A multilevel sparse kernel-based stochastic collocation finite element method for elliptic problems with random coefficients

Zhaonan Dong[a], Emmanuil H. Georgoulis[a,b], Jeremy Levesley[a], Fuat Usta[a]

[a]*Department of Mathematics, University of Leicester, Leicester, United Kingdom*
[b]*Department of Mathematics, School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Zografou 15780, Greece*

Abstract

A new stochastic collocation finite element method is proposed for the numerical solution of elliptic boundary value problems (BVP) with random coefficients, assuming that the randomness is well-approximated by a finite number of random variables with given probability distributions. The proposed method consists of a finite element approximation in physical space, along with a stochastic collocation quadrature approach utilizing the recent *Multilevel Sparse Kernel-Based Interpolation (MuSIK)* technique (Georgoulis, Levesley & Subhan, *SIAM J. Sci. Comput.*, 35(2), pp. A815–A831, 2013). MuSIK is based on a multilevel sparse grid-type algorithm with the basis functions consisting of directionally anisotropic Gaussian radial basis functions (kernels) placed at *directionally-uniform* grid-points. We prove that MuSIK is interpolatory at these nodes, and, therefore, can be naturally used to define a quadrature scheme. Numerical examples are also presented, assessing the performance of the new algorithm in the context of high-dimensional stochastic collocation finite element methods.

Keywords: Stochastic collocation, PDE with random data, sparse grids, radial basis functions, high dimensional approximation, interpolation.

1. Introduction

Uncertainty is often embedded in the mathematical modelling of physical systems via introducing randomness in the coefficients of continuum models involving partial differential equations (PDEs). Uncertainty may come from incompleteness of understanding the nature and/or the difficulty of accurately modelling a phenomenon. Several approaches have appeared in addressing these challenges, such as worst-case scenario analysis (see, e.g., [3]), or via the introduction of various probabilistic expansions. Popular families of methods of the latter kind are so-called the polynomial chaos expansions [36, 21, 49, 30] and the Karhunen-Loève expansions [40, 41, 5, 23, 6, 4, 20, 44, 45, 51, 34] for parametrisation of the random fields. In this work, we focus on elliptic PDEs with random diffusion and load coefficients which are modelled as random fields with given spatial correlation structure. Typically, numerical methods for this class of problems seek to predict statistical moments of the numerical solution (such as mean, variance, etc.) with a given probability distribution of random field which, if unknown, is approximated by a Karhunen-Loève-type expansion [40, 41] as a combination of a finite number of random variables. This is done either by resorting to high-dimensional deterministic settings, see, e.g., [5, 23, 6, 4, 20, 44, 45, 29, 11, 10, 19] and the references therein, or using quasi Monte-Carlo or Multi-level Monte-Carlo techniques, see, e.g., [7, 15, 38, 18, 39, 31, 52, 28] and the references therein.

This work is concerned with a new stochastic collocation finite element method, whereby the collocation step is based on a novel sparse-type quadrature scheme utilizing the recent *Multilevel Sparse Kernel-Based Interpolation (MuSIK)* technique [27]. To focus on this novel quadrature, we consider a fairly standard setting for the finite element discretization and noise approximation in the context of stochastic collocation. We believe, nevertheless, that the new collocation step is also applicable in more elaborate stochastic collocation approaches. We also mention [56], which contains an interesting computational verification of the proposed approach.

MuSIK is a low complexity interpolation scheme, constructed via a sparse grid-type combination technique [22, 33], with the basis functions consisting of directionally anisotropic Gaussian kernels (radial basis functions) centred

at *uniform* grid-points in each spatial direction. The use of Gaussian kernels, instead of polynomials (as typically done in the literature [4, 44, 45],) enables, crucially, the use of directionally uniform grids. Therefore, the proposed MuSIK-type quadrature is both *multilevel* and *hierarchical* with respect to the quadrature nodes between consecutive levels. This means that improvement of the quadrature accuracy is done in a "multiresolution" fashion, whereby the next level improves only on higher frequencies than the previous levels, i.e., adding "details". Hence, if higher accuracy is found to be needed after the (typically lengthy) computation, more levels are added to further improve the accuracy without the need to discard and restart the computation.

To ensure the suitability of MuSIK in the context of numerical quadrature, we prove that MuSIK is, indeed, an interpolant at the directionally uniform (hierarchical) sparse-grid nodes. We stress, however, that, contrary to the standard polynomial or piecewise polynomial/spline sparse-grid interpolants [14, 4, 25], the respective MuSIK basis functions do *not* form nested linear spaces at each level. Indeed, we show that the nestedness of the nodes along with tensor-type basis functions is sufficient for a sparse-grid/combination-type scheme to be interpolatory. This result could be of independent interest.

Another key attribute of the proposed approach is the unified treatment of random variables with both bounded and unbounded ranges. Indeed, as MuSIK employs globally supported Gaussian kernels, they are suitable to approximate also unbounded random variables (e.g., Gaussian random variables) upon suitable selection of the kernel center locations. In contrast, polynomial stochastic collocation approaches employ special polynomial expansions (e.g., Hermite polynomial expansions) to treat Gaussian random variables, which are, crucially, not hierarchical. In this context, the proposed stochastic collocation based on MuSIK, which is naturally hierarchical by construction renders itself as an attractive alternative.

The probability space is determined by the random variables retained in the truncated Karhunen-Loève expansions [40, 41]. The proposed method is tested numerically for an elliptic problem with up to 21 dimensions in probability space, and for both bounded and unbounded random variables. The numerical results, highlight the good performance of the proposed method in terms of accuracy versus degrees of freedom and dimension.

The remainder of this work is organized as follows. In Section 2, the model problem is introduced, along with the stochastic collocation FEM framework. In Section 3, the MuSIK is reviewed and we also prove that MuSIK scheme is interpolatory on the grid points, while in Section 4 the MuSIK-based stochastic collocation FEM is introduced. Numerical examples are given in Section 5, while some conclusions and further directions of research are discussed in Section 6.

2. Problem Setting

2.1. Model problem

Let $D \subset \mathbb{R}^d$ be an open polygonal domain, henceforth termed as the *physical domain*. Let also $(\Omega, \mathcal{F}, P)$ be a complete probability space, with Ω being the set of outcomes, $\mathcal{F} \subset 2^\Omega$ the corresponding σ-algebra of events, and $P : \mathcal{F} \to [0, 1]$ the probability measure. We consider the stochastic elliptic boundary value problem: find the random field, $u : \Omega \times \overline{D} \to \mathbb{R}$, such that

$$-\nabla \cdot (a(\omega, x)\nabla u(\omega, x)) = f(\omega, x) \quad \text{for } x \in D, \qquad u(\omega, x) = h(x) \quad \text{for } x \in \partial D, \tag{1}$$

$\omega \in \Omega$, P-almost everywhere in Ω, with a denoting the diffusivity coefficient, whose detailed properties are given below and $h \in H^{1/2}(\partial D) := \{v \in L^2(D) : \|v\|_{H^{1/2}(\partial D)} < +\infty\}$ where $\|v\|_{H^{1/2}(\partial D)} := \inf_{v=\gamma w} \|w\|_{H^1(D)}$ and γ denotes the classical Sobolev trace operator. We stress, however, that more general elliptic problems involving lower order terms and mixed boundary conditions are by all means possible to be included in the present framework. We refrain from doing so in the interest of simplicity of the presentation.

To ensure ellipticity, we assume that

$$\exists a_{\min}, a_{\max} \in (0, \infty), \text{ s.t. } P(\omega \in \Omega : a(\omega, x) \in [a_{\min}, a_{\max}] \, \forall x \in \overline{D}) = 1, \tag{2}$$

i.e., $a(\omega, \cdot)$ is uniformly bounded and coercive. Also, we introduce the Hilbert spaces

$$V_{P,a} := \{v \in L^2_P(\Omega) \otimes H^1(D) : \int_D E_P[a|\nabla v|^2]dx < \infty\},$$

with respective norm $\|v\|_{P,a} := (\int_D E_P[a|\nabla v|^2 + |v|^2]dx)^{1/2}$, where E_P denotes the P-mean of a random field. From the above assumptions, we can see that $V_{P,a} \subset V_{P,1}$, and that $\|v\|_{P,1} \leq C(a_{\min}, a_{\max})\|v\|_{P,a}$. The weak form of the problem (1) then reads: there exists a unique $u \in V_{P,a}$, such that, for all $v \in L_P^2(\Omega) \otimes H_0^1(D)$, we have

$$\int_D E_P[a\nabla u \cdot \nabla v]\mathrm{d}x = \int_D E_P[fv]\mathrm{d}x.$$

The well-posedness of (1) follows from the Lax-Milgram lemma, by assuming $\int_D E_P[f^2]\mathrm{d}x < \infty$. Then, the problem has a unique solution $u \in V_{P,a}$. For brevity, we also set $E \equiv E_P$.

2.2. Finite dimensional noise assumption

A central question for this class of PDE problems is the parametrisation of the randomness in the diffusion a and in the forcing term f. This has been studied in a number of works (see, e.g., [5, 42, 23] and the references therein). When randomness can be completely described through a finite number of random variables of known probability densities, the original (stochastic) problem (1) can be transformed into a deterministic one. When randomness is not completely defined through such a finite number of random variables, the predominant way forward in the literature has been to represent the random coefficients as expansions of infinitely many random variables and truncate these expansions.

A widely used tool in this context is the Karhunen-Loève expansion [40, 41], which can be used to represent the random diffusion a and load f coefficients as an expansion of infinite number of random variables, whose careful truncation after N terms, say, results in approximations a_N and f_N, defined through a finite family of known random variables.

More specifically, let g be a second order random field, with covariance $\mathrm{cov}[g] : \overline{D} \times \overline{D} \to \mathbb{R}$. We introduce the compact and self-adjoint operator $T_g : L^2(D) \to L^2(D)$ by

$$T_g v(\cdot) := \int_D \mathrm{cov}[g](x, \cdot)v(x)\,\mathrm{d}x , \quad v \in L^2(D),$$

which admits a sequence of non-negative, decreasing eigenvalues $\{\lambda_k\}_{k=1}^\infty$. The corresponding sequence of orthonormal eigenfunctions is denoted by $\{\phi_k\}_{k=1}^\infty$.

Using the above orthonormal set, we define a family of mutually uncorrelated real random variables, given by

$$Y_k(\omega) := \frac{1}{\lambda_k} \int_D (g(w, x) - E[g](x))\phi_k(x)\,\mathrm{d}x, \tag{3}$$

$k \in \mathbb{N}$, with $\mathrm{E}[Y_k] = 0$ and $\mathrm{E}[Y_i Y_j] = \delta_{ij}$ for $i, j, k \in \mathbb{N}$. We can now further define the N-term partial sum of the Karhunen-Loève expansion g_N of g, by

$$g_N(\omega, x) := E[g](x) + \sum_{k=1}^N \sqrt{\lambda_k}\phi_k(x)Y_k(\omega), \tag{4}$$

P-a.e. in Ω, with $N \in \mathbb{N}$. Mercer's theorem [41] then implies

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

We stress that the practical construction of eigenpairs (λ_k, ϕ_k), along with the study of the eigenvalue decay are involved topics, which we shall refrain from exploring in detail in this work; instead, we refer to [16, 53].

For the remainder of this work, we shall assume that we have at our disposal known approximations a_N and f_N of the random diffusion a and load f random fields, respectively. We shall be seeking to find a random field $u_N : \Omega \times \overline{D} \to \mathbb{R}$, such that

$$-\nabla \cdot (a_N(\omega, x)\nabla u_N(\omega, x)) = f_N(\omega, x) \quad \text{for } x \in D, \quad u_N(\omega, x) = g(x) \quad \text{for } x \in \partial D, \tag{5}$$

3

P-almost everywhere in Ω, such that $u_N \in V_{P,a}$ is unique and is described by a set of N random variables $[Y_1, \ldots, Y_N]$, viz., $u_N(\omega, x) \equiv u_N(Y_1(\omega), \ldots, Y_N(\omega), x)$.

As the focus of this work is in proposing a new framework for sparse high-dimensional stochastic collocation procedures, we shall make use of the Karhunen-Loève expansion for the $\log(a - a_{\min})$ under the assumption that $a > a_{\min} \forall x \in D$, almost surely in Ω to guarantee the diffusivity coefficient is bounded away from zero [45, 44], thereby constructing an a_N as in (4), with

$$\log(a_N - a_{\min})(\omega, x) = \phi_0(x) + \sum_{1 \leq n \leq N} \sqrt{\lambda_n} \phi_n(x) Y_n(\omega), \tag{6}$$

with $\phi_0(x)$ denoting the mean. For conditions on the well-posedness of (5), we refer to [2, 24]. We stress, however, that the new stochastic collocation technique presented below is independent of the expansion used for a. The right hand-side of (1) is also assumed to be well approximated by a truncated Karhunen-Loève expansion:

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

The diffusion tensor a and forcing term f are often independent. In such a situation, the truncated Karhunen-Loève expansions of a and f will depend on a distinct subset of $[Y_1, \ldots, Y_N]$ each. Nonetheless, we retain the same index N for both in the interest of notational simplicity.

Denoting by $\Gamma_k(\Omega)$ the images of the random variables Y_k (which may be unbounded), we set $\Gamma^N = \prod_{n=1}^N \Gamma_n(\Omega)$. On Γ^N, we define $\rho : \Gamma^N \to \mathbb{R}_+$, with $\rho \in L^\infty(\Gamma^N)$ to be the joint probability density function of the random variables $[Y_1, \ldots, Y_N]$, and we denote by $L^2_\rho(\Gamma^N)$ the ρ-weighted L^2-space over Γ^N.

Hence, (5) can be transformed to the deterministic problem: find $u_N \in L^2_\rho(\Gamma^N) \otimes H^1(D)$, such that

$$\int_{\Gamma^N} \rho(y) \int_D a_N \nabla u_N \cdot \nabla v \, \mathrm{d}x \mathrm{d}y = \int_{\Gamma^N} \rho(y) \int_D f_N v \, \mathrm{d}x \mathrm{d}y, \tag{7}$$

for all $v \in L^2_\rho(\Gamma^N) \otimes H^1_0(D)$. The proof of existence and uniqueness for the first two statistical moments can be found in [4]. Also, in [17] the discretization of different statistical moments for nonlinear parametric operator equations is discussed.

2.3. Stochastic collocation FEM

The next step is to design a numerical method to approximate the solution to (7), taking into account that $\dim(\Gamma^N)$ can be large.

To this end, we consider a standard finite element space $V_h \subset H^1_0(D)$ of degree p, with total dimension N_h, defined on a regular triangulation $\mathcal{T}_h$ with maximum mesh-diameter $h > 0$ (see, e.g., [12]).

To arrive at a practical numerical method, one needs to approximate the integral over the probability space also. The stochastic collocation and related approaches [5, 23, 6, 4, 20, 44, 45] proceed by approximating the integral via carefully selected quadrature rules for high dimensional problems. A typical such quadrature rule is of the form

$$\int_{\Gamma^N} G(y) \mathrm{d}y \approx \sum_{q \in \mathcal{N}} w_q G(y_q),$$

for some integrable integrand G, where $\mathcal{N}$ is an index set, $\{y_q : q \in \mathcal{N}\}$ is the set of abscissae and $\{w_q : q \in \mathcal{N}\}$ the set of weights, respectively. Such quadrature rules can arise from considering a corresponding interpolation problem for the integrand G on $\{y_q : q \in \mathcal{N}\}$.

Indeed, let $\{\theta_q : q \in \mathcal{N}\}$ be a Lagrange basis for the class of interpolants S at the points $\{y_q : q \in \mathcal{N}\}$. The interpolation problem for the integrand seeks a function S from the space spanned by $\{\theta_q : q \in \mathcal{N}\}$, such that $G(y_q) = S(y_q)$ for all $q \in \mathcal{N}$. Hence, we have

$$S(y) = \sum_{q \in \mathcal{N}} G(y_q) \theta_q(y).$$

4

Therefore, going back to the integral over Γ^N, we consider the approximation

$$\int_{\Gamma^N} G(y)\mathrm{d}y \approx \int_{\Gamma^N} S(y)\mathrm{d}y = \sum_{q \in \mathcal{N}} w_q S(y_q) = \sum_{q \in \mathcal{N}} w_q G(y_q), \tag{8}$$

with

$$w_q := \int_{\Gamma^N} \theta_q(y)\,\mathrm{d}y. \tag{9}$$

Using this quadrature formula on both sides of (7), we deduce formally

$$\sum_{q \in \mathcal{N}} w_q \rho(y_q) \int_D \Big(a_N(y_q, x)\nabla u_N(y_q, x) \cdot \nabla v(x) - f_N(y_q, x)v(x) \Big)\mathrm{d}x \approx 0. \tag{10}$$

These formal considerations give rise to the stochastic collocation finite element method, whereby we require (10) to hold with strict equality. This is satisfied upon finding $u_N^{h,q} \in V_h$ approximations to $u_N(y_q, \cdot)$, such that

$$\int_D a_N(y_q, x)\nabla u_N^{h,q}(x) \cdot \nabla v(x)\,\mathrm{d}x = \int_D f_N(y_q, x)v(x)\,\mathrm{d}x \qquad \forall v \in V_h,$$

for each $q \in \mathcal{N}$; the respective approximate solution to (7) is, thus, given by

$$u_N^h(y, x) := \sum_{q \in \mathcal{N}} u_N^{h,q}(x)\theta_q(y).$$

A crucial numerical challenge in this approach is the cardinality of $\mathcal{N}$. If we use tensor-product-type quadrature rules based on univariate quadratures with r points in each direction, say, then the cardinality of $\mathcal{N}$ will be r^N, i.e., the cardinality of the finite element problems to be solved grows *exponentially* with the dimension of Γ^N. This is a manifestation of what is typically referred to in the literature as *curse of dimensionality*. To address this, reduced-complexity quadrature rules have been employed in the literature. Prominent methods in this setting are the Smolyak-type quadratures [50], based on sparse tensor products of Gauss-Lobatto or Clenshaw-Curtis univariate rules [4, 45, 44], for bounded Γ^N. The sparse quadrature node distribution of Gauss-type rules is non-hierarchical as the number of nodes increases in each direction, while Clenshaw-Curtis based rules are hierarchical upon doubling of the nodes in each direction. For unbounded Γ^N, Smolyak-type quadratures based on univariate Hermite polynomials have been typically used [4]. The latter do *not* give rise to hierarchical nodes. If more resolution is found to be required for aforementioned quadrature rules with non-hierarchical node distributions, one has to recompute essentially a completely different set of nodal values. Thus, within an adaptive quadrature stochastic collocation setting, Clenshaw-Curtis type methods are preferred for bounded Γ^N in the literature [44]; for unbounded Γ^N, the nested Gauss-Hermite rule in [26, 35] is preferred.

In the next section, we shall introduce a new sparse quadrature rule based on the recent multilevel sparse kernel-based interpolation (MuSIK) method [27], which uses Gaussian kernels instead of polynomials and can be applied to *both* bounded and unbounded Γ^N. An attractive attribute of the MuSIK approach is its ability to provide stable interpolation on *hierarchical directionally-uniform interpolation nodes*. In particular, MuSIK is able to interpolate in a stable fashion on the standard sparse-grid node sequences for piecewise linear sparse interpolation (cf., Figure 3 for an illustration of two dimensional node distributions at different levels.) Crucially, these sparse grid node distributions are *hierarchical* between two consecutive levels of refinement and are *uniform in each axiparallel direction*.

This hierarchical configuration of the interpolation nodes is combined within a multilevel iteration, whereby a more accurate interpolant is computed by adding "detail" to the previous one. Recalling that standard Gaussian kernels are able to interpolate at an exponential rate of convergence [55], such good performance is not counter-intuitive or unexpected.

3. Multilevel sparse kernel-based interpolation

3.1. Review of MuSIK

Here we review the multilevel sparse kernel-based interpolation method (MuSIK) introduced in [27] (where it was abbreviated as MLSKI). We point out that the proposed MuSIK algorithm in this work is different from the original

MuSIK algorithm, and is more suitable for high-dimensional quadrature (cf. Remark 3). Moreover, we shall prove that the MuSIK method is, indeed, interpolatory at the respective nodes.

Definition 1. *A continuous function* $\Phi : \mathbb{R}^N \to \mathbb{R}$ *is said to be a* positive definite kernel *if* $\sum_{i,j=1}^{M} c_i c_j \Phi(\mathbf{x}_i - \mathbf{x}_j) > 0$ *for any M pairwise different points $\mathbf{x}_i$, $i = 1, \dots, M$ and for any $\mathbf{c} := [c_1, \dots, c_M]^T$ with $\mathbf{c} \neq \mathbf{0}$. Moreover, Φ is said to be* radial, *if there exists a function $\phi : [0, \infty) \to \mathbb{R}$ such that $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|)$ for all $\mathbf{x} \in \mathbb{R}^N$, for some norm $\|\cdot\|$ of $\mathbb{R}^N$. Finally, a radial positive definite kernel Φ will be referred to as* radial basis function *(RBF)*.

Examples of widely used positive definite RBFs are the Gaussians with $\phi(r) := \exp(-c^2 r^2)$, inverse multiquadrics $\phi(r) := 1/\sqrt{c^2 + r^2}$, and the family of compactly supported RBFs [54]. We shall focus on Gaussian kernels in this work, due to their spectral accuracy in conjunction with their tensor product nature: multivariate Gaussians are tensor products of univariate Gaussian kernels. The interpolation of data with anisotropic distribution of data sites in the domain should be considered specially.

Definition 2. *Given an RBF $\varphi(\|\cdot - \mathbf{x}_i\|)$ centred at $\mathbf{x}_i \in \mathbb{R}^N$ and an invertible matrix $A \in \mathbb{R}^{N \times N}$, we define the anisotropic RBF φ_A by $\varphi_A(\|\cdot - \mathbf{x}_i\|) = \varphi(\|A(\cdot - \mathbf{x}_i)\|)$.*

For any given data sites $X := \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ contained in a bounded domain $\Gamma \subset \mathbb{R}^N$, we consider the interpolation data $\{(\mathbf{x}_i, z_i) : \mathbf{x}_i \in X, i = 1, \dots, M\}$. Then, for φ positive-definite radial function, the anisotropic RBF interpolant S_A is given by

$$S_A(\mathbf{x}) = \sum_{i=1}^{M} c_i \varphi_A(\|\mathbf{x} - \mathbf{x}_i\|), \quad \mathbf{x} \in \Gamma;$$

the c_i are chosen to satisfy the interpolation condition $S_A(\mathbf{x}_i) = z_i$, $i = 1, \dots, M$. The invertibility of the scaling matrix A guarantees the well-posedness of the interpolation problem for positive definite kernels [9].

To address the computational challenge posed by high dimensionality, the MuSIK is constructed as a multilevel interpolation method on sparse grids in the sense of [14], whose node cardinality grows favourably with dimension N; see Figure 3 for an illustration of the node distribution in standard sparse grids in two dimensions. To ensure the well-posedness of the interpolation problem, a multivariate Boolean interpolation procedure [22], also known as combination technique in sparse grid terminology [33], is employed.

The idea behind such classical combination/extrapolation procedures is to construct an interpolant on sparse grids as a linear combination of smaller interpolation sub-problems. To illustrate this, we refer to Figure 1, depicting a sparse grid as the union of *directionally uniform* partial grids. In the present context, anisotropically scaled Gaussians are employed by the proposed MuSIK method as interpolants of each partial grid; the anisotropic scaling used is such to make the mapped interpolation nodes to be globally uniform on the mapped domain. The use of Gaussian kernels, although not crucial for the MuSIK method [27], is preferable as their tensor-product nature results to the combination technique being interpolatory at the nodes; this is proven in Theorem 5 below.

Once each partial grid interpolant is computed, all such interpolants are linearly combined in an appropriate fashion (to be specified below) to form a *sparse kernel-based interpolant* (SIK). Due to the essentially stationary nature of the SIK method with Gaussians, convergence is accelerated via a multilevel procedure: the SIK is computed on the coarsest grid used in the computation, which is subsequently enriched by SIK computations of the residuals at the sparse grid nodes on the next level. The hierarchical nature of the sparse grids of different levels is particularly convenient at this point; cf., Figure 1.

We start by defining SIK on a bounded domain $\Gamma := [0, 1]^N$, to aid accessibility; the discussion on how these results can be extended to general, possibly unbounded, tensor-product domains Γ is postponed until Section 4.2. To this end, consider a function $u : \Gamma \to \mathbb{R}$ to be interpolated. For a multi-index $\boldsymbol{\ell} = (l_1, \dots, l_N) \in \mathbb{N}^N$, we define the family of directionally uniform grids $\{\mathbb{X}_{\boldsymbol{\ell}} : \boldsymbol{\ell} \in \mathbb{N}^N\}$ consisting of the nodes $\mathbf{x}_{\boldsymbol{\ell},\mathbf{i}} := (x_{l_1,i_1}, \dots, x_{l_N,i_N})$, with $x_{l_j,i_j} = i_j 2^{-(l_j-1)}$, for $i_j = 0, 1, \dots, 2^{l_j-1}$, $j = 1, \dots, N$, with the convention that $l_j = 1$ contains 1 node at the center of the jth direction for $j = 1, \dots, N$. We also define the vector of directional meshsizes of $\mathbb{X}_{\boldsymbol{\ell}}$ by $h_{\boldsymbol{\ell}} = 2^{-(\mathbf{l}-\mathbf{1})} := (2^{-(l_1-1)}, \dots, 2^{-(l_N-1)})$. Hence, if $h_{l_i} = 2^{-(k-1)}$, for all $i = 1, \cdots, N$, $\mathbb{X}_{\boldsymbol{\ell}}$ is the uniform *full grid* of level k; this will be denoted by $\mathbb{X}^{k,N}$. We consider the subset of $\mathbb{X}^{k,N}$, given by

$$\tilde{\mathbb{X}}^{k,N} := \bigcup_{|\boldsymbol{\ell}|_1 = k+N} \mathbb{X}_{\boldsymbol{\ell}}, \tag{11}$$

with $|\boldsymbol{\ell}|_1 := l_1 + \cdots + l_N$, which will be referred to as the *sparse grid of level k in N dimensions*, for $k = 0, 1, \ldots$. We refer to Figure 1 for an illustration of (11) for $k = 4$ and $N = 2$.
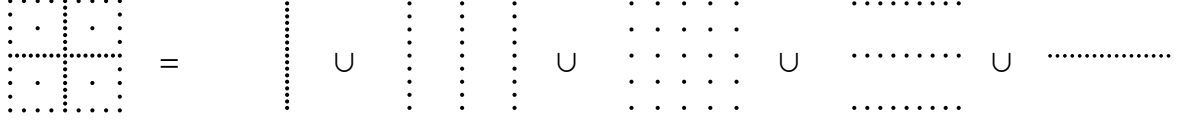

Figure 1: Sparse grid $\tilde{\mathbb{X}}^{4,2}$ via (11).

We want to evaluate the interpolant at the constituent partial grids $\mathbb{X}_\ell$. As the constituent grids admit different density in each coordinate direction, we shall make use of the anisotropic RBFs. To this end, for each multi-index $\boldsymbol{\ell} = (l_1, \ldots, l_N)$, we consider the transformation matrix $A_\ell \in \mathbb{R}^{N \times N}$ with

$$A_\ell := \operatorname{diag}(2^{l_1-1}, \ldots, 2^{l_N-1}).$$

The anisotropic RBF interpolant S_{A_ℓ} of u_N^h at the points of $\mathbb{X}_\ell$ is then given by

$$S_{A_\ell}(\mathbf{x}) := \sum_{j=1}^{M_\ell} c_j \varphi(\|A_\ell(\mathbf{x} - \mathbf{x}_j)\|), \tag{12}$$

for $\mathbf{x} \in \Gamma^N$, where M_ℓ is the cardinality of $\mathbb{X}_\ell$ and $c_j \in \mathbb{R}$ are chosen so that the interpolation conditions

$$S_{A_\ell}|_{\mathbb{X}_\ell} = u_N^h|_{\mathbb{X}_\ell},$$

are satisfied.

To construct the *sparse kernel-based interpolant* (SIK) $\tilde{S}_k$ on the sparse grid $\tilde{\mathbb{X}}^{k,N}$, the partial grid interpolants S_{A_ℓ} are linearly combined using the formula

$$\tilde{S}_k(\mathbf{x}) = \sum_{q=0}^{N-1} (-1)^q \binom{N-1}{q} \sum_{|\boldsymbol{\ell}|_1 = k+N-q} S_{A_\ell}(\mathbf{x}). \tag{13}$$

We refer to Figure 2 for an illustration when $N = 2$ and $k = 4$; the interpolant, then, reads

$$\tilde{S}_4(\mathbf{x}) = \sum_{|\boldsymbol{\ell}|_1 = 6} S_{A_\ell}(\mathbf{x}) - \sum_{|\boldsymbol{\ell}|_1 = 5} S_{A_\ell}(\mathbf{x}).$$
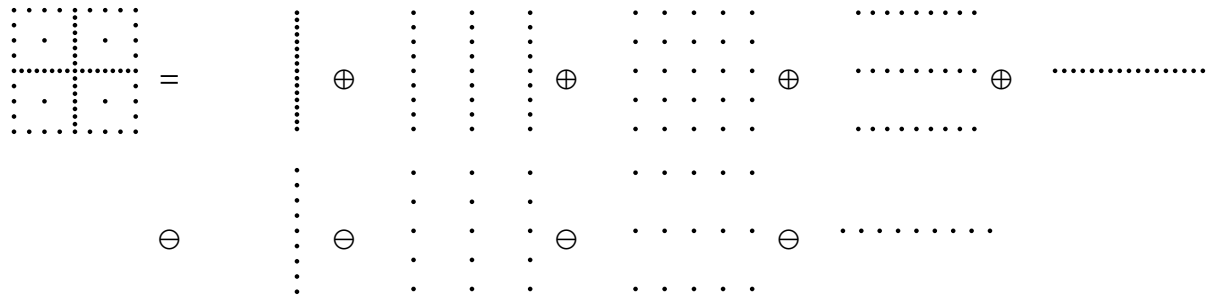

Figure 2: The construction of the sparse kernel-based $\tilde{S}_4$ interpolant on $\tilde{\mathbb{X}}^{4,2}$.

For more details on the combination formula (13), we refer to [22, 33, 32]. Moreover, in Section 3.3 below, we shall prove that (13) is an interpolant at the sparse grid nodes for the case of Gaussian kernels, or, indeed, any other basis function constructed as tensor-product of univariate basis functions.

It is important to note that the SIK can be implemented in a straightforward fashion using the existing *standard* RBF interpolation algorithms. Each of the $O(k^{N-1})$ such partial-grid constituent RBF problems can be solved independently. The size of each partial-grid problem is $O(2^k)$, where k is the number of levels; in particular, it is *independent* of the dimension N.

The next step is to consider a multilevel implementation of SIK [37, 43], in order to accelerate convergence and to overcome the essentially stationary nature of the SIK interpolant with Gaussians.
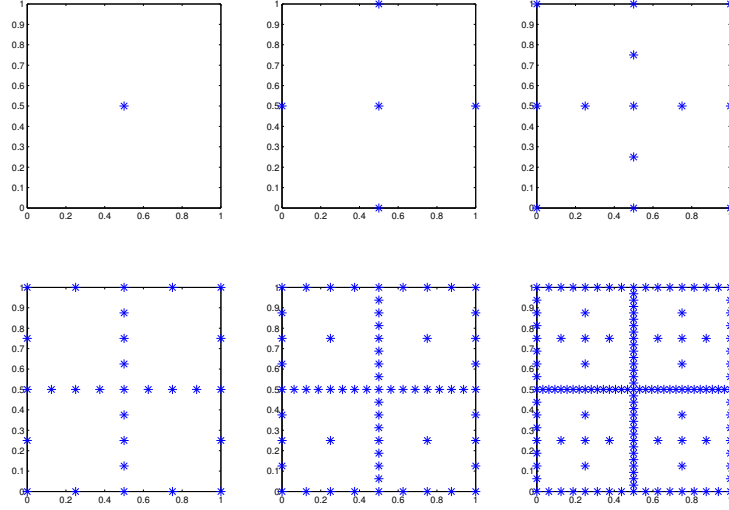

Figure 3: Nested-ness of sparse grids $\tilde{\mathbb{X}}^{k,2}$, $k = 0, 1, \ldots, 5$.

The MuSIK algorithm is initialized by computing the SIK $\tilde{S}_{k_0}$ at the coarsest designated sparse grid $\tilde{\mathbb{X}}^{k_0,N}$ and setting $\Delta_0 := \tilde{S}_{k_0}$. Then, for $j = 1, \ldots k$, we compute Δ_k to be the SIK of the residual $u_N^h - \sum_{i=0}^{k-1} \Delta_j$ on $\tilde{\mathbb{X}}^{k,N}$. The resulting MuSIK is then given by

$$\tilde{S}_k^{\mathrm{ML}} := \sum_{j=0}^{k} \Delta_j. \tag{14}$$

Remark 3. *The MuSIK method defined above is a variant to the original one presented in [27]. Indeed, as we shall be concerned with devising a quadrature formula below, it is possible to start the MuSIK at level 0, involving one point only; we refer to Figure 3 for an illustration. This means that the computational overhead is reduced significantly, allowing for more levels of the multi-level methods in high dimensional problems, than for the respective interpolation problem, leading to, potentially, greater accuracy.*

The numerical stability of RBF interpolation is, generally speaking, quite challenging due to the ill-conditioning of the respective interpolation matrices in standard bases; we refer to [47, 48, 55, 13] for detailed discussions on this issue. Nevertheless, in the MuSIK setting, the resulting partial-grid interpolation problems are typically sufficiently well conditioned for computations, due to the anisotropic scaling introduced [27]. Indeed, this anisotropic scaling exactly counteracts the anisotropy of the grid points of each partial-grid, resulting in quasi-uniform grids in the mapped domains, on which classical *radial* basis functions are employed. The situation is more favourable in the current setting of Gaussian kernels, as we shall see below, whereby the tensor product nature of the kernel can be utilised to accurately construct Lagrange (cardinal) bases efficiently.

8

3.2. Tensor products of univariate cardinal functions

Gaussian kernels on $\mathbb{R}^d$ can be viewed as tensor products of univariate Gaussians; this is evident also for the anisotropic versions of Gaussians. This crucial property of Gaussians will be instrumental in both the proof of interpolation of the MuSIK algorithm, and the development of fast procedures to evaluate the respective MuSIK interpolants.

We begin by considering the set of cardinal (also known as Lagrange) functions for each interpolation sub-problem in the SIK interpolant $S_{A_\ell}(\cdot)$ in (13), with the particular choice of Gaussian RBFs on the set $\mathbb{X}_\ell$. Let $\chi_{\ell,\mathbf{i}}$ be the cardinal function on the grid indexed by ℓ for the point $\mathbf{x}_{\ell,\mathbf{i}} \in \mathbb{X}_\ell$. Let also χ_{l_j,i_j} denote the univariate cardinal function in one variable for the node x_{l_j,i_j}, with respect to the set of nodes $\{x_{l_j,i_j} : i_j = 0, 1, \ldots, 2^{l_j-1}\}$. Note that χ_{l_j,i_j} is *not* necessarily cardinal on a grid of different level l_k, $k \neq j$. Hence, for $j = 1, \cdots, d$, there exist $b_{l_j,i_j} \in \mathbb{R}$, such that

$$\chi_{l_j,i_j}(y) = \sum_{i_j=0}^{2^{l_j-1}} b_{l_j,i_j} \exp(-c^2 h_{l_j}^2 (y - x_{l_j,i_j})^2),$$

for $y \in [0, 1]$. Then, we have

$$z(\mathbf{y}) := \prod_{j=1}^{d} \chi_{l_j,i_j}(y_j) = \sum_{\mathbf{x}_{\ell,\mathbf{i}} \in \mathbb{X}_\ell} b_{\ell,\mathbf{i}} \exp\left(-c^2 \sum_{j=i}^{d} h_{l_j}^2 (y_i - x_{l_j,i_j})^2\right),$$

where $b_{\ell,\mathbf{i}} = \prod_{j=1}^{d} b_{l_j,i_j}$. This is exactly the form of the cardinal function based on the points in the grid $\mathbb{X}_\ell$, which implies $z(\mathbf{y}) = \chi_{\ell,\mathbf{i}}$, due to the uniqueness of Gaussian interpolation.

Hence, it is possible to compute the cardinal functions for multivariate approximation by computing *ab initio* (to arbitrarily high precision, e.g., by using symbolic calculators) the cardinal functions for univariate approximation up to (for instance) $5, 9, 17, \cdots, 129$ equally spaced points, and store these. The approximation process would then require *no* solution of linear systems, thereby, *massively increasing the speed of the algorithm.*

3.3. Tensor product kernels give interpolatory schemes

The combination formula (13) for polynomial-based sparse grid schemes has been proven to be interpolatory [8]; crucially, the proof relies on the nestedness of polynomial spline subspaces. Classical (isotropic) radial basis functions on sparse grids have also been shown to be interpolatory in [48], albeit in a non-multilevel setting. The multilevel sparse kernel-based interpolation algorithm considered in the present work, however, is characterised by multilevel spaces which are *not* nested, as they are generated by different dilations of an infinitely differentiable function, in particular, the Gaussian on each level. Note that the interpolation nodes are nested, but in the case of kernel-based interpolation, nestedness of nodes does *not* imply nestedness of spanned approximation spaces. Therefore, in order to prove that the multilevel sparse kernel-based interpolation algorithm is interpolatory, we will generalise the aforementioned earlier results by showing that *the tensor-product nature of the kernel only* is sufficient to obtain interpolation on the directionally uniform sparse-grid nodes.

In what follows, we will need to compute sums of products of binomial coefficients in a number of different situations. One will be where we are applying the full $(N-1)$th order difference operator in the definition of S_k to the number of ways of filling $r \leq N$ positions in the multiindex with components which sum to j. This number is

$$\binom{j+r-1}{j} = \frac{(j+1)\cdots(j+r-1)}{(r-1)!},$$

which is a polynomial of degree $r-1$ with respect to j, with leading term $j^{r-1}/(r-1)!$. Hence, the difference operator applied to this will result in 0 if $r < N$ and be equal to 1, if $r = N$. On the other hand, we might be in the situation where the point in question only appears for the first time in some intermediate partial grid within the MuSIK scheme. In this case, we might not get the full difference operator applied. Let us suppose we only get a partial operator applied:

$$C_{N,s} := \sum_{j=s}^{N-1} (-1)^{N-1-j} \binom{N-1}{j} \binom{r-1+j-s}{j-s}.$$

The second binomial coefficient is in this form because the point will appear once in the first grid ($j = s$) and in the number of ways multi-indices of length $j - s$ can be put into r positions thereafter.

Let us rewrite

$$C_{N,s} = (-1)^{N-1-t} \sum_{j=0}^{N-1-t} (-1)^j \binom{N-1}{j+t}\binom{r-1+j}{j}. \tag{15}$$

If we now set

$$a_j := (-1)^j \binom{N-1}{j+t}\binom{r-1+j}{j},$$

we have

$$b_j = \frac{a_{j+1}}{a_j} \quad = \quad -\frac{(r+j)(N-1-j-t)}{(j+t+1)(j+1)},$$

so that, since $a_j = 0$ when $j > N - 1 - t$,

$$\sum_{j=0}^{N-1-t} a_j \quad = \quad a_0 \sum_{j=0}^{\infty} \prod_{n=0}^{j} b_n$$

$$= \quad \binom{N-1}{t} \sum_{j=0}^{\infty} \frac{(r)_j (t+1-N)_j}{j!(t+1)_j}$$

$$= \quad \binom{N-1}{t} {}_2F_1(r, t+1-N; t+1; 1), \tag{16}$$

where $(a)_j = a(a+1)\cdots(a+j-1)$ is the Pochhammer symbol, and ${}_2F_1$ is the Gauss hypergeometric function; see [1, Chapter 15]. Using [1, 15.1.20]

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}, \quad c-a-b > 0,$$

we have

$${}_2F_1(r, t+1-N; t+1; 1) \quad = \quad \frac{\Gamma(t+1)\Gamma(N-r)}{\Gamma(t+1-r)\Gamma(N)}. \tag{17}$$

Substituting this into (15) and (16) we see that

$$\sum_{j=t}^{N-1} (-1)^{N-1-j} \binom{N-1}{j}\binom{r-1+j-t}{j-r} \quad = \quad (-1)^{N-1}\binom{N-1}{t}\frac{\Gamma(t+1)\Gamma(N-r)}{\Gamma(t+1-r)\Gamma(N)}.$$

If $t \leq r-1$, then $\Gamma(t+1-r) = \infty$, and the sum is equal to zero. If $r = N$, then we also have a $\Gamma(0)$ in the numerator, and then by using continuity of the Γ functions, we have

$$\lim_{\epsilon \to 0} \frac{\Gamma(\epsilon)}{\Gamma(t+1-N+\epsilon)} = (-1)^{N-1-t}(N-1-t)!;$$

here we have used the Euler's reflection formula; see, e.g., [1, 6.1.17]. Thus, with a slight (valid) abuse of the applicability of (17),

$$\sum_{j=t}^{N-1} (-1)^{N-1-j} \binom{N-1}{j}\binom{N-1+j-t}{j-t}$$

$$= (-1)^{N-1-t}(-1)^{N-1}\binom{N-1}{t}\frac{t!(N-1-t)!}{(N-1)!} = 1.$$

We summarise the above development in the following result.

Lemma 4. *If $0 \le t < r \le N$, then*

$$\sum_{j=t}^{N-1}(-1)^{N-1-j}\binom{N-1}{j}\binom{r-1+j-t}{j-t} = \begin{cases} 1, & r = N, \\ 0, & 1 \le r < N. \end{cases}$$

We also refer to [46] for a related result.

We will say $\mathbf{m} < \mathbf{n}$ for two multi-indices $\mathbf{m}, \mathbf{n} \in \mathbb{N}_0^d$ if $m_j < n_j$ for $j = 1, \dots, d$, and analogously for $>, \le, \ge$. The key challenge we face in demonstrating that the combination scheme is interpolatory is that cardinal functions for points on one grid may not be equal to zero at points on a different grid. So let us fix a grid $\mathbb{X}_{\mathbf{m}}$ and compute the value of the SIK interpolant $S_k(\mathbf{x}_{\mathbf{m},\mathbf{i}})$ for some $0 \le \mathbf{i} \le 2^{\mathbf{m}}$.

For $\boldsymbol{\ell} \in I_{k,N} = \{\boldsymbol{\ell} \in \mathbb{N}_0^d : k \le |\boldsymbol{\ell}| \le k + N - 1\}$, consider the cardinal functions $\chi_{\boldsymbol{\ell},\mathbf{k}}$ at the point $\mathbf{x}_{\boldsymbol{\ell},\mathbf{k}}$ for some $\mathbf{k} \le 2^{\boldsymbol{\ell}}$. In order that $\chi_{\boldsymbol{\ell},\mathbf{k}}(\mathbf{x}_{\mathbf{m},\mathbf{i}})$ be non zero we require that

$$\chi_{l_j,k_j}(x_{m_j,i_j}) \quad \ne \quad 0, \quad j = 1, 2, \cdots, N,$$

which implies that

$$k_j \quad = \quad i_j 2^{l_j - m_j}, \quad l_j \ge m_j.$$

If $m_j < l_j$, $\chi_{l_j,k_j}(x_{m_j,i_j})$ can be non zero.

Since $\mathbf{x}_{\mathbf{m},\mathbf{k}} \in \mathbb{X}_{\tilde{\mathbf{m}}}$ for all $\tilde{\mathbf{m}} \ge \mathbf{m}$ if $\boldsymbol{\ell} \ge \mathbf{m}$, then the only non-zero cardinal function we get at $\mathbf{x}_{\mathbf{m},\mathbf{k}} \in \mathbb{X}_{\boldsymbol{\ell}}$ is $\chi_{\boldsymbol{\ell},\tilde{\mathbf{k}}}$, where $\tilde{k}_j = k_j 2^{l_j - m_j}$, $j = 1, 2, \cdots, N$, which takes the value 1, since $\mathbf{x}_{\mathbf{m},\mathbf{k}} = \mathbf{x}_{\boldsymbol{\ell},\tilde{\mathbf{k}}}$.

We will need to consider different subsets of indices as suggested above. Let $A = \{1, 2, \cdots, N\}$. Furthermore, suppose that $\omega = \{\omega_1, \cdots, \omega_s\}$, $s \le N$, is an enumeration of ω, and $\boldsymbol{\ell}(\omega) = (l_{\omega_1}, l_{\omega_2}, \cdots, l_{\omega_s}) \in \mathbb{N}_0^s$. Let $\tilde{\omega} = A \setminus \omega$, i.e., all the elements of A that are not elements of ω.

Given $\mathbf{c} \in \mathbb{N}_0^s$, with $c := |\mathbf{c}|_1$, we fix a point $\mathbf{y} \in I_{k,N}$, with $\mathbf{y}(\omega) = \mathbf{c}$; $\mathbf{y}$ is the grid point which we are interested to know in how many grids it belongs to. To do so, we seek the lowest index grid that $\mathbf{y}$ is in. To this end, there is a unique $\boldsymbol{\ell}_{\mathbf{y}} \in \mathbb{N}_0^N$ with minimum size such that $\mathbf{y} = \mathbf{x}_{\boldsymbol{\ell}_{\mathbf{y}},\mathbf{k}_{\mathbf{y}}} \in \mathbb{X}_{\boldsymbol{\ell}_{\mathbf{y}}}$, for some $\mathbf{k}_{\mathbf{y}} \in \mathbb{N}_0^N$ with $\mathbf{k}_{\mathbf{y}} \le 2^{\boldsymbol{\ell}_{\mathbf{y}}}$. Let $|\boldsymbol{\ell}_{\mathbf{y}}|_1 = u \ge 0$. Then, $\mathbf{y} \in \mathbb{X}_{\boldsymbol{\ell}}$ for all $\boldsymbol{\ell} \ge \boldsymbol{\ell}_{\mathbf{y}}$, with $\boldsymbol{\ell}(\omega) = \mathbf{c}$, and in what follows we compute the contribution at $\mathbf{x}_{\mathbf{m},\mathbf{i}}$ from cardinal (Lagrange) functions at $\mathbf{y}$ arising from all grids containing $\mathbf{y}$. If we only vary elements indexed by elements of ω (recall that ω has length s), then, at level $u + 1$, we have multi-indices which sum to 1 in $N - s$ positions, i.e., $N - s$ grids. For the next level we have multi-indices which sum to 2 in $N - s$ positions, i.e., $(N - s + 1)(N - s)/2$ grids, and so on. Letting

$$K(\mathbf{m}, \omega, \mathbf{c}, j) = \{\boldsymbol{\ell} \in I_{k,N} : \boldsymbol{\ell}(\omega) = \mathbf{c}, \boldsymbol{\ell}(\tilde{\omega}) \ge \mathbf{m}(\tilde{\omega}), |\boldsymbol{\ell}|_1 = \tilde{m} + c + j\}, \quad j = 0, 1, \cdots,$$

where $\tilde{m} = |\mathbf{m}(\tilde{\omega})|_1$, we seek to compute its cardinality, denoted by card $K(\mathbf{m}, \omega, \mathbf{c}, j)$. When we get to level $\tilde{m} + c + j$ we are restricting ourselves to the part of the sum without $\mathbf{c}$ which sums to c *and* without $\boldsymbol{\ell}_{\mathbf{y}}$ which sums up to u, because this is the first level we are looking at. Since k is the level of the initial grid we are interested in, hence $k - c - u$ is the multi-index size we initially have for the number of grids at this level with $y \in I_{k,N}$. Then, we add one each time with j. Then, we can see that

$$\text{card } K(\mathbf{m}, \omega, \mathbf{c}, j) \quad = \quad \binom{N - s - 1 + k - c - u + j}{N - s - 1}, \tag{18}$$

since we are selecting multi-indices which sum to $k - c - u + j$ to put in the $N - s$ positions where $\boldsymbol{\ell}(\tilde{\omega}) \ge \mathbf{m}(\tilde{\omega})$.

The contribution to $S_{k,N}$ coming from the point $\mathbf{y}$ from grids indexed by elements of $K(\mathbf{m}, \omega, \mathbf{c}, j)$ is

$$(-1)^{N-1} f(\mathbf{y}) \chi_{\boldsymbol{\ell}_{\mathbf{y}},\mathbf{k}_{\mathbf{y}}}(\mathbf{x}_{\mathbf{m},\mathbf{i}}) \sum_{j=0}^{N-1}(-1)^j \binom{N-1}{j} \text{card } K(\mathbf{m}, \omega, \mathbf{c}, j)$$

$$= \quad (-1)^{N-1} f(\mathbf{y}) \chi_{\boldsymbol{\ell}_{\mathbf{y}},\mathbf{k}_{\mathbf{y}}}(\mathbf{x}_{\mathbf{m},\mathbf{i}}) \sum_{j=0}^{N-1}(-1)^j \binom{N-1}{j}\binom{N - s - 1 + k - c - u + j}{N - s - 1},$$

using (18).

We now have two situations. Either $c + u \leq k$, or $c + u > k$. In the former case, card $K(\mathbf{m}, \omega, \mathbf{c}, j)$ is a polynomial of degree $N - s - 1$ in j. Therefore, Lemma 4 with $t = 0$, yields that the interpolant vanishes in this case. In the latter case, we set $t = c + u - k$ in Lemma 4. Then, the above sum becomes

$$\sum_{j=t}^{N-1} (-1)^j \binom{N-1}{j} \binom{N-s-1+j-t}{N-s-1},$$

which is equal to zero by Lemma 4 if $t < N - s$. Since $c_i \leq m_{\omega_i} - 1$, $i = 1, 2, \cdots, s$, we have $c \leq \sum_{i=1}^{s}(m_{\omega_i} - 1) = \sum_{i=1}^{s} m_{\omega_i} - s$. Thus, $|m|_1 - s \geq c + u$, and since $|m|_1 \leq k + N - 1$, we have $c + u + s \leq k + N - 1$. Rearranging we have $t = c + u - k \leq N - 1 - s < N - s$ as required. Hence, in either situation, those contributions to the interpolant coming from points different to $\mathbf{x}_{\mathbf{m},\mathbf{k}}$ are zero.

We are left with computing the contribution to the interpolant at $\mathbf{y} = \mathbf{x}_{\mathbf{m},\mathbf{k}}$ from grids containing $\mathbf{x}_{\mathbf{m},\mathbf{k}}$. As before, we look for grid of lowest multi-index size $\boldsymbol{\ell}_{\mathbf{y}} \in \mathbb{N}_0^N$, such that $\mathbf{x}_{\mathbf{m},\mathbf{k}} \in \mathbb{X}_{\boldsymbol{\ell}}$. Suppose that $|\boldsymbol{\ell}_{\mathbf{y}}|_1 = u$. Let

$$K(\mathbf{y}, j) = \{|\boldsymbol{\ell}|_1 = j : \mathbf{y} \geq \boldsymbol{\ell}_{\mathbf{y}}\}.$$

Then $\mathbf{y} \in K(\mathbf{y}, j)$, for all $j \geq u$ and, hence

$$\text{card } K(\mathbf{y}, j) = \binom{N-1+j-u}{N-1}.$$

Then, since for $\boldsymbol{\ell} \in K(\mathbf{y}, j)$ there is unique $\mathbf{i_y} \leq 2^{\boldsymbol{\ell}}$ such that $\mathbf{x}_{\boldsymbol{\ell},\mathbf{i_y}} = \mathbf{y}$, and

$$\chi_{\boldsymbol{\ell},\mathbf{i_y}}(\mathbf{y}) = 1,$$

we have the contribution to $S_{k,N}$ from cardinal functions centred at $\mathbf{y}$ as

$$f(\mathbf{y}) \sum_{j=0}^{N-1} (-1)^{N-1-j} \binom{N-1}{j} \text{card } K(\mathbf{y}, k+j)$$

$$= f(\mathbf{y}) \sum_{j=0}^{N-1} (-1)^{N-1-j} \binom{N-1}{j} \binom{N-1+k+j-u}{N-1}.$$

As before we have two cases, and similarly, by Lemma 4 the sum is 1. Hence the contribution coming from $\mathbf{y}$ is $f(\mathbf{y})$.

Thus we have already proven the following result.

Theorem 5. *Assuming that the interpolation kernel has the form*

$$\psi(\mathbf{y}) = \prod_{i=1}^{d} \phi(y_i),$$

then SIK and the respective MuSIK with this kernel are interpolatory.

4. Stochastic Collocation based on MuSIK

4.1. Bounded random variables

If we wish to apply MuSIK in the context of (8), we only need to calculate the weights for the anisotropic Gaussian interpolants for each partial grid. This is an immediate consequence of the linearity of both the combination (13) and of the multilevel enrichment procedures (14).

To this end, let $\Gamma^N := \bigotimes_{i=1}^{N}[\alpha_i, \beta_i]$, and we consider the interpolant

$$S_{A_\ell}(\mathbf{y}) = \sum_{j=1}^{N^\ell} c_j \varphi(\|A_\ell(\mathbf{y} - \mathbf{y}_j)\|), \tag{19}$$

of $G(\mathbf{y})$ in (8) onto a (generic) partial grid. Using the tensor form of the N-variate Gaussian Lagrange basis, we deduce

$$\int_{\Gamma^N} S_{A_\ell}(\boldsymbol{\xi})d\boldsymbol{\xi} = \sum_{j=1}^{N^\ell} c_j \prod_{i=1}^{N} \int_{\alpha_i}^{\beta_i} e^{-a_i^2 c^2 (\xi_i - y_j^i)^2} d\xi_i = \sum_{j=1}^{N^\ell} c_j \prod_{i=1}^{N} g_{ij}, \tag{20}$$

with A_ℓ is the $N \times N$ scaling diagonal matrix with diagonal elements a_i, ξ_i is the ith dimensional component of $\boldsymbol{\xi}$, y_j^i is the ith dimensional component of $\mathbf{y}_j$, and g_{ij} stands for the integral from the ith dimensional component of jth basis function. A simple calculation reveals that g_{ij} is given by

$$g_{ij} = \frac{1}{a_i c} \cdot \frac{\sqrt{\pi}}{2} \Big[\mathrm{erf}(a_i c(\beta_i - y_j^i)) - \mathrm{erf}(a_i c(\alpha_i - y_j^i)) \Big], \tag{21}$$

with $\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

Therefore the integral on the left-hand side of (20) can be constructed by the linear combination of products of g_{ij}'s, from which the quadrature weights can be calculated.

4.2. Unbounded random variables

We now discuss the possible extension of the above MuSIK stochastic collocation method to random variables with unbounded ranges Γ^N and, in particular for Gaussian random variables. To this end, let $\{Y_i(w) \sim \mathcal{N}(\mu_i, \sigma_i^2)\}_{i=1}^N$ to be N independent Gaussian random variables.

The key observation is the probability density function of Gaussian distribution decays exponentially fast on tails. Therefore, the MuSIK quadrature rule for (10) over $\Gamma^N = \mathbb{R}^N$, with ρ involving Gaussian functions, can be arbitrarily well approximated by interpolating the integrand on MuSIK nodes contained in the hypercubes $H_k := \bigotimes_{i=1}^N [\mu_i - r\sigma_i, \mu_i + r\sigma_i]$, instead, with $r \in \mathbb{N}$ a user-defined constant. The resulting interpolant S_{A_ℓ} is a good approximation to the integrands over $\Gamma^N = \mathbb{R}^N$: multivariate Gaussians used by MuSIK also decay exponentially fast outside H_k. In this context, we set $\alpha_i = -\infty$ and $\beta_i = \infty$ for $i = 1, \dots, N$, in (20).

Therefore, the proposed MuSIK stochastic collocation approach results to hierarchical node distributions also in the case of unbounded random variables. To substantiate further this claim, we present a numerical example for the MuSIK stochastic collocation method for Gaussian random variables in Section 5.2.

5. Numerical examples

We present a series of numerical examples highlighting the good performance of the proposed MuSIK stochastic collocation method for the approximation of elliptic PDEs with random coefficients.

5.1. Example 1

We shall now illustrate the convergence of the new stochastic collocation method based on MuSIK for the stochastic linear elliptic boundary value problem with two physical spatial dimensions ($d = 2$). The example here is take from [45],

$$-\nabla \cdot (a(\omega, x_1, x_2)\nabla u(\omega, x_1, x_2)) = f(\omega, x_1, x_2) \quad \text{in} \quad \Omega \times D,$$
$$u(\omega, x_1, x_2) = 0 \quad \text{on} \quad \Omega \times \partial D, \tag{22}$$

with $D = [0,1]^2$, a deterministic load $f(\omega, x_1, x_2) = \cos(x_1)\sin(x_2)$ and the random diffusion coefficient $a_N(\omega, x_1)$ is independent of spatial variable x_2, constructed by:

$$\log(a_N(\omega, x_1) - 0.5) = 1 + Y_1(\omega)(\frac{\sqrt{\pi}L}{2})^{1/2} + \sum_{i=2}^N \xi_i \phi_i(x_1) Y_i(\omega), \tag{23}$$

where

$$\xi_i := (\sqrt{\pi}L)^{1/2} \exp(\frac{-(\lfloor \frac{i}{2} \rfloor \pi L)^2}{8}), \quad \text{if} \quad i \ge 1, \tag{24}$$

13

and

$$\phi_i(x_1) := \begin{cases} \sin(\frac{\lfloor \frac{i}{2} \rfloor \pi x_1}{L_p}), & \text{for } i \text{ even}; \\ \cos(\frac{\lfloor \frac{i}{2} \rfloor \pi x_1}{L_p}), & \text{for } i \text{ odd}, \end{cases} \tag{25}$$

for $\{Y_n(\omega)\}_{n=1}^{\infty}$ independent random variables, uniformly distributed in the interval $[-\sqrt{3}, \sqrt{3}]$, with zero mean and unit variance, viz., $E[Y_n] = 0$ and $E[Y_n Y_m] = \delta_{mn}$ for $n, m \in \mathbb{N}$. The corresponding joint probability density function is denoted by ρ and (23) is the truncation of an one-dimensional random field with stationary covariance

$$\text{cov}[\log(a_N - 0.5)](\xi, \eta)$$

$$= E[(\log(a)(\xi) - E[\log(a)(\xi)])(\log(a)(\eta) - E[\log(a)(\eta)])] = \exp(\frac{-(\xi - \eta)^2}{L_c^2}),$$

For $x_1 \in [0, 1]$, let L_c be the desired physical correlation length for the coefficient a, meaning that for $|\xi - \eta| >> L_c$ two random variables $a(\xi)$ and $a(\eta)$ become uncorrelated. Then, set $L_p := \max\{1, 2L_c\}$ in (25), and $L := L_c/L_p$ in (24) and (23).

For the solution of each independent finite element problem, standard cubic Lagrange elements over a uniform triangulation of D with 4225 unknowns are used. To study the convergence rate of the MuSIK stochastic collocation method, we consider the problem with a fixed dimension N and investigate the error behavior with respect to the different levels k. The implementation of this algorithm is performed in MATLAB using the ALICE high performance computing facility at the University of Leicester. We shall focus only on the quadrature-induced part of the error.

To estimate the $L^2(D)$-norm of the computational error ϵ in the k-th level of MuSIK, we use the approximation

$$\|\text{E}[\epsilon]\|_{L^2(D)} \approx \|\text{E}[\mathcal{M}(K, N)\Pi_h u_N - \mathcal{M}(k, N)\Pi_h u_N]\|_{L^2(D)}, \tag{26}$$

with $\mathcal{M}(k, N)$ denoting the MuSIK collocation projection at level k over dimension N probability space with $k = 0, 1, 2, \ldots, K$, and Π_h denoting finite element projection over the spatial domain D. We begin by assessing the performance of the method with respect to the probability space dimension N for $N = 5, 10, 14, 21$, selecting correlation length $L_c := 1/64$. The $L^2(D)$-norm of the computational error ϵ against 'Dof' (standing for the total number of collocation points) is given in Figure 4. As the level k increases, the observed convergence rate appears to be (at least) algebraic. We also observe the (expected) slow deterioration of convergence rate as N increases (cf. [4, 45] for the corresponding computations using polynomially exact quadratures). Further, to investigate the performance of MuSIK approximation under different correlation lengths L_c, we set $L_c = 1/64, 1/16, 1/2$ for $N = 5, 10$. The results are given in Figure 5.

The Gaussian building blocks of MuSIK include a user-defined so-called *shape parameter c* which affects the convergence of the MuSIK method as well as the numerical stability of the Lagrange function computations. For a detailed discussion on this issue for we refer to [27] and the references therein. All the above results are based on the set of chosen shape parameters $\{0.01, 0.01, 0.03, 0.06, 0.085, 0.09, 0.1\}$ for levels $1, 2, \ldots, 7$, which result to good convergence rate and sufficient numerical stability. We have found that a good practical range for c to guarantee both good convergence and numerical stability is $[0.01, 0.4]$. This range can be enlarged if one uses higher precision arithmetic for the calculation of the Lagrange basis, thereby circumventing the effect of bad conditioning of the cardinal function computations. Figure 6 shows convergence plots for the method for $N = 5, 10$ with two different choices of shape parameter c: one choice is the variable shape parameter detailed above, with the second choice being selecting a uniform shape parameter $c = 0.1$ for all levels.

5.2. Example 2

Next, we consider the elliptic boundary value problem (22), over the L-shaped domain $D := (-1, 1)^2 \backslash [0, 1) \times (-1, 0]$, with random diffusion coefficient and load function given by:

$$\log(a_N(\omega, x_1, x_2) - 0.5) = 1 + Y_1(\omega) \sin(\pi x_1) + Y_2(\omega) \sin(\pi x_2)$$
$$+ Y_3(\omega) \cos(\pi x_1) + Y_4(\omega) \cos(\pi x_1), \tag{27}$$

and $f(\omega, x_1, x_2) = 1 + \exp\{Y_5(\omega) \sin(\pi x_1) \sin(\pi x_2) + Y_6(\omega) \cos(\pi x_1) \cos(\pi x_2)\}$, respectively. We make two different choices of independent identically distributed random variables $\{Y_n(\omega)\}_{n=1}^{6}$: a) uniform random variables with zero
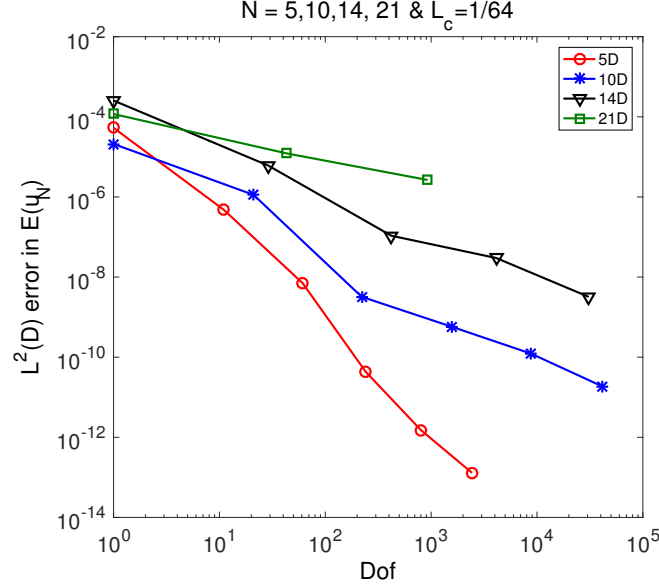
14

Figure 4: Convergence for $N = 5, 10, 14, 21$ with $L_c = 1/64$.

mean and unit variance distributed over the interval $[-\sqrt{3}, \sqrt{3}]$, and b) Gaussian random variables satisfying the normal distribution $\mathcal{N}(0, 1)$, respectively. Note that the diffusion coefficient depends on 4 random variables and the forcing term on 2 random variables.

The solution u has low regularity at the D-origin, reflecting the typical (singular) behavior in the vicinity of re-entrant corners. To resolve this, we use 1536 cubic elements with the physical space mesh geometrically graded towards the origin. We are interested in the numerical results for the $L^2(D)$ approximation error of the expected value $E[u_N]$ and of the variance $\text{Var}[u_N]$, with $N = 6$, under different choices of shape parameter c; the error is computed in a completely analogous fashion to (26). We set $r = 3$ for the domain truncation for the case of Gaussian random variables.

As the level k of the MuSIK algorithm increases, the observed convergence rate in both expected value and variance are at least algebraic for a number of choices of shape parameters for both uniform and Gaussian random variables, respectively; the respective convergence histories are given in Figures 7 and 8. As expected, the smaller shape parameter gives smaller error independent of the levels. Nonetheless, the shape parameter appears to have an influence on the convergence rate of the proposed MuSIK algorithm when the level is low, which appears to gradually diminish as the level grows. The choice of shape parameter in this numerical example is such that the univariate Lagrange basis computations admit condition numbers of 10^5 or less.

6. Concluding remarks

A new stochastic collocation finite element method based on the recent multilevel sparse kernel-based interpolation for the, typically high-dimensional, quadrature step for the solution elliptic PDEs with random coefficients is proposed and tested. The implementation of the method is fully parallelizable, while an interesting attribute in the use of directionally uniform multilevel nested quadrature nodes for the approximation of both bounded and unbounded random variables. As such, the development of adaptive anisotropic versions of the MuSIK-based quadrature rule, which could lead to further complexity reduction, is feasible, since adding new (anisotropic) levels can reuse lower level finite element computations. This, of course, could be combined with the tremendous advances in finite element adaptive algorithms to lead to a fully adaptive hierarchical algorithmic framework; we refer to [29, 11] for some recent results in this direction for stochastic Galerkin finite element methods. This is an interesting future direction of research.
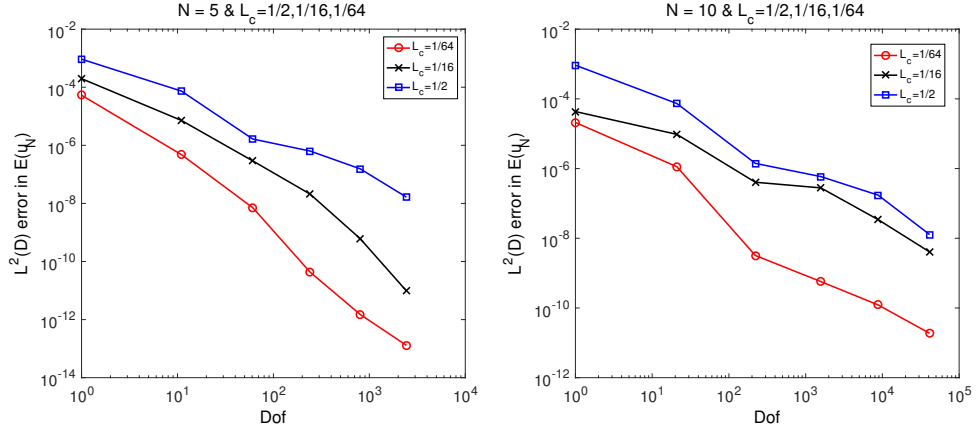
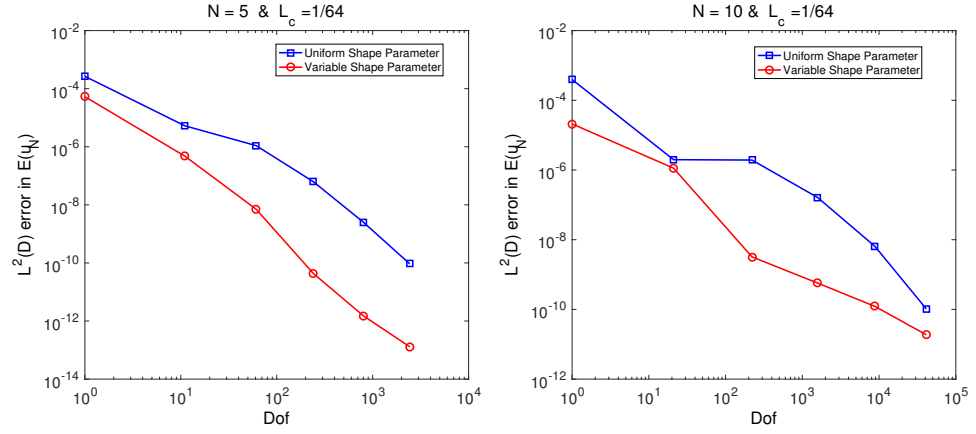Figure 5: $N = 5$ (left) and $N = 10$ (right) with $L_c = 1/64, 1/16, 1/2$.

Figure 6: Uniform against variable shape parameter for $N = 5, 10$, $L_c = 1/64$.
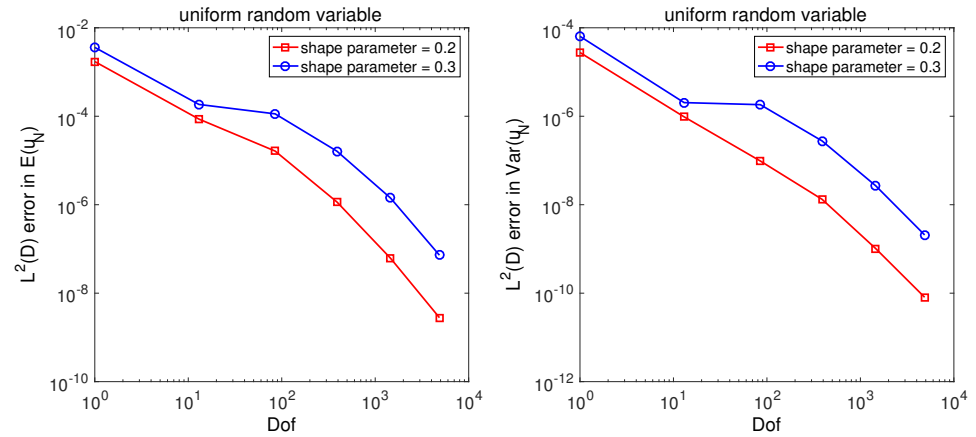

Figure 7: The error decay in expectation and variance under shape parameter $c = 0.2$ and $c = 0.3$.
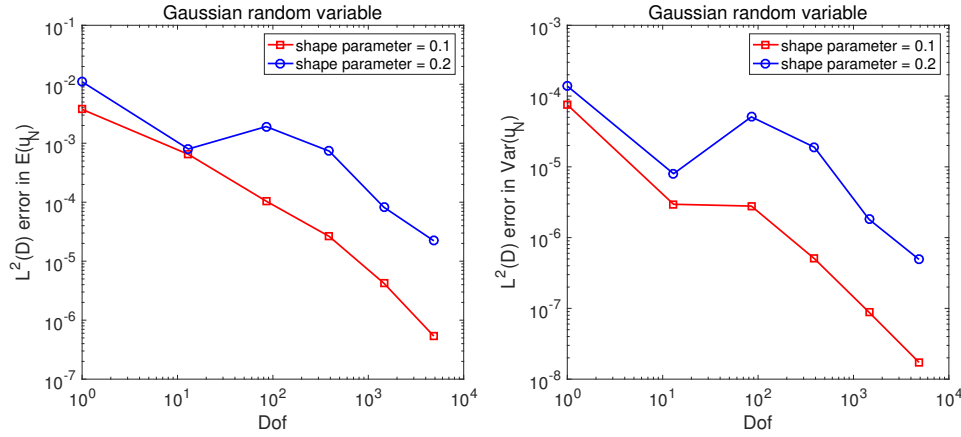
16

Figure 8: The error decay in expectation and variance under shape parameter $c = 0.1$ and $c = 0.2$.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of mathematical functions : with formulas, graphs, and mathematical tables*. Dover, New York, 1965. Unaltered, unabridged republication of 55, National Bureau of Standards, Applied mathematics series (1964) corrected edition.–p. iv of cover.

[2] I. Babuška and P. Chatzipantelidis. On solving elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 191(37-38):4093–4122, 2002.

[3] I. Babuška, F. Nobile, and R. Tempone. Worst case scenario analysis for elliptic problems with uncertainty. *Numerische Mathematik*, 101(2):185–219, 2005.

[4] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45:1005–1034, 2007.

[5] I. 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, 2004.

[6] I. 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.

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

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

[9] R. Beatson, O. Davydov, and J. Levesley. Error bounds for anisotropic RBF interpolation. *J. Approx. Theory*, 162(3):512–527, 2010.

[10] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal stochastic Galerkin methods for a class of PDES with random coefficients. *Comput. Math. Appl.*, 67(4):732–751, 2014.

[11] A. Bespalov, C. E. Powell, and D. Silvester. Energy norm a posteriori error estimation for parametric operator equations. *SIAM J. Sci. Comput.*, 36(2):A339–A363, 2014.

[12] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15. Springer, 2008.

[13] M. D. Buhmann. *Radial basis functions: theory and implementations*, volume 12. Cambridge university press, 2003.

[14] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[15] J. Charrier, R. Scheichl, and A. L. Teckentrup. Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods. *SIAM J. Numer. Anal.*, 51(1):322–352, 2013.

[16] A. Chernov. Sparse polynomial approximation in positive order Sobolev spaces with bounded mixed derivatives and applications to elliptic problems with random loading. *Applied Numerical Mathematics*, 62(4):360–377, 2012.

[17] A. Chernov and C. Schwab. First order k-th moment finite element analysis of nonlinear operator equations with stochastic data. *Math. Comp.*, 82(284):1859–1888, 2013.

[18] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput. Vis. Sci.*, 14(1):3–15, 2011.

[19] A. Cohen and R. DeVore. Approximation of high-dimensional parametric pdes. *Acta Numerica*, 24:1–159, 2015.

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

[21] A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Analysis and Applications*, 9(01):11–47, 2011.

[22] F.-J. Delvos. d-variate boolean interpolation. *Journal of Approximation Theory*, 34(2):99–114, 1982.

[23] 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.

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

[25] J. Garcke. Sparse grids in a nutshell. In *Sparse grids and applications*, pages 57–80. Springer, 2012.

[26] A. Genz and B. D. Keister. Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight. *J. Comput. Appl. Math.*, 71(2):299–309, 1996.

[27] E. H. Georgoulis, J. Levesley, and F. Subhan. Multilevel sparse kernel-based interpolation. *SIAM J. Sci. Comput.*, 35(2):A815–A831, 2013.

[28] Michael B Giles, Tigran Nagapetyan, and Klaus Ritter. Multilevel monte carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):267–295, 2015.

[29] C. J. Gittelson. An adaptive stochastic Galerkin method for random elliptic operators. *Math. Comp.*, 82(283):1515–1541, 2013.

[30] C. J. Gittelson. Convergence rates of multilevel and sparse tensor approximations for a random elliptic PDE. *SIAM J. Numer. Anal.*, 51(4):2426–2447, 2013.

[31] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, C. Schwab, and I. H. Sloan. Quasi-monte carlo finite element methods for elliptic pdes with lognormal random coefficients. *Numerische Mathematik*, 131(2):329–368, 2015.

[32] M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.

[33] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems. 1992.

[34] A. Haji-A, F. Nobile, L. Tamellini, and R. Tempone. Multi-index stochastic collocation for random pdes. *Computer Methods in Applied Mechanics and Engineering*, 306:95–122, 2016.

[35] F. Heiss and V. Winschel. Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1):62–80, 2008.

[36] V. H. Hoang and C. Schwab. N-term wiener chaos approximation rates for elliptic pdes with lognormal Gaussian random inputs. *Mathematical Models and Methods in Applied Sciences*, pages 1–30, 2013.

[37] A. Iske and J. Levesley. Multilevel scattered data approximation by adaptive domain decomposition. *Numer. Algorithms*, 39(1-3):187–198, 2005.

[38] F. Y. Kuo, C. Schwab, and I. H. Sloan. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM J. Numer. Anal.*, 50(6):3351–3374, 2012.

[39] F. Y. Kuo, C. Schwab, and I. H. Sloan. Multi-level quasi-monte carlo finite element methods for a class of elliptic pdes with random coefficients. *Foundations of Computational Mathematics*, 15(2):411–449, 2015.

[40] M. Lòeve. *Probability Theory I*, volume 45 of *Graduate Texts in Mathematics*. Springer-Verlag, 4th ed. edition, 1977.

[41] M. Lòeve. *Probability Theory II*, volume 46 of *Graduate Texts in Mathematics*. Springer-Verlag, 4th ed. edition, 1978.

[42] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194(12):1295–1331, 2005.

[43] F. J. Narcowich, R. Schaback, and J. D. Ward. Multilevel interpolation and approximation. *Applied and Computational Harmonic Analysis*, 7(3):243–261, 1999.

[44] F. Nobile, R. Tempone, and C. G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442., 2008.

[45] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345., 2008.

[46] M. S. Ruiz. An algebraic identity leading to wilson's theorem. *The Mathematical Gazette*, 80(489):579–582, 1996.

[47] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math.*, 3(3):251–264, 1995.

[48] A. Schreiber. *The method of Smolyak in multivariate interpolation*. PhD thesis, der Mathematisch-Naturwissenschaftlichen Fakultäten, der Georg-August-Universität zu Göttingen, 2000.

[49] C. Schwab, E. Süli, and R. A. Todor. Sparse finite element approximation of high-dimensional transport-dominated diffusion problems. *M2AN Math. Model. Numer. Anal.*, 42(5):777–819, 2008.

[50] S. A. Smolyak. Quadrature and interpolation of formulas for tensor product of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963.

[51] A. L. Teckentrup, P. Jantsch, C. G. Webster, and M. Gunzburger. A multilevel stochastic collocation method for partial differential equations with random input data. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1046–1074, 2015.

[52] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann. Further analysis of multilevel monte carlo methods for elliptic pdes with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.

[53] R. A. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, 27(2):232–261, 2007.

[54] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in computational Mathematics*, 4(1):389–396, 1995.

[55] H. Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

[56] P. Zaspel. *Parallel RBF Kernel-Based Stochastic Collocation for Large-Scale Random PDEs*. PhD thesis, Universitäts-und Landesbibliothek Bonn, 2015.