the fast convergence of the Spectral Galerkin method, even coinciding in particular cases, while maintaining an ensemble-based approach, just as Monte Carlo. Hence, Stochastic collocation seems to be an ideal method for computing realistic statistics from solutions of PDEs with random input data, but the challenge comes when solving problems with a relatively large number of input random variables. Therefore, our goal is investigating variants of this procedure to attain highly accurate solutions while reducing the *curse of dimensionality*. To extend the applicability of the collocation method to cases with large dimensional random inputs, several

sparse-approximation-based methods have been recently explored [32, 25, 13]. Aligned with this research effort, the main contribution of this work is to propose and analyze a novel anisotropic stochastic collocation method that is based on a weighted version of the Smolyak algorithm. As our numerical and theoretical results indicate, this method seems very promising and worth exploring further, both from the implementation and the theoretical point of view.

The outline of the paper is the following: in Section 1 we introduce the mathematical problem and the main notation used throughout. There we also state assumptions on the parametrization of the random inputs, which is useful when later transforming the original stochastic problem into a deterministic parametric one, and on the problem solution's regularity, which is used to later prove the error estimates in Section 5.

In Section 2 we focus on applications to linear elliptic PDEs with random input data. The main idea here is to verify the assumptions from Section 1 in a particular setting, showing that they are justified and that one can extend them to a variety of problems.

In Section 3 we introduce the approximation spaces and provide an overview of various collocation techniques. We also describe the anisotropic sparse approximation method to be considered as well as the different interpolation techniques to be employed. Next, in Section 4, we provide *a priori* and *a posteriori* procedures for tuning the anisotropy of our sparse grid method to the problem at hand. In Section 5 we provide a detailed error analysis of the method, including cases where the sparse interpolant uses both Clenshaw-Curtis and Gaussian abscissas. This analysis relies on the regularity of the solution and exploits the behavior of sparse approximations from the anisotropic Smolyak method. Finally, in Section 6 we present numerical results, including a comparison with the Monte Carlo method, showing the efficiency of the proposed method.

1 Problem setting

Similarly to [25] we begin by focusing our attention on an elliptic operator $\mathcal{L}$, linear or nonlinear, on a domain $D \subset \mathbb{R}^d$, which depends on some coefficients $a(\omega, x)$ with $x \in D$, $\omega \in \Omega$ and $(\Omega, \mathcal{F}, P)$ a complete probability space. Here Ω is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is the σ-algebra of events and $P : \mathcal{F} \to [0, 1]$ is a probability measure. Analogously, the forcing term $f = f(\omega, x)$ can be assumed random as well.

Consider the stochastic elliptic boundary value problem: find a random function, $u : \Omega \times \overline{D} \to \mathbb{R}$, such that P-almost everywhere in Ω, or in other words almost surely (a.s.), the following equation holds:

$$\mathcal{L}(a)(u) = f \quad \text{in } D \tag{1.1}$$

equipped with suitable boundary conditions. Before introducing some assumptions we denote by $W(D)$ a Banach space of functions $v : D \to \mathbb{R}$ and define,

for $q \in [1, \infty]$, the stochastic Banach spaces

$$L_P^q(\Omega; W(D)) := \Big\{ v : \Omega \to W(D) \mid v \text{ is strongly measurable} \\ \text{and } \int_\Omega \|v(\omega, \cdot)\|_{W(D)}^q dP(\omega) < +\infty \Big\}$$

and

$$L_P^\infty(\Omega; W(D)) := \Big\{ v : \Omega \to W(D) \mid v \text{ is strongly measurable} \\ \text{and } P - \operatorname*{ess\,sup}_{\omega \in \Omega} \|v(\omega, \cdot)\|_{W(D)}^2 < +\infty \Big\}.$$

Of particular interest is the space $L_P^2(\Omega; W(D))$, consisting of Banach valued functions that have finite second moments.

We will now make the following assumptions:

A_1) the solution to (1.1) has realizations in the Banach space $W(D)$, i.e. $u(\cdot, \omega) \in W(D)$ almost surely and $\forall \omega \in \Omega$

$$\|u(\cdot, \omega)\|_{W(D)} \leq C \|f(\cdot, \omega)\|_{W^*(D)}$$

where we denote $W^*(D)$ to be the dual space of $W(D)$, and C is a constant independent of the realization $\omega \in \Omega$.

A_2) the forcing term $f \in L_P^2(\Omega; W^*(D))$ is such that the solution u is unique and bounded in $L_P^2(\Omega; W(D))$.

Here we give two example problems that are posed in this setting:

Example 1.1 *The linear problem*

$$\begin{cases} -\nabla \cdot (a(\omega, \cdot) \nabla u(\omega, \cdot)) = f(\omega, \cdot) \text{ in } \Omega \times D, \\ u(\omega, \cdot) = 0 \text{ on } \Omega \times \partial D, \end{cases} \tag{1.2}$$

with $a(\omega, \cdot)$ uniformly bounded and coercive, i.e. there exists $a_{min}, a_{max} \in (0, +\infty)$ such that $P(\omega \in \Omega : \ a(\omega, x) \in [a_{min}, a_{max}] \, \forall x \in \overline{D}) = 1$ and $f(\omega, \cdot)$ square integrable with respect to P, satisfies assumptions A_1 and A_2 with $W(D) = H_0^1(D)$ (see [1]).

Example 1.2 *Similarly, for $k \in \mathbb{N}^+$, the nonlinear problem*

$$\begin{cases} -\nabla \cdot (a(\omega, \cdot) \nabla u(\omega, \cdot)) + u(\omega, \cdot) |u(\omega, \cdot)|^k = f(\omega, \cdot) \quad \text{in } \Omega \times D, \\ u(\omega, \cdot) = 0 \quad \text{on } \Omega \times \partial D, \end{cases} \tag{1.3}$$

with $a(\omega, \cdot)$ uniformly bounded and coercive and $f(\omega, \cdot)$ square integrable with respect to P, satisfies assumptions A_1 and A_2 with $W(D) = H_0^1(D) \cap L^{k+2}(D)$.

1.1 On Finite Dimensional Noise

In some applications, the coefficient a and the forcing term f appearing in (1.1) can be described by a random vector $[Y_1, \ldots, Y_N] : \Omega \to \mathbb{R}^N$, as in the following examples. In such cases, we will emphasize such dependence by writing a_N and f_N.

Example 1.3 (Piecewise constant random fields) *Let us consider again problem* (1.2) *where the physical domain D is the union of non-overlapping subdomains D_i, $i = 1, \ldots, N$. We consider a diffusion coefficient that is piecewise constant and random on each subdomain, i.e.*

$$a_N(\omega, x) = a_{min} + \sum_{i=1}^{N} \sigma_i \, Y_i(\omega) 1_{D_i}(x).$$

Here 1_{D_i} is the indicator function of the set D_i, σ_i, a_{min} are positive constants, and the random variables Y_i are nonnegative with unit variance.

In other applications the coefficients and forcing terms in (1.1) may have other type of spatial variation that is amenable to describe by an expansion. Depending on the decay of such expansion and the desired accuracy in our computations we may retain just the first N terms.

Example 1.4 (Karhunen-Loève expansion) *We recall that any second order random field $g(\omega, x)$, with continuous covariance function $cov[g] : \overline{D \times D} \to \mathbb{R}$, can be represented as an infinite sum of random variables, by means, for instance, of a Karhunen-Loève expansion [21]. To this end, introduce the compact and self-adjoint operator $T_g : L^2(D) \to L^2(D)$, which is defined by*

$$T_g v(\cdot) := \int_D cov[g](x, \cdot) \, v(x) \, dx \quad \forall v \in L^2(D).$$

Then, consider the sequence of non-negative decreasing eigenvalues of T_g, $\{\lambda_i\}_{i=1}^{\infty}$, and the corresponding sequence of orthonormal eigenfunctions, $\{b_i\}_{i=1}^{\infty}$, satisfying

$$T_g b_i = \lambda_i b_i, \quad (b_i, b_j)_{L^2(D)} = \delta_{ij} \ \text{ for } \ i, j \in \mathbb{N}_+.$$

In addition, define mutually uncorrelated real random variables

$$Y_i(\omega) := \frac{1}{\sqrt{\lambda_i}} \int_D (g(\omega, x) - E[g](x)) \, b_i(x) dx, \; i = 1, \ldots$$

with zero mean and unit variance, i.e. $E[Y_i] = 0$ and $E[Y_i Y_j] = \delta_{ij}$ for $i, j \in \mathbb{N}_+$. The truncated Karhunen-Loève expansion g_N, of the stochastic function g, is defined by

$$g_N(\omega, x) := E[g](x) + \sum_{i=1}^{N} \sqrt{\lambda_i} \, b_i(x) \, Y_i(\omega) \quad \forall N \in \mathbb{N}_+.$$

Then by Mercer's theorem (cf [27, p. 245]), it follows that

$$\lim_{N\to\infty}\left\{\sup_D E\left[(g-g_N)^2\right]\right\} = \lim_{N\to\infty}\left\{\sup_D\left(\sum_{i=N+1}^{\infty}\lambda_i b_i^2\right)\right\} = 0.$$

Observe that the N random variables in (1.4)*, describing the random data, are then weighted differently due to the decay of the eigen-pairs of the Karhunen-Loève expansion. The decay of eigenvalues and eigenvectors has been investigated e.g. in the works [12] and [29].*

The above examples motivate us to consider problems whose coefficients are described by finitely many random variables. Thus, we will seek a random field $u_N : \Omega \times \overline{D} \to \mathbb{R}$, such that a.s., the following equation holds:

$$\mathcal{L}(a_N)(u_N) = f_N \quad \text{in } D, \tag{1.4}$$

We assume that equation (1.4) admits a unique solution $u_N \in L^2_P(\Omega; W(D))$. We then have, by the Doob–Dynkin's lemma (cf. [26]), that the solution u_N of the stochastic elliptic boundary value problem (1.4) can be described by $u_N = u_N(\omega, x) = u_N(Y_1(\omega), \ldots, Y_N(\omega), x)$. We underline that the coefficients a_N and f_N in (1.4) may be an exact representation of the input data as in Example 1.3 or a suitable truncation of the input data as in Example 1.4. In the latter case, the solution u_N will also be an approximation of the exact solution u in (1.1) and the truncation error $u - u_N$ has to be properly estimated, see for instance [25, Section 4.2].

Remark 1.5 (Nonlinear coefficients) *In certain cases, one may need to ensure qualitative properties on the coefficients a_N and f_N and may be worth while to describe them as nonlinear functions of Y. For instance, in Example 1.1 one is required to enforce positiveness on the coefficient $a_N(\omega, x)$, say $a_N(\omega, x) \geq a_{min}$ for all $x \in D$, a.s. in Ω. Then a better choice is to expand $\log(a_N - a_{min})$. The following standard transformation guarantees that the diffusivity coefficient is bounded away from zero almost surely*

$$\log(a_N - a_{min})(\omega, x) = b_0(x) + \sum_{1\leq n\leq N}\sqrt{\lambda_n}b_n(x)Y_n(\omega), \tag{1.5}$$

i.e. one performs a Karhunen-Loève expansion for $\log(a_N - a_{min})$, assuming that $a_N > a_{min}$ almost surely. On the other hand, the right hand side of (1.4) *can be represented as a truncated Karhunen-Loève expansion*

$$f_N(\omega, x) = c_0(x) + \sum_{1\leq n\leq N}\sqrt{\mu_n}c_n(x)Y_n(\omega).$$

Remark 1.6 *It is usual to have f_N and a_N independent, because the forcing terms and the parameters in the operator $\mathcal{L}$ are seldom related. In such a situation we have $a_N(Y(\omega), x) = a_N(Y_a(\omega), x)$ and $f_N(Y(\omega), x) = f_N(Y_f(\omega), x)$, with $Y = [Y_a, Y_f]$ and the vectors Y_a, Y_f are independent.*

For this work we denote $\Gamma_n \equiv Y_n(\Omega)$ the image of Y_n, where we assume $Y_n(\omega)$ to be bounded. Without loss of generality we can assume $\Gamma_n = [-1, 1]$. We also let $\Gamma^N = \prod_{n=1}^N \Gamma_n$ and assume that the random variables $[Y_1, Y_2, \ldots, Y_n]$ have a joint probability density function $\rho : \Gamma^N \to \mathbb{R}_+$, with $\rho \in L^\infty(\Gamma^N)$. Thus, the goal is to approximate the function $u_N = u_N(y, x)$, for any $y \in \Gamma^N$ and $x \in \overline{D}$. (see [1], [2])

Remark 1.7 (Unbounded Random Variables) *By using a similar approach to the work [1] we can easily deal with unbounded random variables, such as Gaussian or exponential ones. For the sake of simplicity in the presentation we focus our study on bounded random variables only.*

1.2 Regularity

Before discussing various collocation techniques and going through the convergence analysis of such methods, we need to state some regularity assumptions on the data of the problem and consequent regularity results for the exact solution u_N. We will perform a one-dimensional analysis in each direction $y_n,\ n = 1, \ldots, N$. For this, we introduce the following notation: $\Gamma_n^* = \prod_{\substack{j=1 \\ j \neq n}}^N \Gamma_j$, $\mathbf{y}_n^*$ will denote an arbitrary element of Γ_n^*. We require the solution to problem (1.1) to satisfy the following estimate:

Assumption 1.8 *For each $y_n \in \Gamma_n$, there exists $\tau_n > 0$ such that the function $u_N(y_n, \mathbf{y}_n^*, x)$ as a function of y_n, $u_N : \Gamma_n \to C^0(\Gamma_n^*; W(D))$ admits an analytic extension $u(z, \mathbf{y}_n^*, x)$, $z \in \mathbb{C}$, in the region of the complex plane*

$$\Sigma(\Gamma_n; \tau_n) \equiv \{z \in \mathbb{C}, \quad dist(z, \Gamma_n) \leq \tau_n\}. \tag{1.6}$$

Moreover, $\forall z \in \Sigma(\Gamma_n; \tau_n)$,

$$\|u_N(z)\|_{C^0(\Gamma_n^*; W(D))} \leq \lambda \tag{1.7}$$

with λ a constant independent of n.

The previous assumption should be verified for each particular application. In particular, this has implications on the allowed regularity of the input data, e.g. coefficients, loads, etc., with respect to $y_n, n = 1, 2, \ldots, N$, of the PDE under study. In the next Section we recall some theoretical results, which were proved in [1, Section 3], for the linear problem introduced in Example 1.1.

2 Applications to linear elliptic PDEs with random input data

In this Section we give more details concerning the linear problem described in Example 1.1. Problem (1.2) can be written in a weak form as: find $u \in$

$L^2_P(\Omega; H^1_0(D))$ such that

$$\int_D E[a\nabla u \cdot \nabla v]\, dx = \int_D E[fv]\, dx \quad \forall\, v \in L^2_P(\Omega; H^1_0(D)). \tag{2.1}$$

A straightforward application of the Lax-Milgram theorem allows one to state the well posedness of problem (2.1). Moreover, the following a priori estimates hold

$$\|u\|_{H^1_0(D)} \leq \frac{C_P}{a_{min}} \|f(\omega, \cdot)\|_{L^2(D)} \quad \text{a.s.} \tag{2.2}$$

and

$$\|u\|_{L^2_P(\Omega; H^1_0(D))} \leq \frac{C_P}{a_{min}} \left(\int_D E[f^2]\, dx \right)^{1/2}, \tag{2.3}$$

where C_P denotes the constant appearing in the Poincaré inequality:

$$\|w\|_{L^2(D)} \leq C_P \|\nabla w\|_{L^2(D)} \qquad \forall w \in H^1_0(D).$$

Once we have the input random fields described by a finite set of random variables, i.e. $a(\omega, x) = a_N(Y_1(\omega), \ldots, Y_N(\omega), x)$, and similarly for $f(\omega, x)$, the "finite dimensional" version of the stochastic variational formulation (2.1) has a "deterministic" equivalent which is the following: find $u_N \in L^2_\rho(\Gamma^N; H^1_0(D))$ such that

$$\int_{\Gamma^N} \rho\, (a_N \nabla u_N, \nabla v)_{L^2(D)}\, dy = \int_{\Gamma^N} \rho\, (f_N, v)_{L^2(D)}\, dy, \quad \forall\, v \in L^2_\rho(\Gamma^N; H^1_0(D)). \tag{2.4}$$

Observe that in this work the gradient notation, ∇, always means differentiation with respect to $x \in D$ only, unless otherwise stated. The stochastic boundary value problem (2.1) now becomes a deterministic Dirichlet boundary value problem for an elliptic partial differential equation with an $N-$dimensional parameter. For convenience, we consider the solution u_N as a function $u_N : \Gamma^N \to H^1_0(D)$ and we use the notation $u_N(y)$ whenever we want to highlight the dependence on the parameter y. We use similar notations for the coefficient a_N and the forcing term f_N. Then, it can be shown that problem (2.1) is equivalent to

$$\int_D a_N(y) \nabla u_N(y) \cdot \nabla \phi\, dx = \int_D f_N(y) \phi\, dx, \quad \forall \phi \in H^1_0(D), \ \ \rho\text{-a.e. in } \Gamma^N. \tag{2.5}$$

For our convenience, we will suppose that the coefficient a_N and the forcing term f_N admit a smooth extension on the ρ-zero measure sets. Then, equation (2.5) can be extended a.e. in Γ^N with respect to the Lebesgue measure (instead of the measure ρdy).

It has been proved in [1] that problem (2.5) satisfies the analyticity result stated in Assumption 1.8. For instance, if we take the diffusivity coefficient as

in Example 1.3 and a deterministic load we can take the size of the analyticity region as

$$\tau_n = \frac{a_{min}}{4\sigma_n}. \tag{2.6}$$

On the other hand, if we take the diffusivity coefficient as a truncated expansion like in Remark 1.5, then a suitable analyticity region $\Sigma(\Gamma_n; \tau_n)$ is given by

$$\tau_n = \frac{1}{4\sqrt{\lambda_n}\|b_n\|_{L^\infty(D)}}. \tag{2.7}$$

Observe that, in the latter case, as $\sqrt{\lambda_n}\|b_n\|_{L^\infty(D)} \to 0$ for a regular enough covariance function (see [12]) the analyticity region increases as n increases. This fact introduces, naturally, an anisotropic behavior with respect to the "direction" n. This effect will be exploited in the numerical methods proposed in the next sections.

3 Collocation techniques

We seek a numerical approximation to the exact solution of (1.4) in a suitable finite dimensional subspace. To describe such a subspace properly, we introduce some standard approximation subspaces, namely:

- $W_h(D) \subset W(D)$ is a standard finite element space of dimension N_h, which contains continuous piecewise polynomials defined on regular triangulations $\mathcal{T}_h$ that have a maximum mesh-spacing parameter $h > 0$. We suppose that W_h has the following deterministic approximability property: for a given function $\varphi \in W(D)$,

$$\min_{v \in W_h(D)} \|\varphi - v\|_{W(D)} \leq C(s;\varphi)\, h^s, \tag{3.1}$$

where s is a positive integer determined by the smoothness of φ and the degree of the approximating finite element subspace and $C(s;\varphi)$ is independent of h.

Example 3.1 *Let D be a convex polygonal domain and $W(D) = H_0^1(D)$. For piecewise linear finite element subspaces we have*

$$\min_{v \in W_h(D)} \|\varphi - v\|_{H_0^1(D)} \leq c\, h\, \|\varphi\|_{H^2(D)}.$$

That is, $s = 1$ and $C(s;\varphi) = \|\varphi\|_{H^2(D)}$, see for example [5].

We will also assume that there exists a finite element operator $\pi_h : W(D) \to W_h(D)$ with the optimality condition

$$\|\varphi - \pi_h\varphi\|_{W(D)} \leq C_\pi \min_{v \in W_h(D)} \|\varphi - v\|_{W(D)}, \quad \forall \varphi \in W(D), \tag{3.2}$$

where the constant C_π is independent of the mesh size h.

- $\mathcal{P}_{\mathbf{p}}(\Gamma^N) \subset L^2_\rho(\Gamma^N)$ is the span of tensor product polynomials with degree at most $\mathbf{p} = (p_1, \ldots, p_N)$ i.e. $\mathcal{P}_{\mathbf{p}}(\Gamma^N) = \bigotimes_{n=1}^N \mathcal{P}_{p_n}(\Gamma_n)$, with

$$\mathcal{P}_{p_n}(\Gamma_n) = \text{span}(y_n^k,\, k = 0, \ldots, p_n), \quad n = 1, \ldots, N.$$

  Hence the dimension of $\mathcal{P}_{\mathbf{p}}(\Gamma^N)$ is $N_p = \prod_{n=1}^N (p_n + 1)$.

Stochastic collocation entails the sampling of approximate values $\pi_h u_N(y_k) = u_h^N(y_k) \in W_h(D)$, to the solution u_N of (1.4) on a suitable set of abscissas $y_k \in \Gamma^N$.

Example 3.2 *If we examine the linear PDE for example, then we introduce the semi-discrete approximation $u_h^N : \Gamma^N \to W_h(D)$, obtained by projecting equation* (2.5) *onto the subspace $W_h(D)$, for each $y \in \Gamma^N$, i.e.*

$$\int_D a_N(y) \nabla u_h^N(y) \cdot \nabla \phi_h \, dx = \int_D f_N(y) \phi_h \, dx, \quad \forall \phi_h \in W_h(D), \;\; \textit{for a.e. } y \in \Gamma^N. \tag{3.3}$$

Notice that the finite element functions $u_h^N(y)$ satisfy the optimality condition (3.2)*, for all $y \in \Gamma^N$.*

Then the construction of a fully discrete approximation, $u_{h,\mathbf{p}}^N \in C^0(\Gamma^N; W_h(D))$, is based on a suitable interpolation of the sampled values. That is

$$u_{h,\mathbf{p}}^N(y, \cdot) = \sum_k u_h^N(y_k, \cdot) l_k^{\mathbf{p}}(y), \tag{3.4}$$

where, for instance, the functions $l_k^{\mathbf{p}}$ can be taken as the Lagrange polynomials (see Section 3.1, 3.2 and 3.3).

This formulation can be used to compute the mean value or variance of u, as:

$$E[u] \approx \overline{u}_h^N \equiv \sum_k u_h^N(y_k, \cdot) \int_{\Gamma^N} l_k^{\mathbf{p}}(y) \rho(y) dy$$

and

$$\text{Var}[u] \approx \sum_k \left(u_h^N(y_k, \cdot)\right)^2 \int_{\Gamma^N} l_k^{\mathbf{p}}(y) \rho(y) dy - \left(\overline{u}_h^N\right)^2.$$

Several choices are possible for the interpolation points. We will discuss two of them, namely Clenshaw-Curtis and Gaussian in Sections 3.3.1 and 3.3.2 respectively. See the work [30] for an insightful comparison between these two choices. Regardless of the choice of interpolating knots, the interpolation can be constructed by using either full tensor product polynomials, see Section 3.1, or the space of sparse polynomials, see Sections 3.2 and 3.3.

3.1 Full tensor product interpolation

In this Section we briefly recall interpolation based on Lagrange polynomials. Let $i \in \mathbb{N}_+$ and $\{y_1^i, \ldots, y_{m_i}^i\} \subset [-1,1]$ be a sequence of abscissas for Lagrange interpolation on $[-1,1]$.

For $u \in C^0(\Gamma^1; W(D))$ and $N = 1$ we introduce a sequence of one-dimensional Lagrange interpolation operators $\mathscr{U}^i : C^0(\Gamma^1; W(D)) \to V_{m_i}(\Gamma^1; W(D))$

$$\mathscr{U}^i(u)(y) = \sum_{j=1}^{m_i} u(y_j^i) \cdot l_j^i(y), \quad \forall u \in C^0(\Gamma^1; W(D)), \tag{3.5}$$

where $l_j^i \in \mathcal{P}_{m_i-1}(\Gamma^1)$ are Lagrange polynomials of degree $p_i = m_i - 1$ and

$$V_m(\Gamma^1; W(D)) = \left\{ v \in C^0(\Gamma^1; W(D)) \,:\, v(y,x) = \sum_{k=1}^{m} \widetilde{v}_k(x) l_k(y),\ \{\widetilde{v}_k\}_{k=1}^m \in W(D) \right\}.$$

Here of course we have, for $i \in \mathbb{N}_+$,

$$l_j^i(y) = \prod_{\substack{k=1 \\ k \neq j}}^{m_i} \frac{(y - y_k^i)}{(y_j^i - y_k^i)}$$

and formula (3.5) reproduces exactly all polynomials of degree less than m_i. Now, in the multivariate case $N > 1$, for each $u \in C^0(\Gamma^N; W(D))$ and the multi-index $\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{N}_+^N$ we define the full tensor product interpolation formulas

$$\mathscr{I}_{\mathbf{i}}^N u(y) := \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N}\right)(u)(y) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_N=1}^{m_{i_N}} u\left(y_{j_1}^{i_1}, \ldots, y_{j_N}^{i_N}\right) \cdot \left(l_{j_1}^{i_1} \otimes \cdots \otimes l_{j_N}^{i_N}\right). \tag{3.6}$$

Clearly, the above product needs $(m_{i_1} \cdots m_{i_N})$ function values, sampled on a grid. These formulas will also be used as the building blocks for the Smolyak method, described next.

3.2 The isotropic Smolyak method

Here we follow closely the work [4] and describe the Smolyak *isotropic* formulas $\mathscr{A}(w, N)$. The Smolyak formulas are just linear combinations of product formulas (3.6) with the following key properties: only products with a relatively small number of knots are used and the linear combination is chosen in such a way that an interpolation property for $N = 1$ is preserved for $N > 1$. With $\mathscr{U}^0 = 0$ and for $i \in \mathbb{N}_+$ define

$$\Delta^i := \mathscr{U}^i - \mathscr{U}^{i-1}. \tag{3.7}$$

Moreover, for integers $w \in \mathbb{N}$, we define the sets

$$X(w,N) := \left\{ \mathbf{i} \in \mathbb{N}_+^N, \mathbf{i} \geq \mathbf{1} \,:\, \sum_{n=1}^{N} (i_n - 1) \leq w \right\}$$

and

$$Y(w,N) := \left\{ \mathbf{i} \in \mathbb{N}_+^N, \mathbf{i} \geq \mathbf{1} \,:\, w - N + 1 \leq \sum_{n=1}^{N} (i_n - 1) \leq w \right\}$$

and for $\mathbf{i} \in X(w,N)$ or $\mathbf{i} \in Y(w,N)$ we put $|\mathbf{i}| = i_1 + \cdots + i_N$. Then the isotropic Smolyak algorithm is given by

$$\mathscr{A}(w,N) = \sum_{\mathbf{i} \in X(w,N)} \left(\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N} \right). \tag{3.8}$$

Equivalently, formula (3.8) can be written as (see [31])

$$\mathscr{A}(w,N) = \sum_{\mathbf{i} \in Y(w,N)} (-1)^{w+N-|\mathbf{i}|} \binom{N-1}{w+N-|\mathbf{i}|} \cdot \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N} \right). \tag{3.9}$$

To compute $\mathscr{A}(w,N)(u)$, one only needs to know function values on the "sparse grid"

$$\mathscr{H}(w,N) = \bigcup_{\mathbf{i} \in Y(w,N)} \left(\vartheta^{i_1} \times \cdots \times \vartheta^{i_N} \right) \tag{3.10}$$

where $\vartheta^i = \left\{ y_1^i, \ldots, y_{m_i}^i \right\} \subset [-1,1]$ denotes the set of points used by $\mathscr{U}^i$. Note that the Smolyak algorithm, as presented in this Section, is isotropic, since all directions are treated equally. This can be seen from (3.8) where the multi-index $\mathbf{i} \in \mathbb{N}_+^N$ that determines the number of sample points in each dimension, i_n, $n = 1, 2, \ldots, N$ is sampled from the set $X(w,N)$. This ensures that if $(i_1, i_2, \ldots, i_N)$ is a valid index, then any permutation of it is also a valid index.

Examples of isotropic sparse grids, for $N = 2$, constructed from the nested Clenshaw-Curtis abscissas, described in Section 3.3.1, and the non-nested Gaussian abscissas, described in Section 3.3.2, are shown in Figure 1. To see the reduction in function evaluations we also include a plot of the corresponding full tensor product grid computed from the Clenshaw-Curtis abscissas. We note that since the Gaussian abscissas and the Clenshaw-Curtis abscissas tend to accumulate towards the boundary, plots of full tensor product grids will look superficially the same.

We will next discuss improvements that can be made to further reduce the number of points used to compute $\mathscr{U}^i$ by considering an anisotropic version of the method.

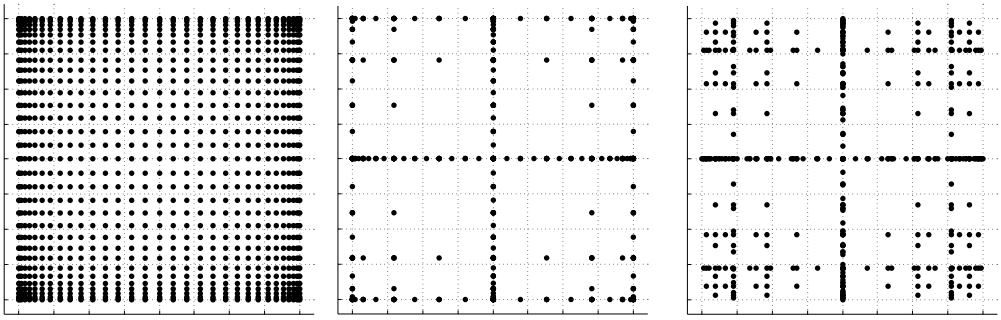

Figure 1: For a finite dimensional Γ^N with $N = 2$ we plot the full tensor product grid using the Clenshaw-Curtis abscissas and isotropic Smolyak sparse grids utilizing the Clenshaw-Curtis abscissas and the Gaussian abscissas for $w = 5$.

3.3 The anisotropic Smolyak method

It was shown in [25] that the conventional Smolyak construction is very effective for problems whose input data depend on a moderate number of random variables, which "weigh equally" in the solution. Upon the assumption that the solution is analytic with respect to each stochastic direction we show that for such isotropic situations the displayed convergence is faster than standard collocation techniques built upon full tensor product spaces.

On the other hand, the convergence rate deteriorates when we attempt to solve highly anisotropic problems, such as those appearing when the input random variables come e.g., from Karhunen-Loève -type truncations of "smooth" random fields described by Example 1.4. In this particular case we can rely on a priori information related to the decay of the noise coefficients to develop an anisotropic Smolyak algorithm.

To generalize the traditional sparse grid method with respect to different stochastic dimensions we propose an algorithm that considers a different index set rather than the unit simplex $|\mathbf{i}| \leq w + N$. Similar to [14, 16] we let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \mathbb{R}^N_+$ be N-dimensional weight vector for the different stochastic dimensions. Furthermore, we define $\underline{\alpha} := \min_{1 \leq n \leq N} \alpha_n$ and consider the follow general class of simplicies defined by the index set

$$X_{\boldsymbol{\alpha}}(w, N) = \left\{ \mathbf{i} \in \mathbb{N}^N_+, \mathbf{i} \geq \mathbf{1} \, : \, \sum_{n=1}^{N} (i_n - 1)\alpha_n \leq w\underline{\alpha} \right\}.$$

The strategy to this approach relies on constructing the weight vector $\boldsymbol{\alpha} \in \mathbb{R}^N_+$ from either *a priori* knowledge or *a posteriori* information. See Section 4 respectively for a more detailed description.

Then, using (3.7) we described the anisotropic Smolyak formulas, $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$,

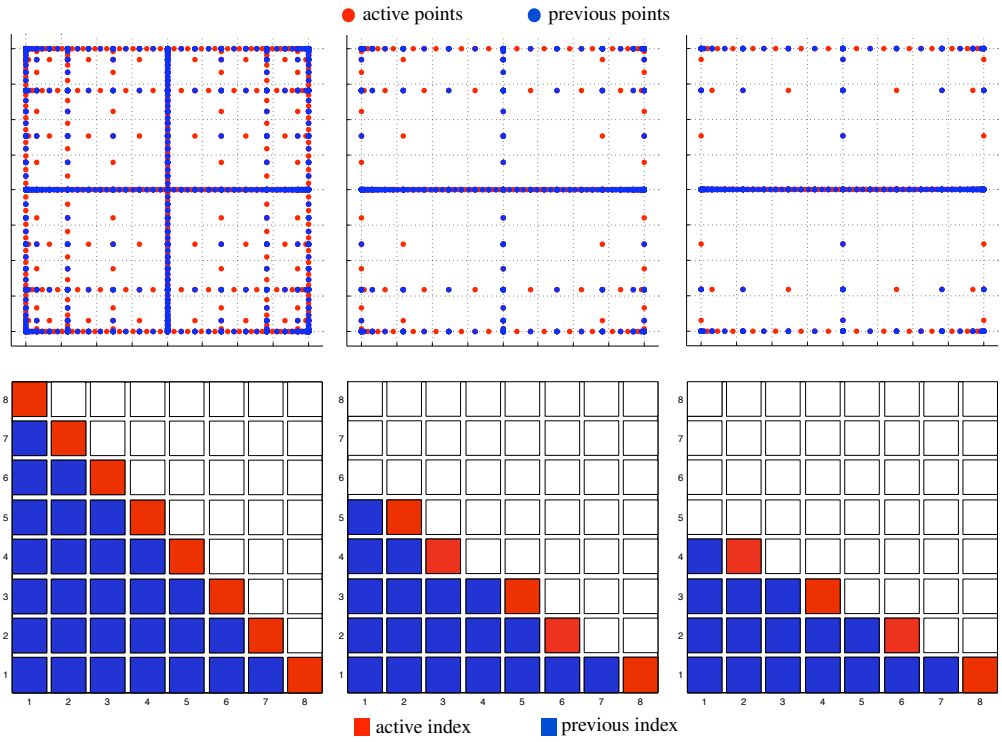

Figure 2: For a finite dimensional Γ^N with $N = 2$ and $w = 7$ we plot: on the top, the isotropic Smolyak grid and the anisotropic Smolyak grids with $\alpha_2/\alpha_1 = 3/2$ and $\alpha_2/\alpha_1 = 2$, utilizing the Clenshaw-Curtis abscissas and the bottom, the corresponding indices $(i_1, i_2) \in X_{\boldsymbol{\alpha}}(7, 2)$.

given by

$$\mathscr{A}_{\boldsymbol{\alpha}}(w, N) = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \left(\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N}\right) \tag{3.11}$$

for $w \in \mathbb{N}$. Equivalently, (3.11) can be written as

$$\mathscr{A}_{\boldsymbol{\alpha}}(w, N) = \sum_{\mathbf{i} \in Y_{\boldsymbol{\alpha}}(w,N)} c_{\boldsymbol{\alpha}}(\mathbf{i}) \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N}\right) \tag{3.12}$$

with

$$c_{\boldsymbol{\alpha}}(\mathbf{i}) := \sum_{\substack{\mathbf{j} \in \{0,1\}^N \\ \mathbf{i}+\mathbf{j} \in X_{\boldsymbol{\alpha}}(w,N)}} (-1)^{|\mathbf{j}|}$$

and

$$Y_{\boldsymbol{\alpha}}(w, N) := X_{\boldsymbol{\alpha}}(w, N) \setminus X_{\boldsymbol{\alpha}}\left(w - \frac{|\boldsymbol{\alpha}|}{\underline{\alpha}}, N\right).$$

Similarly to the isotropic case, to compute $\mathscr{A}_{\boldsymbol{\alpha}}(w,N)(u)$, one only needs to know function values on the "sparse grid"

$$\mathscr{H}_{\boldsymbol{\alpha}}(w,N) = \bigcup_{\mathbf{i}\in Y_{\boldsymbol{\alpha}}(w,N)} \left(\vartheta^{i_1}\times\cdots\times\vartheta^{i_N}\right). \tag{3.13}$$

We note that the isotropic Smolyak method presented in Section 3.2 is a special case of the anisotropic algorithm. This can be observed by simply taking the components of weight vector to be equal, i.e. $\alpha_1 = \alpha_2 = \cdots = \alpha_N$. In this case formula (3.11) is equivalent to (3.8) and it can be easily shown that (3.12) reduces to (3.9). Finally, in Figure 2, for $N = 2$ and $w = 7$, we show anisotropic Smolyak sparse grids utilizing the Clenshaw-Curtis points described in Section 3.3.1, corresponding to the anisotropy ratio $\alpha_2/\alpha_1 = 1$, $\alpha_2/\alpha_1 = 3/2$ and $\alpha_2/\alpha_1 = 2$ respectively. We also show the indices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(7,2)$ that were used to construct the anisotropic sparse interpolant $\mathscr{A}_{\boldsymbol{\alpha}}(7,2)$. In accordance with the nested structure of the Clenshaw-Curtis abscissas, in Figure 2 we point out the difference between active points/indices and previously computed points/indices.

3.3.1 Clenshaw-Curtis Formulas

We first suggest to use the Smolyak algorithm based on polynomial interpolation at the extrema of Chebyshev polynomials. For any choice of $m_i > 1$ these knots are given by

$$y_j^i = -\cos\left(\frac{\pi(j-1)}{m_i-1}\right), \quad j = 1,\ldots,m_i. \tag{3.14}$$

In addition, we define $y_1^i = 0$ if $m_i = 1$. It remains to specify the numbers m_i of knots that are used in formulas $\mathscr{U}^i$. In order to obtain nested sets of points, i.e., $\vartheta^i \subset \vartheta^{i+1}$ and thereby $\mathscr{H}_{\mathscr{A}}(w,N) \subset \mathscr{H}_{\mathscr{A}}(w+1,N)$, we choose

$$m_1 = 1 \text{ and } m_i = 2^{i-1}+1, \text{ for } i > 1. \tag{3.15}$$

For such a choice of m_i we arrive at Clenshaw-Curtis formulas, see [8]. It is important to choose $m_1 = 1$ if we are interested in optimal approximation in relatively large N, because in all other cases the number of points used by $\mathscr{A}(w,N)$ and $\mathscr{A}_{\boldsymbol{\alpha}}(w,N)$ increases too fast with N.

A variant of the Clenshaw-Curtis formulas are the Filippi formulas in which the abscissas at the boundary of the interval are omitted [15]. In this case only the smaller degree $m_i - 1$ of exactness is obtained.

3.3.2 Gaussian formulas

We also propose to apply the Smolyak formulas based on polynomial interpolation at the zeros of the orthogonal polynomials with respect to a weight ρ. This naturally leads to the Gauss formulas that have a maximum degree of exactness

of $2m_i - 1$. However, these Gauss-Legendre formulas are in general not nested. Regardless, as in the Clenshaw-Curtis case, we choose

$$m_1 = 1 \text{ and } m_i = 2^{i-1} + 1, \text{ for } i > 1.$$

The natural choice of the weight ρ should be the probability density of the random variables $Y_i(\omega)$ for all i. Yet, in the general multivariate case, if the random variables Y_i are not independent, the density ρ does not factorize, i.e.

$$\rho(y_1, \ldots, y_n) \neq \prod_{n=1}^{N} \rho_n(y_n).$$

To this end, we first introduce an auxiliary probability density function $\hat{\rho} : \Gamma^N \to \mathbb{R}^+$ that can be seen as the joint probability of N independent random variables, i.e. it factorizes as

$$\hat{\rho}(y_1, \ldots, y_n) = \prod_{n=1}^{N} \hat{\rho}_n(y_n), \quad \forall y \in \Gamma^N, \qquad \text{and is such that} \quad \left\| \frac{\rho}{\hat{\rho}} \right\|_{L^\infty(\Gamma^N)} < \infty. \tag{3.16}$$

For each dimension $n = 1, \ldots, N$ let the m_n Gaussian abscissas be the roots of the m_n degree polynomial that is $\hat{\rho}_n$-orthogonal to all polynomials of degree $m_n - 1$ on the interval $[-1, 1]$.

4 Selection of the α weights for Anisotropic Smolyak

The ability to evaluate the stochastic dimensions differently is a necessity since many practical problems exhibit highly anisotropic behavior. The rationale behind our anisotropic sparse grid approach is based on an examination of the total error

$$\varepsilon = \| u_N - \mathcal{I}^N_{\mathbf{p}} u_N \|_{L^2_{\hat{\rho}}(\Gamma^N; W(D))}, \tag{4.1}$$

produced by anisotropic full tensor product polynomial interpolation on Gaussian abscissas.

Each stochastic dimension contributes to this total error. When the total error is divided equally among the random variables our earlier work [25] revealed that the isotropic Smolyak method, described in Section 3.2, displays a fast convergence rate and is very effective in considerably reducing the *curse of dimensionality*. We are also able to treat effectively problems that depend on a moderately large number of random variables, while keeping a high level of accuracy. When the error is dominated by certain directions we utilize the anisotropic Smolyak algorithm, described in Section 3.3 which combines an optimal treatment of the anisotropy of the problem while minimizing function evaluations via the use of sparse grids. Similar to an adaptive interpolation method, we place more points in the directions with the largest contribution to the total error.

The main idea is to link the α_n coefficients with the rate of exponential convergence in the corresponding direction, which for functions that satisfy Assumption 1.8 is described by the following Lemma, whose proof can be found in [1, Lemma 7] and which is an immediate extension of the result given in [9, Chapter 7, Section 8]:

Lemma 4.1 *Given a function $v \in C^0(\Gamma^1; W(D))$ which admits an analytic extension in the region of the complex plane $\Sigma(\Gamma^1;\tau) = \{z \in \mathbb{C}, \ \ dist(z,\Gamma^1) \leq \tau\}$ for some $\tau > 0$, there holds*

$$E_{m_i} \equiv \min_{w \in V_{m_i}} \|v - w\|_{C^0(\Gamma^1;W(D))} \leq \frac{2}{\varrho - 1} e^{-m_i \log(\varrho)} \max_{z \in \Sigma(\Gamma^1;\tau)} \|v(z)\|_{W(D)}$$

where

$$1 < \varrho = \frac{2\tau}{|\Gamma^1|} + \sqrt{1 + \frac{4\tau^2}{|\Gamma^1|^2}}. \tag{4.2}$$

Remark 4.2 (Approximation with unbounded random variables) *A related result with*
weighted norms holds for unbounded random variables whose probability density decays as the Gaussian density at infinity (see [1]).

In the multidimensional case, the value of τ_n will depend, in general, on the direction n, cf. (2.7) and (2.6). As a consequence of this variation and (4.2) the decay coefficients ϱ_n will also depend on the direction, $n = 1, \ldots, N$. We now assume that we know positive numbers $0 < g(n)$, $n = 1, \ldots, N$, such that

$$\varrho_n \geq e^{g(n)}. \tag{4.3}$$

Then, we will choose the anisotropic Smolyak weights as

$$\alpha_n = g(n) \quad \text{for all } n = 1, 2, \ldots, N, \tag{4.4}$$

On what follows we will use the notations

$$\underline{\alpha} = \underline{g} = \min_{1 \leq n \leq N} \{g(n)\} \quad \text{and} \quad \mathscr{G}(N) = \sum_{n=1}^{N} g(n). \tag{4.5}$$

Observe that we have now transformed the problem of choosing $\boldsymbol{\alpha}$ into the one of estimating the decay coefficients $\mathbf{g} = (g(1), \ldots, g(N))$.

Remark 4.3 (Optimality of $\boldsymbol{\alpha}$ choice) *As will be seen in Remark 5.12 the choice $\boldsymbol{\alpha} = \mathbf{g}$ is optimal with respect to the error bound derived in Theorem 5.7.*

In principle, it is possible to consider two kinds of estimation strategies. The first uses *a priori* knowledge, while in the second approach we use *a posteriori* information from computations, i.e. we fit the values of $\mathbf{g}$. The remainder of the

Section will explain the choice of $\boldsymbol{\alpha} \in \mathbb{R}_+^N$, for the construction of set of indices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$, using these procedures.

A priori selection
The estimation of the $\boldsymbol{\alpha} = \mathbf{g} \in \mathbb{R}_+^N$ coefficients can be achieved using *a priori* information, i.e. by estimating first suitable values for τ_n, then using (4.2), and taking $g(n)$ as in (4.3), namely

$$g(n) = \log\left(\frac{2\tau_n}{|\Gamma_n|} + \sqrt{1 + \frac{4{\tau_n}^2}{|\Gamma_n|^2}}\right),$$

for $n = 1, 2, \ldots, N$. In the numerical examples presented in Section 6 we have used the simpler and more conservative relation

$$g(n) = \log\left(1 + \frac{2\tau_n}{|\Gamma_n|}\right). \tag{4.6}$$

On the other hand, a priori estimates for τ_n can be derived for certain classes of problems, see for instance the linear problem (2.1) and the estimates (2.7) and (2.6). In Section 6 we implement numerically this *a priori* choice of $\boldsymbol{\alpha} = \mathbf{g} \in \mathbb{R}_+^N$ for the construction of the general simplices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$ when solving problem (1.2).

In some practical applications it may not be possible to sharply estimate the vector $\mathbf{g}$ using *a priori* information. Therefore, in the next Section we propose a computational alternative.

A posteriori selection
In order to explain the *a posteriori* estimation of $\boldsymbol{\alpha} = \mathbf{g} \in \mathbb{R}_+^N$ for the construction of the general simplices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$ we will describe particular cases with $N = 11$ and focus on both an isotropic and highly anisotropic version the example described in Section 6, corresponding to $L_c = 1/64$ and $L_c = 1/2$ respectively.

Using Lemma 4.1, we expect an error decay in the direction n of the form

$$\varepsilon_n \approx d_n {\varrho_n}^{-p_n}, \text{ for all } n = 1, 2, \ldots, N, \tag{4.7}$$

where p_n is the number of collocation points in the direction n. In order to compute the weight vector $\mathbf{g} = \boldsymbol{\alpha} \in \mathbb{R}_+^N$, $\mathbf{g} \approx \log(\boldsymbol{\varrho})$, we first observe from (4.7) that

$$\log_{10}(\varepsilon_n) \approx \log_{10}(d_n) - p_n \log_{10}(\varrho_n) \approx \log_{10}(d_n) - p_n \log_{10}(e) g(n).$$

Therefore, if we plot $\log_{10}(\varepsilon_n)$ versus p_n for each stochastic dimension $n = 1, 2, \ldots, N$ and use a linear least squares approximation to fit the data, the slope of each line will give an estimate of $g(n)$. When solving problem (6.1), for the cases $L_c = 1/2$ and $1/64$, we utilize the same finite element space as in Section 6

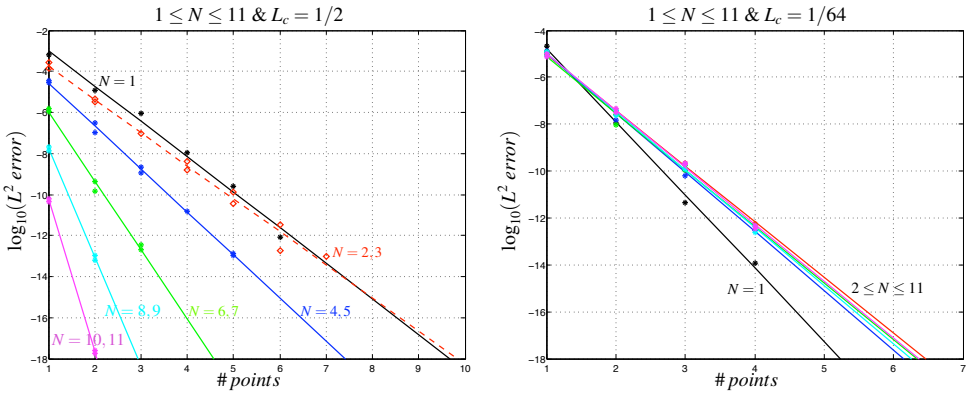

Figure 3: A linear least square approximation to fit $\log_{10}(\|E[\varepsilon_n]\|_{L^2(D)})$ versus p_n with ε_n defined by (4.7). For $n = 1, 2, \ldots, N = 11$ we plot: on the left, the highly anisotropic case $L_c = 1/2$ and on the right, the isotropic case $L_c = 1/64$.

and we approximate $\mathbb{E}[\varepsilon_n]$ in the n-th direction, corresponding to a multi index $\mathbf{p} = (1, 1, \ldots, p_n, 1, \ldots, 1)$ by $\mathbb{E}[\varepsilon_n] \approx \mathbb{E}[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]$, with $\widetilde{\mathbf{p}} = (1, 1, \ldots, p_n + 1, 1, \ldots, 1)$. The computational results for the $L^2(D)$ approximation error in the expected value, $\mathbb{E}[\varepsilon_n]$, are shown on Figure 3.

	$g(1)$	$g(2), g(3)$	$g(4), g(5)$	$g(6), g(7)$	$g(8), g(9)$	$g(10), g(11)$
$L_c = 1/2$	1.7	1.6	2.1	3.3	5.3	7.4
$L_c = 1/64$	3.1	2.4	2.5	2.4	2.5	2.4

Table 1: The $N = 11$ values of the function $g(n) = \alpha_n$ constructed from *a posteriori* information. The values $g(n)$, $n = 1, 2, \ldots, N$ are the slopes of N linear least squares fits to the error of an univariate anisotropic method when solving problem (6.1) with correlation lengths $L_c = 1/2$ and $1/64$.

The results for $g(n)$, $n = 1, 2, \ldots, N = 11$ can be seen in Table 1. Table 1 reveals that the a posteriori selection of $\boldsymbol{\alpha} \in \mathbb{R}^N_+$ performs well at dictating the behavior of problem (6.1) for the cases $L_c = 1/2$ and $1/64$. In the former case the vector $\boldsymbol{\alpha}$ weighs heavily in the higher dimensions as opposed to the latter that approximately weighs equally in all directions. Also, both cases $L_c = 1/2$ and $1/64$, reveal that $\underline{\alpha} = \alpha_2 = \alpha_3$.

Observe that in general the rate ϱ_n also depends on $\mathbf{y}^*_n \in \Gamma^*_n$, i.e. on the values of the other random variables. Our way to estimate the decay coefficient $g(n)$ is not conservative since we only estimate ϱ_n at the single point $\mathbb{E}[\mathbf{y}^*_n]$. A more conservative estimate will imply estimating the worst value of $\mathbf{y}^*_n$, i.e. the one that minimizes ϱ_n. This may be critical for nearly singular cases.

Remark 4.4 (Applications to piecewise constant random fields) *We also comment that the a posteriori selection of $\boldsymbol{\alpha} \in \mathbb{R}^N_+$, described above, is not only restricted to random fields related to the Karhunen-Loève expansion described in Example 1.4 but also can be easily applied in other cases, for instance to the piecewise constant random fields described in Example 1.3.*

5 Error Analysis

Collocation methods can be used to approximate the solution $u_N \in C^0(\Gamma^N; W(D))$ using finitely many function values. By Assumption 1.8, u_N admits an analytic extension. Further, each function value will be computed by means of a finite element technique. In general, the semi-dicrete finite element solution $\pi_h u_N$ also satisfies the regularity assumpion 1.8. In particular this is true for the linear problem presented in Section 2. We define the numerical approximation $u^N_{h,\mathbf{p}} = \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N$. Our aim is to give a priori estimates for the total error

$$\epsilon = u - u^N_{h,\mathbf{p}} = u - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N$$

where the operator $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$ is described by (3.11) and π_h is the finite element projection operator described by (3.2). We will investigate the error

$$\|u - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N\| \leq \underbrace{\|u - u_N\|}_{(I)} + \underbrace{\|u_N - \pi_h u_N\|}_{(II)} + \underbrace{\|\pi_h u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N\|}_{(III)} \tag{5.1}$$

evaluated in the natural norm $L^2_P(\Omega; W(D))$. Since the error functions in (II) and (III) are finite dimensional the natural norm is equivalent to $L^2_\rho(\Gamma^N; W(D))$. By controlling the error in this natural norm we also control the error in the expected value of the solution, for example:

$$\left\|E[u - u^N_{h,\mathbf{p}}]\right\|_{W(D)} \leq E\left[\left\|u - u^N_{h,\mathbf{p}}\right\|_{W(D)}\right] \leq \left\|u - u^N_{h,\mathbf{p}}\right\|_{L^2_P(\Omega; W(D))}.$$

The quantity (I) controls the truncation error for the case where the input data a_N and f_N are suitable truncations of random fields. This contribution to the total error was considered in [25, Section 4.2]. The quantity (I) is otherwise zero if the representation of a_N and f_N is exact, as in Example 1.3. The second term (II) controls the convergence with respect to h, i.e. the finite element error, which will be dictated by standard approximability properties of the finite element space $W_h(D)$, given by (3.1), and the regularity in space of the solution u (see e.g. [7, 5]). Specifically,

$$\|u_N - \pi_h u_N\|_{L^2_\rho(\Gamma^N; W(D))} \leq C_\pi h^s \left(\int_{\Gamma^N} C(s; u)^2 \rho(y)\, dy\right)^{1/2}$$

by the finite element approximability property (3.1).

The full tensor product convergence results are given by [1, Theorem 1] while the sparse tensor product convergence results for the isotropic Smolyak method are given by [25, Theorem 4.6 and 4.10]. Therefore, we will only concern ourselves with the convergence results when implementing the anisotropic Smolyak algorithm described in Section 3.3. Namely, our primary concern will be to analyze the interpolation error (III)

$$\|\pi_h u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N\|_{L^2_\rho(\Gamma^N; W(D))}, \tag{5.2}$$

for both the Clenshaw-Curtis and Gaussian versions of the anisotropic Smolyak algorithm.

Under the very reasonable assumption that the semi-discrete finite element solution $\pi_h u_N$ admits an analytic extension as described in Assumption 1.8 with the same analyticity region as for u_N, the behavior of the error (5.2) will be analogous to $\|u_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N)u_N\|_{L^2_\rho(\Gamma^N; W(D))}$. For this reason, in the next sections we will analyze the latter.

5.1 Analysis of the interpolation error

In order to to get error bounds for Smolyak's algorithm in the multidimensional case, we will connect the general case to the case $N = 1$ and then apply successively Lemma 4.1 .

As stated in Section 3.3, the sparse grid construction treats all directions differently and is therefore an anisotropic algorithm. Moreover, the convergence analysis presented in Section 5.1.1 exploits the possible anisotropic behaviors of problem (1.1). Therefore, we can expect a faster convergence rate when compared to our previous isotropic Smolyak algorithm [25], for such problems that exhibit strong anisotropic effects. Since the algorithm exploits this behavior and is a sparse interpolation technique, a similar conclusion can be drawn when making convergence comparisons with the anisotropic full tensor product method, introduced in the work [1]. See Section 6 where we explore numerically the consequences of introducing an anisotropy into the model problem described by Example 1.1.

5.1.1 Clenshaw-Curtis interpolation estimates

For $n = 1, 2, \ldots, N$ we begin by letting E_m be the error of the best approximation to functions $u \in C^0(\Gamma_n; W(D))$ by functions $w \in V_m$. Similarly to [4], for $n = 1, \ldots, N$, since $\mathscr{U}^{i_n}$ is exact on $V_{m_{i_n}-1}$ we can apply the general formula

$$\left\|u - \mathscr{U}^{i_n}(u)\right\|_{\infty,1} \leq E_{m_{i_n}-1}(u) \cdot \left(1 + \Lambda_{m_{i_n}}\right) \tag{5.3}$$

where Λ_m is the Lebesgue constant for our choice (3.14). It is known that

$$\Lambda_m \leq \frac{2}{\pi}\log(m-1) + 1 \tag{5.4}$$

for $m \geq 2$, see [10].

Using Lemma 4.1 and with the assumption (4.3) on the decay coefficients, the best approximation to functions $u \in C^0(\Gamma^n; W(D))$ that admit an analytic extension as described by Assumption 1.8 is bounded by:

$$E_{m_{i_n}}(u) \leq C\, \varrho_n^{-m_{i_n}} = C\, e^{-g(n)\, m_{i_n}}, \tag{5.5}$$

where C is a constant dependent on τ and u but otherwise independent of n, defined in Lemma 4.1. For $n = 1, 2, \ldots, N$ we define the one-dimensional indentity operator $I_1^{(n)} : \Gamma_n \to \Gamma_n$, then (5.3)-(5.5) implies

$$\begin{aligned}
&\left\| (I_1^{(n)} - \mathscr{U}^{i_n})(u) \right\|_{\infty,1} \leq C\, \log(m_{i_n}) \varrho_n^{-m_{i_n}} \leq C\, i_n \varrho_n^{-2^{i_n}} = C\, i_n e^{-g(n) 2^{i_n}}, \\
&\left\| (\Delta^{i_n})(u) \right\|_{\infty,1} = \left\| (\mathscr{U}^{i_n} - \mathscr{U}^{i_n - 1})(u) \right\|_{\infty,1} \\
&\qquad \leq \left\| (I_1^{(n)} - \mathscr{U}^{i_n})(u) \right\|_{\infty,1} + \left\| (I_1^{(n)} - \mathscr{U}^{i_n - 1})(u) \right\|_{\infty,1} \\
&\qquad \leq E\, i_n \varrho_n^{-2^{i_n - 1}} = E\, i_n e^{-\frac{g(n)}{2} 2^{i_n}}
\end{aligned} \tag{5.6}$$

for all $i \in \mathbb{N}_+$ with positive constants C and E depending on u but not on i or n.

The convergence proof will be split in several pieces, the main results being given in Theorems 5.3 and 5.7 which state the convergence rates in terms of the level w and the total number of collocation points, respectively.

For the purpose of error analysis we first introduce the set

$$\tilde{X}_{\boldsymbol{\alpha}}(w, d) := \left\{ \mathbf{i} \in \mathbb{N}_+^d, \mathbf{i} \geq \mathbf{1} \,:\, \left(\sum_{n=1}^{d} (i_n - 1)\alpha_n - \underline{\alpha} w \right) \in [0, \alpha_d] \right\}.$$

Moreover, we denote by I_d the identity operator applicable to functions which depend on the first d variables $y_1, \ldots, y_d$. Then the following result holds:

Lemma 5.1 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 4.1 with decay coefficients as in* (4.3)*, the anisotropic Smolyak formula* (3.11) *satisfies:*

$$\left\| \left(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w, N) \right)(u) \right\|_{L^2_\rho(\Gamma^N; W(D))} \leq \sum_{d=1}^{N} R(w, d) \tag{5.7}$$

with

$$R(w, d) := \sum_{\mathbf{i} \in \tilde{X}_{\boldsymbol{\alpha}}(w, d)} C E^{d-1} \left(\prod_{n=1}^{d} i_n \right) e^{-h(\mathbf{i}, d)} \tag{5.8}$$

and

$$h(\mathbf{i}, d) = \sum_{n=1}^{d} g(n) 2^{i_n - 1}, \tag{5.9}$$

with constants C and E defined in (5.5) *and* (5.6)*.*

Proof. We start providing and equivalent representation of the anisotropic Smolyak formula:

$$\begin{aligned}
\mathscr{A}_{\boldsymbol{\alpha}}(w,N) &= \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,N)} \bigotimes_{n=1}^{N} \Delta^{i_n} \\
&= \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \sum_{j=1}^{\left\lfloor 1+w\frac{\alpha}{\alpha_N}-\sum_{n=1}^{N-1}(i_n-1)\frac{\alpha_n}{\alpha_N}\right\rfloor} \Delta^{j} \\
&= \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \mathscr{U}^{\left\lfloor 1+w\frac{\alpha}{\alpha_N}-\sum_{n=1}^{N-1}(i_n-1)\frac{\alpha_n}{\alpha_N}\right\rfloor}
\end{aligned}$$

where we have denoted with $\lfloor\cdot\rfloor$ the integer part of a real number.

The error estimate is computed recursively using the previous representation.

$$\begin{aligned}
I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N) &= I_N - \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \left(\mathscr{U}^{\left\lfloor 1+w\frac{\alpha}{\alpha_N}-\sum_{n=1}^{N-1}(i_n-1)\frac{\alpha_n}{\alpha_N}\right\rfloor} - I_1^{(N)}\right) \\
&\quad - \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes I_1^{(N)} \\
&= \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,N-1)} \bigotimes_{n=1}^{N-1} \Delta^{i_n} \otimes \left(I_1^{(N)} - \mathscr{U}^{\left\lfloor 1+w\frac{\alpha}{\alpha_N}-\sum_{n=1}^{N-1}(i_n-1)\frac{\alpha_n}{\alpha_N}\right\rfloor}\right) \\
&\quad + \left(I_{N-1} - \mathscr{A}_{\boldsymbol{\alpha}}(w,N-1)\right) \otimes I_1^{(N)} \\
&= \sum_{d=2}^{N} \left[\tilde{R}(w,d) \bigotimes_{n=d+1}^{N} I_1^{(n)}\right] + \left(I_1^{(1)} - \mathscr{A}_{\boldsymbol{\alpha}}(w,1)\right) \bigotimes_{n=2}^{N} I_1^{(n)}
\end{aligned}$$

where, for a general dimension d, we define

$$\tilde{R}(w,d) = \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,d-1)} \bigotimes_{n=1}^{d-1} \Delta^{i_n} \otimes \left(I_1^{(d)} - \mathscr{U}^{\hat{i}_d-1}\right)$$

and, for any $(i_1,\dots,i_{d-1}) \in X_{\boldsymbol{\alpha}}(w,d-1)$, we have set $\hat{i}_d = \left\lfloor 2 + w\frac{\alpha}{\alpha_d} - \sum_{n=1}^{d-1}(i_n-1)\frac{\alpha_n}{\alpha_d}\right\rfloor$. Observe that with this definition, the d-dimensional vector $\mathbf{j} = (i_1,\dots,i_{d-1},\hat{i}_d) \in \tilde{X}_{\boldsymbol{\alpha}}(w,d)$.

The term $\tilde{R}(w,d)$ can now be bounded as

$$\begin{aligned}
\left\|\tilde{R}(w,d)(u)\right\|_{\infty,d} &\le \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,d-1)} \prod_{n=1}^{d-1} \left\|(\Delta^{i_n})(u)\right\|_{\infty,d} \left\|\left(I_1^{(d)} - \mathscr{U}^{\hat{i}_d-1}\right)(u)\right\|_{\infty,d} \\
&\le \sum_{\mathbf{i}\in X_{\boldsymbol{\alpha}}(w,d-1)} C\,E^{d-1} \left(\prod_{n=1}^{d-1} i_n\right) (\hat{i}_d-1) e^{-\sum_{n=1}^{d-1} g(n) 2^{i_n-1} - g(d) 2^{\hat{i}_d-1}} \\
&\le \sum_{\mathbf{i}\in \tilde{X}_{\boldsymbol{\alpha}}(w,d)} C\,E^{d-1} \left(\prod_{n=1}^{d} i_n\right) e^{-h(\mathbf{i},d)} =: R(w,d).
\end{aligned}$$

Hence, the interpolation error with the anisotropic Smolyak construction can be bounded by

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))(u)\|_{\infty,N} \leq \sum_{d=2}^{N} R(w,d) + \left\|(I_1^{(1)} - \mathscr{A}_{\boldsymbol{\alpha}}(w,1))(u)\right\|_{\infty,1}.$$

Observe that the first term in the recursion (5.7) can also be bounded by (5.8). Indeed, the set $\tilde{X}_{\boldsymbol{\alpha}}(w,1)$ contains only the point $i_1 = \lfloor 2 + \frac{\underline{\alpha} w}{\alpha_1} \rfloor$ and

$$\begin{aligned}
\left\|\left(I_1^{(1)} - \mathscr{A}_{\boldsymbol{\alpha}}(w,1)\right)(u)\right\|_{\infty,1} &= \left\|\left(I_1^{(1)} - \mathscr{U}^{\left\lfloor 1+\frac{\underline{\alpha} w}{\alpha_1}\right\rfloor}\right)(u)\right\|_{\infty,1} \\
&\leq C \left\lfloor 1 + \frac{\underline{\alpha} w}{\alpha_1} \right\rfloor e^{-g(1)\, 2^{\left\lfloor 1+\frac{\underline{\alpha} w}{\alpha_1}\right\rfloor}} \\
&\leq \sum_{i_1 \in \tilde{X}_{\boldsymbol{\alpha}}(w,1)} C\, i_1\, e^{-g(1)\, 2^{i_1 - 1}} =: R(w,1)
\end{aligned}$$

and this concludes the proof. □

Lemma 5.2 *For the choice $\alpha_n = g(n)$ of the weights in the anisotropic Smolyak formula* (3.11)*, the following bound holds for the term $R(w,d)$, $d = 1, \ldots, N$:*

$$R(w,d) \leq C_1(\mathbf{g},d) \exp\left\{ -\frac{\mathscr{G}(d)}{2} 2^{w \frac{\underline{g}}{\mathscr{G}(d)}} + w \frac{d\underline{g}}{\mathscr{G}(d)} \right\}, \tag{5.10}$$

where the function $C_1(\mathbf{g},d)$ does not depend on w.

Proof. First we convert the sum appearing in (5.8) into an integral. For that, we define the two subsets of $\mathbb{R}^d$:

$$\begin{aligned}
\tilde{Y}_{\boldsymbol{\alpha}}(w,d) &:= \left\{ \mathbf{y} \in \mathbb{R}^d_+ :\ \lfloor y \rfloor + 1 \in \tilde{X}_{\boldsymbol{\alpha}}(w,d) \right\}, \\
\tilde{Y}^+_{\boldsymbol{\alpha}}(w,d) &:= \left\{ \mathbf{y} \in \mathbb{R}^d_+ :\ \left(\sum_{n=1}^{d} y_n \alpha_n - w\underline{\alpha} \right) \in [0, \alpha_d + \sum_{n=1}^{d} \alpha_n] \right\},
\end{aligned}$$

and notice that $\tilde{X}_{\boldsymbol{\alpha}}(w,d) \subset \overline{\tilde{Y}_{\boldsymbol{\alpha}}(w,d)} \subset \tilde{Y}^+_{\boldsymbol{\alpha}}(w,d)$. Then the term $R(w,d)$ can be bounded by

$$\begin{aligned}
R(w,d) &= \int_{\tilde{Y}_{\boldsymbol{\alpha}}(w,d)} C E^{d-1} \left(\prod_{n=1}^{d} (\lfloor y_n \rfloor + 1) \right) e^{-h(\lfloor \mathbf{y} \rfloor + 1, d)}\, d\mathbf{y} \\
&\leq \int_{\tilde{Y}^+_{\boldsymbol{\alpha}}(w,d)} C E^{d-1} \left(\prod_{n=1}^{d} (y_n + 1) \right) e^{-h(\mathbf{y},d)}\, d\mathbf{y}.
\end{aligned} \tag{5.11}$$

Next, we define $y^* = w \frac{\underline{g}}{\mathscr{G}(d)}$ and we expand the function $h(\mathbf{y},d)$ up to second order, around the point $\mathbf{y}^* = (y^*, \ldots, y^*)$:

$$h(\mathbf{y},d) = h(\mathbf{y}^*,d) + \underbrace{\nabla h(\mathbf{y}^*,d) \cdot \delta\mathbf{y}}_{\text{(I)}} + \underbrace{\delta\mathbf{y}^T \frac{1}{2} \nabla^2 h(\mathbf{y}^* + s\delta\mathbf{y}, d) \delta\mathbf{y}}_{\text{(II)}}, \quad \text{with } \delta\mathbf{y} = \mathbf{y} - \mathbf{y}^* \text{ and } s \in [0,1].$$

The linear term is positive on the set $\tilde{Y}^+_{\boldsymbol{\alpha}}(w,d)$; indeed

$$\begin{aligned}(\mathrm{I}) &= \log(2)2^{y^*-1}\sum_{n=1}^{d} g(n)\,(y_n - y^*)\\ &= \log(2)2^{y^*-1}\sum_{n=1}^{d}(y_n\alpha_n - w\underline{\alpha}) \geq 0 \qquad \forall \mathbf{y}\in \tilde{Y}^+_{\boldsymbol{\alpha}}(w,d)\end{aligned}$$

Similarly, the second order remainder can be bounded as

$$\begin{aligned}(\mathrm{II}) &= \frac{\log(2)^2}{2}\sum_{n=1}^{d} g(n)2^{[y^*+s(y_n-y^*)-1]}(y_n-y^*)^2\\ &\geq \frac{\log(2)^2}{4}\sum_{n=1}^{d} g(n)(y_n-y^*)^2\\ &= \sum_{n=1}^{d}\frac{(y_n-y^*)^2}{2\sigma_n^2}, \qquad \text{with } \sigma_n^2 = \frac{2}{\log(2)^2 g(n)}\end{aligned}$$

Finally, the bound (5.11) becomes

$$R(w,d) \leq CE^{d-1}e^{-h(\mathbf{y}^*,d)}\underbrace{\int_{\tilde{Y}^+_{\boldsymbol{\alpha}}(w,d)}\prod_{n=1}^{d}(y_n+1)e^{-\sum_{n=1}^{d}\frac{(y_n-y^*)^2}{2\sigma_n^2}}\,d\mathbf{y}}_{I(d)}.$$

We now turn to estimate the term $I(d)$. For that, let us introduce the function $p_n(y) = e^{-\frac{(y-y^*)^2}{2\sigma_n^2}}/\sqrt{2\pi\sigma_n^2}$ which corresponds to the probability density function of a Normal random variable with mean y^* and variance σ_n^2. Then, we have

$$\begin{aligned}I(d) &= \int_{\tilde{Y}^+_{\boldsymbol{\alpha}}(w,d)}\prod_{n=1}^{d}\sqrt{2\pi\sigma_n^2}(y_n+1)p_n(y_n)\,d\mathbf{y} \qquad \{\text{setting } z_n = y_n+1\}\\ &\leq \prod_{n=1}^{d}\left(\sqrt{2\pi\sigma_n^2}\int_0^\infty z_n p_n(z_n-1)\,dz_n\right) \qquad \{\text{by Cauchy-Schwartz ineq.}\}\\ &\leq \prod_{n=1}^{d}\left(\sqrt{2\pi\sigma_n^2}\left(\int_0^\infty z_n^2 p_n(z_n-1)\,dz_n\right)^{\frac{1}{2}}\right)\\ &\leq \prod_{n=1}^{d}\left(\sqrt{2\pi\sigma_n^2}\sqrt{\sigma_n^2+(1+y^*)^2}\right)\\ &\leq \prod_{n=1}^{d}\left(\sqrt{2\pi\sigma_n^2}\,(1+y^*+\sigma_n)\right) \leq \prod_{n=1}^{d}\left(\sqrt{2\pi\sigma_n^2}e^{y^*+\sigma_n}\right)\\ &= C_2(\boldsymbol{\sigma},d)e^{w\frac{dg}{g(d)}}, \qquad \text{with } C_2(\boldsymbol{\sigma},d) = \prod_{n=1}^{d}\left(\sqrt{2\pi\sigma_n^2}e^{\sigma_n}\right).\end{aligned}$$

From this the final result is easily obtained and inequality (5.10) holds with constant

$$C_1(\mathbf{g},d) = CE^{d-1}\left(\frac{2\sqrt{\pi}}{\log(2)}\right)^d\left(\prod_{n=1}^{d}\frac{1}{\sqrt{g(n)}}\right)\exp\left\{\frac{\sqrt{2}}{\log(2)}\sum_{n=1}^{d}\frac{1}{\sqrt{g(n)}}\right\}. \qquad (5.12)$$

□

Theorem 5.3 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 4.1 with decay coefficients as in* (4.3)*, the anisotropic Smolyak formula* (3.11) *with the choice $\alpha_n = g(n)$ of the weights satisfies:*

$$\left\| \left(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N) \right)(u) \right\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}(\mathbf{g},N) e^{w-\lambda(w,N)} \tag{5.13}$$

where

$$\lambda(w,N) := \begin{cases} w\dfrac{\underline{g}\log(2)e}{2}, & \text{if } 0 \leq w \leq \frac{\mathscr{G}(N)}{\underline{g}\log(2)}, \\ \dfrac{\mathscr{G}(N)}{2} 2^{w\frac{\underline{g}}{\mathscr{G}(N)}}, & \text{otherwise} \end{cases}, \tag{5.14}$$

and the function $\hat{C}(\mathbf{g},N)$ does not depend on w.

Proof. From Lemmas 5.1 and 5.2 we obtain the following bound for the interpolation error

$$\left\| \left(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N) \right)(u) \right\|_{L^2_\rho(\Gamma^N;W(D))} \leq \sum_{d=1}^{N} C_1(\mathbf{g},d) \exp\left\{ w - \widehat{\lambda}(w,d) \right\}$$

with $\widehat{\lambda}(w,d) := \frac{\mathscr{G}(d)}{2} 2^{w\frac{\underline{g}}{\mathscr{G}(d)}}$. We now turn our attention to finding a minimum for $\widehat{\lambda}(w,d)$ for $1 \leq d \leq N$. Let us define, for $s \in \left[1, \frac{\mathscr{G}(N)}{\underline{g}}\right]$, the function

$$p(s) := s\, 2^{\frac{w}{s}}.$$

so that $\widehat{\lambda}(w,d) = \frac{\underline{g}}{2} p\left(\frac{\mathscr{G}(d)}{\underline{g}}\right)$ and

$$\min_{1\leq d\leq N} \widehat{\lambda}(w,d) \geq \min_{s\in[1,\mathscr{G}(N)/\underline{g}]} \frac{\underline{g}}{2} p(s).$$

We have

$$\frac{dp}{ds} = 2^{\frac{w}{s}} \left(1 - \left(\frac{w}{s}\right)\log(2)\right) = 0$$

yielding $s = w\log(2)$. For w sufficiently large, the minimum of $p(s)$ falls outside the interval $\left[1, \frac{\mathscr{G}(N)}{\underline{g}}\right]$ and the function $p(s)$ is decreasing on this interval. Therefore, there are two cases to consider. The first being the situation when $w > \frac{\mathscr{G}(N)}{\underline{g}\log(2)}$ and the second when $0 \leq w \leq \frac{\mathscr{G}(N)}{\underline{g}\log(2)}$. In either case

$$\min_{1\leq d\leq N} \widehat{\lambda}(w,d) \geq \lambda(w,N)$$

and hence,

$$\max_{1\leq d\leq N} e^{-\widehat{\lambda}(w,d)} \leq e^{-\lambda(w,N)}.$$

From this the result (5.13) follows, by taking

$$\hat{C}(\mathbf{g},N) = N \max_{d=1,\ldots,N} C_1(\mathbf{g},d). \tag{5.15}$$

□

Remark 5.4 *From the expressions* (5.15) *and* (5.12) *we observe that the constant* $\hat{C}(\mathbf{g}, N)$ *appearing in the convergence estimates of Theorem 5.3 goes to infinity when* $\underline{g}$ *tends to zero. We should note that in such case we loose convergence anyhow.*

On the other hand, for given $\underline{g}$ *and* N*, the more the sequence* $\mathbf{g}$ *is anisotropic (i.e. the larger the ratio* $g_{max}/\underline{g}$ *gets), the smaller becomes the constant* $\hat{C}(\mathbf{g}, N)$*. In any case, such constant can be bounded by*

$$\hat{C}(\mathbf{g}, N) \leq \frac{C}{E} \exp\left\{ \frac{N}{\sqrt{\underline{g}}} \left(\frac{2E\sqrt{\pi} + \sqrt{2}}{\log(2)} \right) \right\},$$

independently of the anisotropy.

Now we relate the number of collocation points $\eta = \eta(w, N) = \#\mathscr{H}_{\boldsymbol{\alpha}}(w, N)$ to the level w of the anisotropic Smolyak algorithm. We state the result in the following lemma:

Lemma 5.5 *Using the anisotropic Smolyak interpolant described by* (3.8) *where the abscissas are the Clenshaw-Curtis knots, described in Section 3.3.1, the total number of points required at level* w *satisfies the following bounds:*

$$2^{w-1} \leq \eta \leq 2^w e^{w \sum_{n=1}^N \underline{\alpha}/\alpha_n}, \tag{5.16}$$

Moreover, as a direct consequence of (5.16) *we get that:*

$$\frac{\log(\eta)}{\sum_{n=1}^N \frac{\underline{\alpha}}{\alpha_n} + \log(2)} \leq w \leq \log_2(2\eta) \tag{5.17}$$

Proof. By using formula (3.8) and exploiting the nested structure of the Clenshaw-Curtis abscissas the number of points $\eta = \eta(w, N) = \#\mathscr{H}_{\boldsymbol{\alpha}}(w, N)$ can be counted in the following way:

$$\eta = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \prod_{n=1}^N r(i_n), \text{ where } r(i) := \begin{cases} 1 & \text{if } i = 1 \\ 2 & \text{if } i = 2 \\ 2^{i-2} & \text{if } i > 2 \end{cases}. \tag{5.18}$$

Begin by noticing that for all $n = 1, 2, \ldots, N$ the following bound holds:

$$2^{i_n - 2} \leq r(i_n) \leq 2^{i_n - 1}. \tag{5.19}$$

Next, let $m \in [1, N]$ be the index corresponding to the minimum α, i.e. $\alpha_m = \min_{1 \leq n \leq N} \{\alpha_n\} = \underline{\alpha}$. A lower bound on the numebr η of points can be obtained considering only the contribution from the tensor grid of indices $i_n = 1$, for $n \neq m$ and $i_m = w + 1$. On the other hand, it is easy to see that $|\mathbf{i} - \mathbf{1}| = \sum_{n=1}^N (i_n - 1) \leq w$ so the following bounds hold:

$$2^{w-1} \leq \eta = \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \prod_{n=1}^N r(i_n) \leq \sum_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} 2^{|\mathbf{i}-\mathbf{1}|} \leq 2^w \# X_{\boldsymbol{\alpha}}(w, N).$$

We need now a bound for the cardinality of the set $X_{\boldsymbol{\alpha}}(w,N)$. We prove by induction that

$$\#X_{\boldsymbol{\alpha}}(w,N) \leq \prod_{n=1}^{N}\left(w\frac{\underline{\alpha}}{\alpha_n}+1\right). \tag{5.20}$$

Indeed, the result is obviously true for $N=1$, and assuming that it holds for $N-1$, we have

$$\begin{aligned}
\#X_{\boldsymbol{\alpha}}(w,N) &= \sum_{j_N=1}^{\lfloor w\underline{\alpha}/\alpha_N\rfloor+1} \#X_{\boldsymbol{\alpha}}\left(w-\frac{(j_N-1)\alpha_N}{\underline{\alpha}},N-1\right)\\
&= \sum_{j_N=1}^{\lfloor w\underline{\alpha}/\alpha_N\rfloor+1}\prod_{n=1}^{N-1}\left(w\frac{\underline{\alpha}}{\alpha_n}-\frac{(j_N-1)\alpha_N}{\alpha_n}+1\right)\\
&\leq \sum_{j_N=1}^{\lfloor w\underline{\alpha}/\alpha_N\rfloor+1}\prod_{n=1}^{N-1}\left(w\frac{\underline{\alpha}}{\alpha_n}+1\right)\\
&\leq \prod_{n=1}^{N}\left(w\frac{\underline{\alpha}}{\alpha_n}+1\right)
\end{aligned}$$

and this finishes the induction proof. Moreover we have

$$\#X_{\boldsymbol{\alpha}}(w,N) \leq \exp\left\{w\sum_{n=1}^{N}\frac{\underline{\alpha}}{\alpha_n}\right\} \tag{5.21}$$

and the inequalities (5.16) and (5.17) follow. □

Remark 5.6 *The bound for the cardinality of the set $\#X_{\boldsymbol{\alpha}}(w,N)$ given in* (5.20) *is not sharp when $w\to\infty$ and actually one has instead the asymptotic behaviour*

$$\#X_{\boldsymbol{\alpha}}(w,N) \lesssim \frac{w^N}{N!}\prod_{n=1}^{N}\frac{\underline{\alpha}}{\alpha_n},$$

which is consistent with the isotropic result given in [25]. Moreover, if we use Stirling's approximation for the factorial term, the previous bound becomes

$$\#X_{\boldsymbol{\alpha}}(w,N) \lesssim \frac{1}{\sqrt{N}}\prod_{n=1}^{N}\frac{ew}{N}\frac{\underline{\alpha}}{\alpha_n} \leq \frac{1}{\sqrt{N}}\exp\left\{w\left(\frac{e}{N}\sum_{n=1}^{N}\frac{\underline{\alpha}}{\alpha_n}\right)\right\},$$

which greatly improves (5.21).

The next Theorem provides an error bound in terms of the total number η of collocation points. The proof follows directly from the results in Theorem 5.3 and Lemma 5.5 and is therefore omitted.

Theorem 5.7 *For functions $u\in L^2_\rho(\Gamma^N;W(D))$ satisfying the assumption of Lemma 4.1 with decay coefficients as in* (4.3)*, the anisotropic Smolyak formula* (3.11) *with the choice $\alpha_n = g(n)$ of the weights satisfies:*

• Algebraic convergence $\left(0 \leq w \leq \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$. *Under the assumption that* $\underline{g} \geq 1/(e\log(2))$,

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}(\mathbf{g},N)\eta^{-\mu_1}, \qquad \mu_1 = \frac{\underline{g}\log(2)e - 1}{\log(2) + \sum_{n=1}^N \underline{g}/g(n)}. \tag{5.22}$$

• Sub-exponential convergence $\left(w > \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}(\mathbf{g},N)(2\eta)^{1/\log(2)} e^{-\frac{\mathscr{G}(N)}{2}\eta^{\mu_2}},$$
$$\text{with} \qquad \mu_2 = \frac{\underline{g}\log(2)}{\mathscr{G}(N)\left(\log(2) + \sum_{n=1}^N \underline{g}/g(n)\right)}. \tag{5.23}$$

and constant $\hat{C}(\mathbf{g},N)$ *defined in* (5.15) *and independent of* η.

Remark 5.8 *The estimates given in* (5.23) *may be improved when* $w \to \infty$. *Such asymptotic estimate is obtained using the better counting result described in Remark 5.6.*

Remark 5.9 *We observe that sub-exponential rate of convergence is always faster than the algebraic one when* $w > \mathscr{G}(N)/(\underline{g}\log(2))$. *Yet, this estimate is of little practical relevance since in practical computations, such a high level* w *is seldom reached.*

Remark 5.10 *The condition* $\underline{g} > 1/(log(2)e)$ *in the algebraic regime can be improved following and* L^2 *analysis. Yet it is largely satisfied in all our numerical tests. In the next Section we present results using Gaussian abscissas where this condition is no longer needed.*

Remark 5.11 *Suppose now that the stochastic input data are truncated expansions of random fields and that we are able to estimate the values* $\{g(n)\}_{n=1}^\infty$. *Whenever the sum* $\sum_{n=1}^\infty \underline{g}/g(n)$ *is finite then the algebraic exponent in* (5.22) *does not deteriorate as the truncation dimension* N *increases. This condition is satisfied for instance by the problem presented in the numerical Section. This is a clear advantage with respect to the isotropic Smolyak method.*

Remark 5.12 (Optimal choice of $\boldsymbol{\alpha}$) *Looking at the exponential term* $e^{-h(\mathbf{i},d)}$ *in* (5.8), *which is the term determining the rate of convergence, we may try to choose the weight* $\boldsymbol{\alpha}$ *for* $X_{\boldsymbol{\alpha}}(w,N)$ *as the solution to the optimization problem*

$$\max_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^d_+ \\ |\boldsymbol{\alpha}|=1}} \min_{\mathbf{i}\in\bar{X}_{\boldsymbol{\alpha}}(w,d)} h(\mathbf{i},d)$$

This problem has the solution $\boldsymbol{\alpha} = \mathbf{g}$ *and hence, our choice of weights* (4.4) *is optimal.*

5.1.2 Gaussian interpolation estimates

By using a similar approach to our previous work [25, Section 4.1.2] we can develop error estimates for interpolating functions $u \in C^0(\Gamma^N; W(D))$ that admit an analytic extension as described by Assumption 1.8 using the anisotropic Smolyak formulations based on Gaussian abscissas described in Section 3.3.2. We remind the reader that in the global estimate (5.1) we need to bound the interpolation error (III) in the norm $L^2_\rho(\Gamma^N; W(D))$. Yet, the Gaussian points defined in Section 3.3.2 are constructed for the more appropriate density $\hat{\rho} = \prod_{n=1}^N \hat{\rho}_n$ and we have the following useful bound:

$$\|v\|_{L^2_\rho(\Gamma^N;W(D))} \leq \left\| \frac{\rho}{\hat{\rho}} \right\|_{L^\infty(\Gamma^N)} \cdot \|v\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))} \qquad \text{for all } v \in C^0(\Gamma^N; W(D)).$$

In what follows we will use the shorthand notation $\|\cdot\|_{\hat{\rho},N}$ for $\|\cdot\|_{L^2_{\hat{\rho}}(\Gamma^N;W(D))}$. Following [25], the 1d interpolation error for Gaussian abscissas satisfies

$$\left\|(I_1^{(n)} - \mathscr{U}^{i_n})(u)\right\|_{\hat{\rho},1} \leq \tilde{C}\, e^{-g(n)2^{i_n}}, \qquad \text{and} \qquad \left\|(\Delta^{i_n})(u)\right\|_{\hat{\rho},1} \leq \tilde{E}\, e^{-\frac{g(n)}{2}2^{i_n}}$$

for all $i \in \mathbb{N}_+$ with positive constants $\tilde{C}$ and $\tilde{E}$ depending on u but not on i or n.

The following Theorem states the rate of convergence for the anisotropic Smolyak formula based on Gaussian abscissas. Since the proof is mostly similar to the one presented in the previous Section for Clenshaw-Curtis points, it will be just sketched and only the main differences will be highlighted.

Theorem 5.13 *For functions $u \in L^2_\rho(\Gamma^N; W(D))$ satisfying the assumption of Lemma 4.1 with decay coefficients as in* (4.3)*, the anisotropic Smolyak formula* (3.11)*, based on Gaussian abscissas and with the choice $\alpha_n = g(n)$ of the weights satisfies:*

- Algebraic convergence $\left(0 \leq w \leq \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$*:*

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}^G(\mathbf{g},N)\eta^{-\mu_1}, \qquad \mu_1 = \frac{\underline{g}\log(2)e}{2\log(2) + \sum_{n=1}^N \underline{g}/g(n)}. \tag{5.24}$$

- Sub-exponential convergence $\left(w > \frac{\mathscr{G}(N)}{\underline{g}\log(2)}\right)$*:*

$$\|(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N))(u)\|_{L^2_\rho(\Gamma^N;W(D))} \leq \hat{C}^G(\mathbf{g},N)e^{-\frac{\mathscr{G}(N)}{2}\eta^{\mu_2}},$$
$$\text{with} \qquad \mu_2 = \frac{\underline{g}\log(2)}{\mathscr{G}(N)\left(2\log(2) + \sum_{n=1}^N \underline{g}/g(n)\right)}. \tag{5.25}$$

and constant $\hat{C}^G(\mathbf{g},N)$ independent of η.

Proof. The recursion formula (5.7) still holds with the term $R(w,d)$ defined now as

$$R(w,d) = \left\| \frac{\rho}{\hat{\rho}} \right\| \sum_{\mathbf{i} \in \tilde{X}_{\boldsymbol{\alpha}}(w,d)} \tilde{C} \tilde{E}^{d-1} e^{-h(\mathbf{i},d)}$$

Notice that the negative effect of the Lebesgue constant for the Clenshaw-Curtis points, which was responsible for the term $\prod_{n=1}^{d} i_n$ in the definition of $R(w,d)$ (see (5.8)) is not present any more. Then, following the guidelines of the proof of Lemma 5.2, this term can be bounded as

$$R(w,d) \leq C_1^G(\mathbf{g},d) \exp\left\{ -\frac{\mathscr{G}(d)}{2} 2^{w \frac{g}{\mathscr{G}(d)}} \right\}.$$

with constant

$$C_1^G(\mathbf{g},d) = \left\| \frac{\rho}{\hat{\rho}} \right\| \tilde{C} \tilde{E}^{d-1} \left(\frac{2\sqrt{\pi}}{\log(2)} \right)^d \left(\prod_{n=1}^{d} \frac{1}{\sqrt{g(n)}} \right).$$

This leads to the estimate in terms of w

$$\left\| \left(I_N - \mathscr{A}_{\boldsymbol{\alpha}}(w,N) \right)(u) \right\|_{L^2_\rho(\Gamma^N; W(D))} \leq \hat{C}^G(\mathbf{g},N) e^{-\lambda(w,N)} \tag{5.26}$$

with the same $\lambda(w,N)$ as in (5.14) and $\hat{C}^G(\mathbf{g},N) = N \max_{d=1,\ldots,N} C_1^G(\mathbf{g},d)$. Again, notice the great improvement with respect to the bound (5.13) holding for Clenshaw-Curtis points. Finally, we observe that for a given level w, the number of Gauss points is larger than the number of Clenshaw-Curtis points (due to the non-nested structure) and, following [25], we know that

$$\eta = \sum_{\mathbf{i} \in Y_{\boldsymbol{\alpha}}(w,N)} \prod_{n=1}^{N} \tilde{r}(i_n), \qquad \text{with} \quad \tilde{r}(i) := \begin{cases} 1 & \text{for } i = 1 \\ 2^{i-1} + 1 & \text{for } i > 1 \end{cases} \tag{5.27}$$

$$\prod_{n=1}^{N} \tilde{r}(i_n) \leq \prod_{n=1}^{N} (2^{i_n - 1} + 1) \leq 2^{2|i-1|} \leq 2^{2w}$$

and the number of points can be bounded as

$$2^{w-1} \leq \eta \leq 2^{2w} \# Y_{\boldsymbol{\alpha}}(w,N) \leq 2^{2w} \# X_{\boldsymbol{\alpha}}(w,N),$$

which, substituted in (5.26) gives the desired result. □

6 Numerical Examples

This Section illustrates the convergence of the anisotropic sparse collocation method for the stochastic linear elliptic problem in two spatial dimensions, as described in Section 2. The computational results are in accordance with the convergence rates predicted by the theory. Actually, we observe a faster convergence than stated in (5.22) and (5.24), which hints that the current estimates may be improved.

We will also use this problem to compare the convergence of the anisotropic Smolyak method with other ensemble-based methods such as: the isotropic Smolyak method described in [25], the anisotropic adaptive full tensor product method described in the work [3, Section 9] and finally, the well-known Monte Carlo method.

The problem is to solve

$$\begin{cases} -\nabla \cdot (a(\omega,\cdot)\nabla u(\omega,\cdot)) = f(\omega,\cdot) \quad \text{in } D\times\Omega, \\ u(\omega,\cdot) = 0 \quad \text{on } \partial D\times\Omega. \end{cases} \tag{6.1}$$

with $D=[0,d]^2$ and $d=1$. For this numerical example we take a deterministic load $f(\omega,x,z)=\cos(x)\sin(z)$ and construct the random diffusion coefficient $a_N(\omega,x)$ with one-dimensional spatial dependence as

$$\log(a_N(\omega,\tilde{x})-0.5) = 1+\mathrm{Y}_1(\omega)\left(\frac{\sqrt{\pi}L}{2}\right)^{1/2} + \sum_{n=2}^{N}\zeta_n\,\varphi_n(x)\,\mathrm{Y}_n(\omega). \tag{6.2}$$

where

$$\zeta_n := \left(\sqrt{\pi}L\right)^{1/2}\exp\left(\frac{-\left(\lfloor\frac{n}{2}\rfloor\pi L\right)^2}{8}\right), \quad \text{if } n>1 \tag{6.3}$$

and

$$\varphi_n(x) := \begin{cases} \sin\left(\frac{\lfloor\frac{n}{2}\rfloor\pi x}{L_p}\right), & \text{if } n \text{ even}, \\ \cos\left(\frac{\lfloor\frac{n}{2}\rfloor\pi x}{L_p}\right), & \text{if } n \text{ odd}. \end{cases} \tag{6.4}$$

For $x\in[0,d]$ let L_c be a desired physical correlation length for the coefficient a, meaning that the random variables $a(x)$ and $a(y)$ become essentially uncorrelated for $|x-y| >> L_c$. Then, the parameter L_p in (6.4) and (6.3) is $L_p=\max\{d,2L_c\}$ and the parameter L in (6.2) and (6.3) is

$$L=\frac{L_c}{L_p}.$$

In this example, the random variables $\{Y_n(\omega)\}_{n=1}^{\infty}$ are independent, have zero mean and unit variance, i.e. $\mathbb{E}[Y_n]=0$ and $\mathbb{E}[Y_nY_m]=\delta_{nm}$ for $n,m\in\mathbb{N}_+$, and are uniformly distributed in the interval $[-\sqrt{3},\sqrt{3}]$. Expression (6.2) is truncation of a one-dimensional random field with stationary covariance

$$\begin{aligned}\mathrm{cov}[\log(a_N-0.5)](x_1,x_2) &= \mathbb{E}\big[(\log(a)(x_1)-\mathbb{E}[\log(a)](x_1))\,((\log(a)(x_2)-\mathbb{E}[\log(a)](x_2))\big] \\ &= \exp\left(\frac{-(x_1-x_2)^2}{L_c^2}\right).\end{aligned}$$

We consider now an a priori selection of the weights $g(n)$ based on formula (4.6). For that, from (2.7) we can estimate the parameters τ_n by

$$\tau_n = \frac{1}{4\zeta_n \|\varphi\|_\infty} = \begin{cases} \sqrt{\frac{1}{8\sqrt{\pi}L}}, & \text{for } n = 1 \\ \sqrt{\frac{1}{16\sqrt{\pi}L}} \exp\left(\frac{\lfloor \frac{n}{2} \rfloor^2 \pi^2 L^2}{8}\right), & \text{for } n > 1. \end{cases}$$

Then, using (4.6) the weight vector $\mathbf{g}$ becomes

$$g(n) = \begin{cases} \log\left(1 + \sqrt{\frac{1}{24\sqrt{\pi}L}}\right), & \text{for } n = 1 \\ \log\left(1 + \sqrt{\frac{1}{48\sqrt{\pi}L}} \exp\left(\frac{\lfloor \frac{n}{2} \rfloor^2 \pi^2 L^2}{8}\right)\right), & \text{for } n > 1 \end{cases} \tag{6.5}$$

and then we have $\varrho_n \geq e^{g(n)}$ for all $n = 1, 2, \ldots, N$.

In Table 2 we show the function values $g(n)$ for $n = 1, 2, \ldots, N = 11$ using (6.5). With this *a priori* information we can construct the simplices $\mathbf{i} \in X_{\boldsymbol{\alpha}}(w, N)$. These become the indices $\mathbf{i} \in \mathbb{N}_+^N$ used for solving problem (6.1) with correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$. This table also yields insight into the anisotropic behavior of each problem. In the case of small correlation lengths, i.e. $L_c = 1/64$, we observe an almost equal weighing of all stochastic directions, except for the first one. The opposite behavior can be seen as we increase the correlation length. For example, when $L_c = 1/2$ the ratio between $g(11)$ and $\underline{g} = g(2) = g(3)$ is approximately $30 : 1$.

	$g(1)$	$g(2), g(3)$	$g(4), g(5)$	$g(6), g(7)$	$g(8), g(9)$	$g(10), g(11)$
$L_c = 1/2$	0.20	0.19	0.42	1.24	3.1	5.8
$L_c = 1/4$	0.27	0.21	0.26	0.36	0.56	0.91
$L_c = 1/16$	0.48	0.36	0.37	0.37	0.38	0.40
$L_c = 1/64$	0.79	0.62	0.62	0.62	0.62	0.62

Table 2: The $N = 11$ values of the function $g(n)$ constructed from *a priori* information given by (6.5) for correlation lengths $L_c = 1/2, 1/4, 1/16, 1/64$.

Since the random variables Y_n are uniformly distributed, in this case the Gaussian abscissas correspond to the root of the Legendre polynomials. Recall from Section 3.3.1 that the Clenshaw-Curtis abscissas are nested and therefore, the number of points $\eta = \eta(w, N) = \#\mathscr{H}_{\boldsymbol{\alpha}}(w, N)$ can be counted as in formula (5.18). On the other hand, the Gaussian abscissas, described in Section 3.3.2, are not nested and hence, we can count the number of points η used by the Smolyak interpolant as in (5.27).

The finite element space for the spatial discretization is the span of continuous functions that are piecewise polynomials with degree two over a uniform triangulation of D with 4225 unknowns.

Observe, in general, that the collocation method only requires the solution of uncoupled deterministic problems over the set of collocation points, even in the presence of a diffusivity coefficient which depends nonlinearly on the random

variables as in (6.2). This is a significant advantage that the collocation method offers compared to the classical Stochastic-Galerkin finite element method as considered, for instance, in [2, 12, 23, 33]. To study the convergence of the anisotropic Smolyak algorithm we consider a problem with a fixed dimension N and investigate the behavior when the level w of the interpolation in the Smolyak algorithm is increased linearly.

The computational results for the $L^2(D)$ approximation error to the expected value, $\mathbb{E}[u]$, are shown in Figures 4. Here we consider the truncated probability space to have dimensions $N = 5$ and $N = 11$ and we compute approximate solutions up to level $\overline{w}$. To estimate the computational error in the w-th level, for $0 \leq w \leq \overline{w}$, first we denote the maximum index utilized by $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$ in each stochastic direction, $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w) \in \mathbb{N}_+^N$, given component-wise as:

$$\kappa_n(\boldsymbol{\alpha}, w) \equiv \max_{\mathbf{i} \in X_{\boldsymbol{\alpha}}(w,N)} \{i_n\}. \tag{6.6}$$

Then we introduce an enriched solution,

$$\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)\pi_h u_N, \quad \text{with } \widehat{\alpha}_n = \left(\frac{\kappa_n(\boldsymbol{\alpha}, \overline{w}) - 1}{\kappa_n(\boldsymbol{\alpha}, \overline{w})}\right)\left(\frac{\alpha_n}{\underline{\alpha}}\right), \tag{6.7}$$

and approximate the computational error for $w = 0, 1, 2, \ldots, \overline{w}$ as

$$\|\mathbb{E}[\epsilon]\| \approx \|\mathbb{E}[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)\pi_h u_N]\| \tag{6.8}$$

By construction, the enriched solution possesses a maximum index $\boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w}+1)$ which is larger by one in each direction, i.e. $\kappa_n(\widehat{\boldsymbol{\alpha}}, \overline{w}+1) = \kappa_n(\boldsymbol{\alpha}, \overline{w}) + \mathbf{1}$, $n = 1, \ldots, N$.

Tables 3 and 4 show the values $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ for computing $\mathscr{A}_{\boldsymbol{\alpha}}(w, N)$, with $w = 0, 1, 2, \ldots, \overline{w}$, as well as the value $\boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w})$ for computing the enriched solution, $\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)$, for the cases $L_c = 1/2$ and $L_c = 1/64$, respectively. The convergence plots shown in Figures 4 and 5 confirm, as expected, that the error decreases sub-exponentially, as the level w increases linearly. For highly anisotropic problems, i.e. $L_c = 1/2$, we observe that the rate of convergence is increased significantly with respect to the isotropic Smolyak method, as we anticipated in Theorem 5.7. We also observe that the convergence rate is dimension dependent and slightly deteriorates as N increases.

To investigate the performance of the algorithm by varying the correlation length L_c we examine Figure 5. We notice that the larger correlation lengths have positive effects on the rate of convergence. This can be explained by examining $g(n)$ defined by (6.5). From this we see that as L_c becomes large the higher dimensions weigh less which greatly reduces the number of function evaluation required by the anisotropic sparse collocation method. On the other hand, the effect of decreasing L_c is a deterioration of the rate of convergence, due to the equal weighing of all directions. In this case, our anisotropic Smolyak algorithm and the isotropic Smolyak obtain an equivalent convergence rate. Our final

w	κ_1	κ_2=κ_3	κ_4=κ_5	κ_6=κ_7	κ_8=κ_9	κ_{10}=κ_{11}
0	1	1	1	1	1	1
1	1	2	1	1	1	1
2	2	3	1	1	1	1
3	3	4	2	1	1	1
4	4	5	2	1	1	1
5	5	6	3	1	1	1
$\widehat{\boldsymbol{\kappa}}$	6	7	4	2	2	2

Table 3: The $N = 11$ components of the maximum indices $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ and $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w} + 1)$, defined by (6.6) and (6.7), respectively, used for solving problem (6.1) with a correlation length $L_c = 1/2$.

w	κ_1	κ_2=κ_3	κ_4=κ_5	κ_6=κ_7	κ_8=κ_9	κ_{10}=κ_{11}
0	1	1	1	1	1	1
1	1	2	2	2	2	2
2	2	3	3	3	3	3
3	3	4	4	4	4	4
4	4	5	5	5	5	5
$\widehat{\boldsymbol{\kappa}}$	5	6	6	6	6	6

Table 4: The $N = 11$ components of the maximum indices $\boldsymbol{\kappa}(\boldsymbol{\alpha}, w)$ and $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}(\widehat{\boldsymbol{\alpha}}, \overline{w})$, defined by (6.6) and (6.7), respectively, used for solving problem (6.1) with a correlation length $L_c = 1/64$.

interest is to compare our sparse tensor product methods, both isotropic and anisotropic, with an anisotropic full tensor product method, proposed in [3] and also with the Monte Carlo method.

The isotropic Smolyak algorithm for solving problem (6.1) was analyzed in [25] and can be constructed by equally weighing all stochastic directions, i.e. $\boldsymbol{\alpha} = \mathbf{1} \in \mathbb{R}^N_+$. To estimate the computational error in the w-th level we approximate $\|\mathbb{E}[\epsilon]\| \approx \|\mathbb{E}[\mathscr{A}(w, N)\pi_h u_N - \mathscr{A}(w+1, N)\pi_h u_N]\|$ using either Gaussian or Clenshaw-Curtis abscissas.

The anisotropic full tensor product algorithm can be described in the following way: given a tolerance *tol* the method computes a multi-index $\mathbf{p} = (p_1, p_2, \ldots, p_N)$, corresponding to the order of the approximating polynomial spaces $\mathcal{P}_{\mathbf{p}}(\Gamma^N)$. This adaptive algorithm increases the tensor polynomial degree with an anisotropic strategy: it increases the order of approximation in one direction as much as possible before considering the next direction. Table 5 and Table 6 show the values of components of the 11-dimensional multi-index $\mathbf{p}$ for different values of *tol*, corresponding to $L_c = 1/2$ and $L_c = 1/64$ respectively. These tables can also give insight into the anisotropic behavior of each particular problem and should be compared with Tables 5 and Table 6, respectively. Ob-

serve, in particular, for the case $L_c = 1/64$ the algorithm predicts a multi-index $\mathbf{p}$ which is equal in all directions, i.e. an isotropic tensor product space. A convergence plot for $L_c = 1/2$ and $L_c = 1/64$ can be constructed by examining each row of the Table 5 and Table 6 respectively, and plotting the number of points in the tensor product grid versus the error in expectation. We estimate the error in expectation by $\|\mathbb{E}[\epsilon]\| \approx \|\mathbb{E}[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]\|$, with $\widetilde{\mathbf{p}} = (p_1+1, p_2+1, \ldots, p_N+1)$. This entails an additional computational cost, which is bounded by the factor $\exp\left(\sum_{n=1}^{N} 1/p_n\right)$ times the work to compute $\mathbb{E}[u^N_{h,\mathbf{p}}]$.

tol	$N=1$	$N=2,3$	$N=4,5$	$N=6,7$	$N=8,9$	$N=10,11$
1.0e-04	1	1	1	1	1	1
1.0e-05	2	1	1	1	1	1
1.0e-06	2	2	1	1	1	1
1.0e-07	3	2	2	1	1	1
1.0e-08	4	3	2	1	1	1
1.0e-09	4	4	3	1	1	1
1.0e-10	5	5	3	2	1	1
1.0e-11	5	5	4	2	1	1
1.0e-12	5	6	4	2	1	1

Table 5: The $N = 11$ components of the multi index $\mathbf{p}$ computed by he anisotropic full tensor product algorithm when solving problem (6.1) with a correlation length $L_c = 1/2$.

tol	$N=1$	$N=2,3$	$N=4,5$	$N=6,7$	$N=8,9$	$N=10,11$
1.0e-03	1	1	1	1	1	1
1.0e-06	2	2	2	2	2	2
1.0e-09	3	3	3	3	3	3
1.0e-12	4	4	4	4	4	4

Table 6: The $N = 11$ components of the multi index $\mathbf{p}$ computed by the anisotropic full tensor product algorithm when solving problem (6.1) with a correlation length $L_c = 1/64$.

The standard Monte Carlo Finite Element Method is the most common choice for anyone solving SPDEs such as (6.1) [6, 3]. If the aim is to compute a functional of the solution such as the expected value, one would approximate $\mathbb{E}[u]$ numerically by sample averages of iid realizations of the stochastic input data. Given a number of realizations, $M \in \mathbb{N}_+$, we compute the sample average as follows: For each $k = 1, \ldots, M$, sample iid realizations of $a(\omega_k, \cdot)$ and $f(\omega_k, \cdot)$, solve problem (6.1) and construct finite element approximations $u^N_h(\omega_k, \cdot)$. We note that once we have fixed $\omega = \omega_k$, the problem is completely deterministic, and may be solved by standard methods as in the collocation approach. Finally, approximate $\mathbb{E}[u]$ by the sample average: $\overline{\mathbb{E}}[u^N_{h,k}; M](\cdot) := \frac{1}{M}\sum_{k=1}^{M} u^N_h(\omega_k, \cdot)$.

For the cases $L_c = 1/2, 1/4, 1/16$ and $1/64$ we take $M = 2^i$, $i = 0, 1, 2, \ldots, 11$ realizations and compute the approximation to the error in expectation by $\|\mathbb{E}[\epsilon]\| \approx \|\overline{E}[u_{h,k}^N; M] - \mathbb{E}[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)\pi_h u_N]\|$, where $\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}+1, N)$ is the enriched anisotropic sparse solution defined previously and $\widehat{\boldsymbol{\alpha}}$ is defined by (6.7).

To study the advantages of an anisotropic sparse tensor product space as opposed to the isotropic sparse tensor product space or an anisotropic full tensor product space we show, in Figure 6, the convergence of these methods when solving problem (6.1), using correlation lengths $L_c = 1/2, 1/4, 1/16$ and $L_c = 1/64$ with $N = 11$. We also include 5 ensembles of the Monte Carlo method described previously. Figure 6 reveals that for the isotropic case with $L_c = 1/64$ the anisotropic and isotropic Smolyak method obtain a comparable convergence rate, both faster than the anisotropic full tensor product method.

On the contrary, opposite behavior can be observed for $L_c = 1/2$. Since, in this case, the rate of decay of the expansion is faster, the anisotropic full tensor method weighs heavily the important modes and, therefore, achieves a faster convergence than the isotropic Smolyak method. Similar conclusions can be made for the anisotropic Smolyak: the increased convergence for this method comes from the fact that it combines an optimal treatment of the problem anisotropy while reducing the *curse of dimensionality* via the use of sparse grids.

In all four cases we observe that all the 3 methods out-perform the Monte Carlo method. We know that the amount of work to reach the accuracy ϵ in the Monte Carlo approach can be approximated by $\epsilon \approx O(M^{-1/2})$ times the amount of work per sample, where M is the number of samples. This is only affected by the problem dimension through the increase of the work per sample. Nevertheless, the convergence rate is quite slow and a high level of accuracy is only achieved when an large amount of function evaluations are required. This can been seen from Figure 6 where we include reference lines with slopes $-1/2$ and -1, respectively, or in Table 7 where, for $N = 11$, we compare the work, proportional to the number of samples, which is the number of collocation points, required by each method to decrease the original error by a factor of 10^4 for all four correlation lengths $L_c = 1/2, 1/4, 1/16$ and $L_c = 1/64$.

L_c	AS	AF	IS	MC
$1/2$	50	2.5×10^2	2.5×10^3	5.0×10^9
$1/4$	1.6×10^2	1.2×10^3	4.0×10^3	2.0×10^9
$1/16$	2.0×10^2	2.0×10^3	5.0×10^2	1.6×10^9
$1/64$	3.1×10^2	2.0×10^5	3.6×10^2	1.3×10^9

Table 7: For $N = 11$, we compare the number of function evaluations required by the Anisotropic Smolyak (AS) using Clenshaw-Curtis abscissas, Anisotropic Full Tensor product method (AF) using Gaussian abscissas, Isotropic Smolyak (IS) using Clenshaw-Curtis abscissas and the Monte Carlo (MC) method using random abscissas, to reduce the original error by a factor of 10^4.

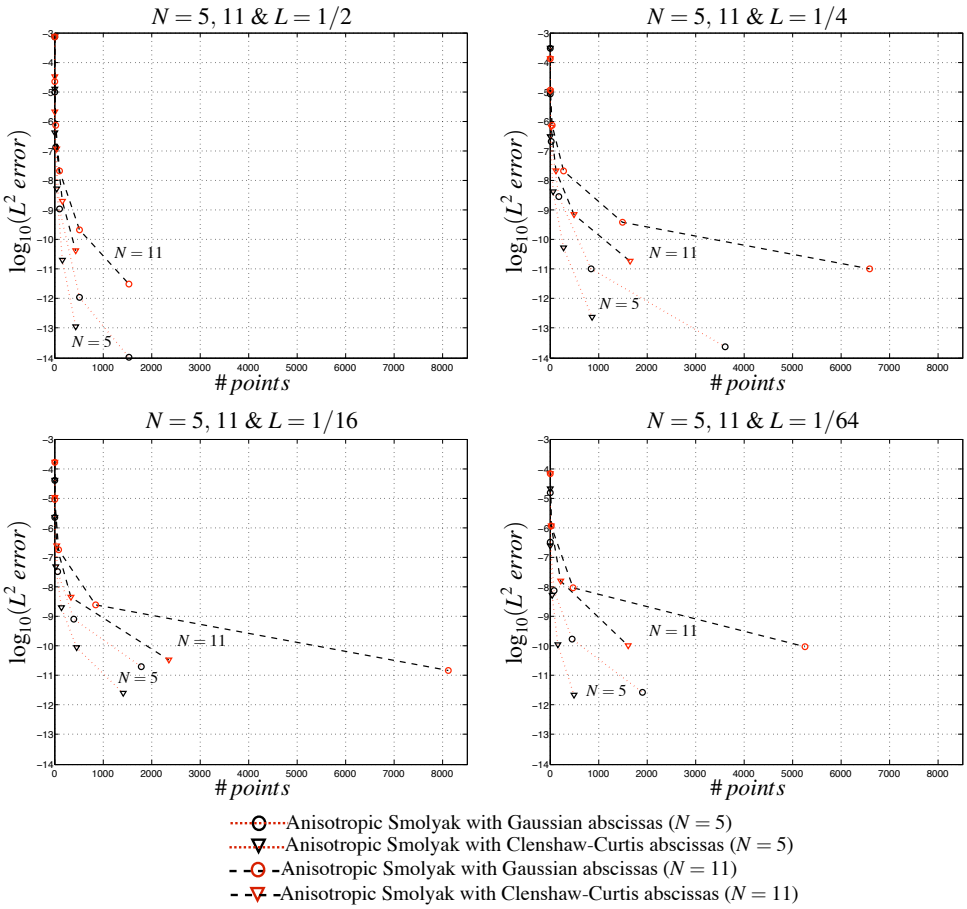

Figure 4: The rates of convergence of the anisotropic Smolyak algorithm for solving problem (6.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (6.7).

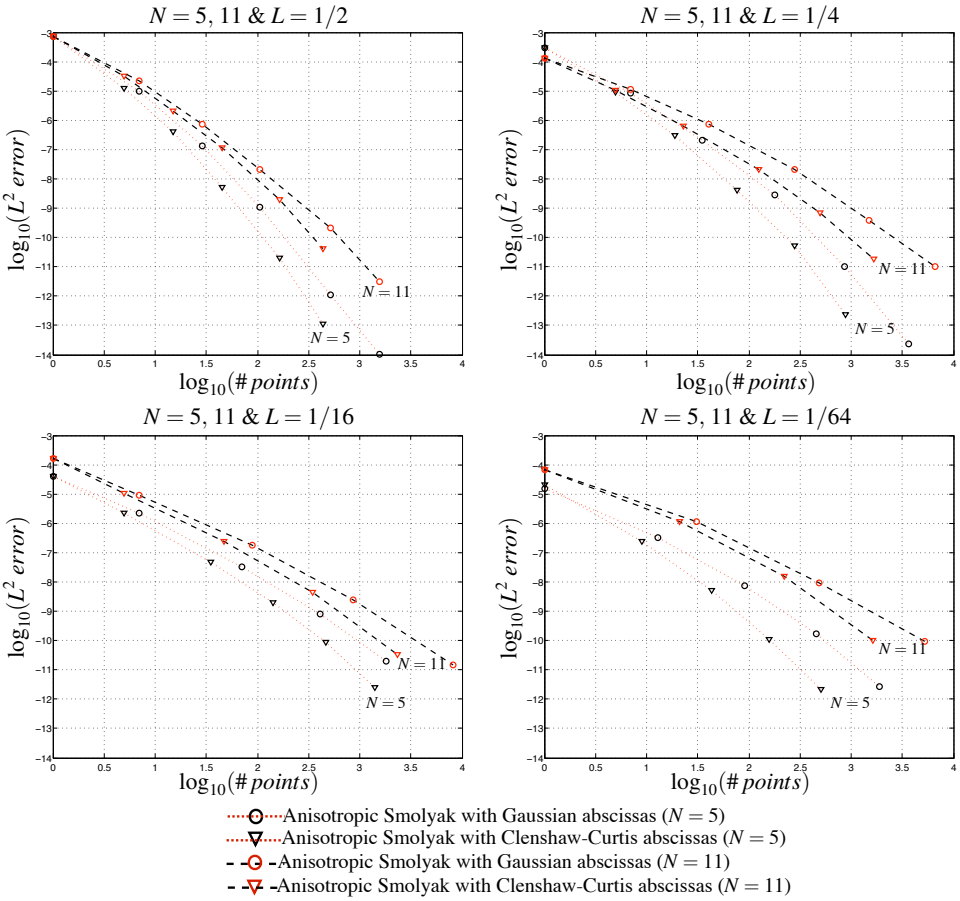

Figure 5: The convergence of the anisotropic Smolyak algorithm for solving problem (6.1) with given correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$ using both the Gaussian and Clenshaw-Curtis abscissas. For a finite dimensional probability space Γ^N with $N = 5$ and $N = 11$ we plot $\log(\epsilon)$ versus the logarithm of the number of collocation points. The $L^2(D)$ approximation error in the expected value for the anisotropic sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[\mathscr{A}_{\boldsymbol{\alpha}}(w, N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w}, N)\pi_h u_N]\|_{L^2(D)}$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (6.7).

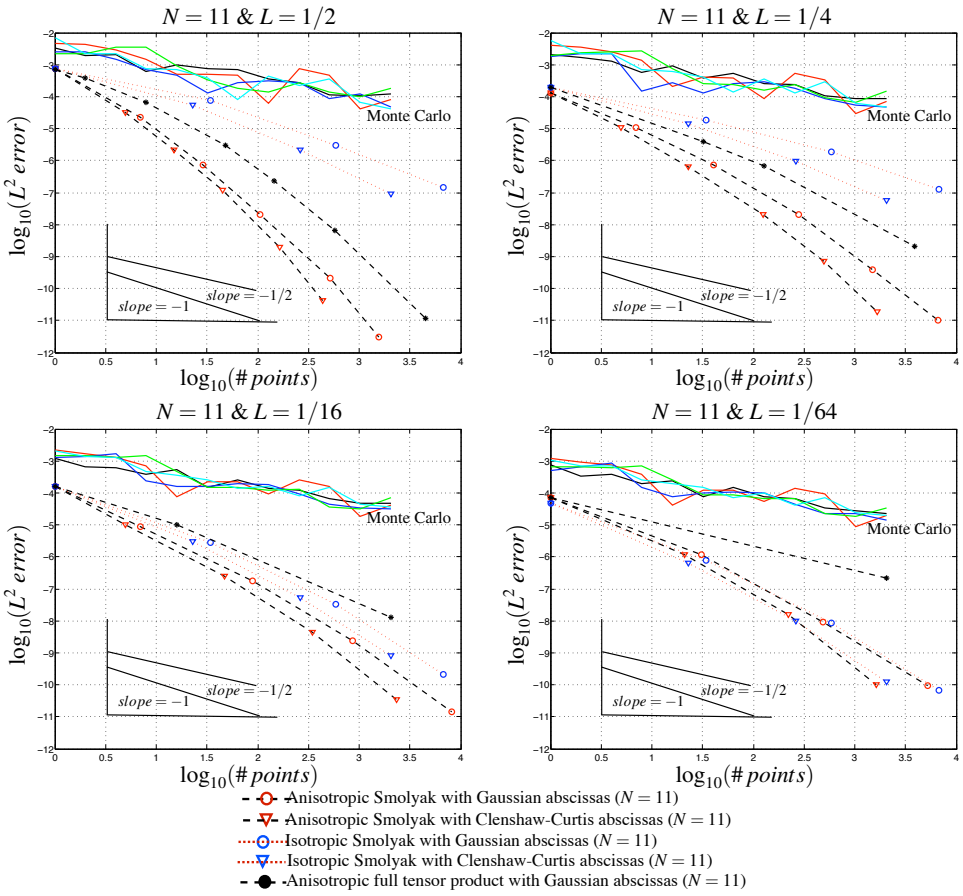

Figure 6: A 11-dimensional comparison of the anisotropic Smolyak method, the isotropic Smolyak method, the anisotropic full tensor product algorithm and Monte Carlo approach for solving problem (6.1) with correlation lengths $L_c = 1/2, 1/4, 1/16$ and $1/64$. The $L^2(D)$ approximation error in the expected value for the sparse collocation methods is given by: $\|E[\epsilon]\|_{L^2(D)} \approx E[\mathscr{A}_{\boldsymbol{\alpha}}(w,N)\pi_h u_N - \mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]$ where $w = 0, 1, \ldots, \overline{w}$ and $\widehat{\boldsymbol{\alpha}}$ is defined by (6.7). The $L^2(D)$ approximation error in the expected value for the anisotropic full tensor product method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|E[u^N_{h,\mathbf{p}} - u^N_{h,\widetilde{\mathbf{p}}}]\|_{L^2(D)}$, where $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ and $\widetilde{\mathbf{p}} = (p_1 + 1, p_2 + 1, \ldots, p_N + 1)$. The $L^2(D)$ approximation error in the expected value for the Monte Carlo method is given by: $\|E[\epsilon]\|_{L^2(D)} \approx \|\overline{E}[u^N_{h,k}; M] - E[\mathscr{A}_{\widehat{\boldsymbol{\alpha}}}(\overline{w},N)\pi_h u_N]\|_{L^2(D)}$, where $M = 2^i$, $i = 0, 1, 2, \ldots, 10$.

7 Conclusions

This work proposed and analyzed a novel weighted Smolyak method, describing an optimal choice for the weight parameters. These weights adaptively tune the anisotropy of the method for each given problem. Their systematic choice can be based both on *a priori* or *a posteriori* information and is motivated by the regularity of the solution and the error estimates derived in this work.

The method proposed here can be viewed as a natural extension of our previous works; the full tensor Stochastic Collocation method [1] and the isotropic Sparse Grid Stochastic Collocation method [25].

The new technique consists of a Galerkin approximation in physical space and an anisotropic collocation in probability space, at the zeros of sparse tensor product polynomials utilizing either Clenshaw-Curtis or Gaussian knots. As a consequence of the collocation approach our techniques naturally lead to the solution of uncoupled deterministic problems that are trivially parallelizable, as in the Monte Carlo method.

The error estimates derived in this work predict a rate of convergence that is at least algebraic with respect to the total number of collocation points. Similarly to the Monte Carlo method, the number of collocation points is directly proportional to the computational work required by the algorithm. Furthermore, the derived estimates provide a glimpse into future directions in this area. We observe that for problems possessing coefficients that are truncated expansions with proper decay in each probability direction, there may not be a deterioration in the exponent of the algebraic convergence with respect to the truncation dimension. This is a substantial advantage when comparing to the previous works [1] and [25].

The numerical examples included in this work gave computational ground to the theoretical results and suggest that the actual convergence rate of the method may be superior to the prediction yielded by the current error estimates. The numerical results include a comparison of the new weighted Smolyak method with the methods from [1], [25] and the standard Monte Carlo. In particular, for moderately large dimensional problems, the anisotropic sparse grid approach seems to be very efficient and superior to all examined methods.

Acknowledgments

The first author was partially supported by M.U.R.S.T. Cofin 2005 "Numerical Modeling for Scientific Computing and Advanced Applications". The second author wants to acknowledge the support of the Dahlquist fellowship at the Royal Institute of Technology, Stockholm and the support of UdelaR in Uruguay. The first and second authors were partially supported by the SANDIA project # 523695. The third author was supported by the School of Computation Science (SCS) at Florida State University and would like to thank Prof. Max Gunzburger

for his insight, guidance and many helpful discussions. Finally, the third author would like to thank Dr. John Burkardt for his assistance in implementing these numerical methods as well as his many useful visualizations.

References

[1] I. M. Babuška, R. Tempone, and F. Nobile. A stochastic collocation method for elliptic partial differential equations with random input data. Technical report, MOX, Dipartimento di Matematica, 2005.

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

[3] I. M. Babuška, R. Tempone, and G. E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1251–1294, 2005.

[4] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12:273–288, 2000.

[5] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer–Verlag, New York, 1994.

[6] J. Burkardt, M. Gunzburger, and C. Webster. Reduced order modeling of some nonlinear stochastic partial differential equations. *International Journal of Numerical Analysis and Modeling*, 2006. To appear.

[7] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, New York, 1978.

[8] C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numer. Math.*, 2:197–205, 1960.

[9] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[10] V. K. Dzjadyk and V. V. Ivanov. On asymptotics and estimates for the uniform norms of the lagrange interpolation polynomials corresponding to the chebyshev nodal points. *Analysis Mathematica*, 9(11):85–97, 1983.

[11] G.S. Fishman. *Monte Carlo*. Springer Series in Operations Research. Springer-Verlag, New York, 1996. Concepts, algorithms, and applications.

[12] P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):205–228, 2005.

[13] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. Technical report, Cornell, 2006.

[14] J. Garcke and M. Griebel. Classification with anisotropic sparse grids using simplicial basis functions. *Intelligent Data Analysis*, 6:483–502, 2002.

[15] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numer. Algorithms*, 18(3-4):209–232, 1998.

[16] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71:65–87, 2003.

[17] R. Ghanem. Ingredients for a general purpose stochastic finite elements implementation. *Comput. Methods Appl. Mech. Engrg.*, 168(1-4):19–34, 1999.

[18] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[19] J.C. Helton and F.J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, 81:23–69, 2003.

[20] O. P. Le Maître, O. M. Knio, H. N. Najm, and R. G. Ghanem. Uncertainty propagation using Wiener-Haar expansions. *J. Comput. Phys.*, 197(1):28–57, 2004.

[21] M. Loève. *Probability theory*. Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45 and 46.

[22] L. Mathelin, M. Y. Hussaini, and T. A. Zang. Stochastic approaches to uncertainty quantification in CFD simulations. *Numer. Algorithms*, 38(1-3):209–236, 2005.

[23] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1295–1331, 2005.

[24] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[25] F. Nobile, R. Tempone, and C. Webster. A sparse grid stochastic collocation method for elliptic partial differential equations with random input data.

Technical report, Technical Report #85, MOX, Dipartimento di Matematica, 2006. Submitted to SIAM J. Numer. Anal. (May 2006).

[26] B. Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[27] F. Riesz and B. Sz-Nagy. *Functional Analysis*. Dover, 1990.

[28] L.J. Roman and M. Sarkis. Stochastic galerkin method for elliptic SPDES: a white noise approach. *DCDS-B Journal*, 6(4):941–955, 2006.

[29] R. A. Todor. *Sparse Perturbation Algorithms for Elliptic PDE's with Stochastic Data*. PhD thesis, Dipl. Math. University of Bucharest, 2005.

[30] L. N. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? SIAM Review, to appear, 2006.

[31] G. W. Wasilkowski and H. Woźniakowski. Explicit cost bounds of algorithms for multivariate tensor product problems. *Journal of Complexity*, 11:1–56, 1995.

[32] D. Xiu and J.S. Hesthaven. High order collocation methods for differential equations with random inputs. submitted to SIAM J. Sci. Comput.

[33] D. Xiu and G. E. Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Comput. Methods Appl. Mech. Engrg.*, 191(43):4927–4948, 2002.