Convergence of Sparse Collocation for Functions of Countably Many Gaussian Random Variables (with Application to Elliptic PDEs)

Oliver G. Ernst[†], Björn Sprungk[†] and Lorenzo Tamellini[*]

[†] Department of Mathematics, TU Chemnitz, Germany

[*] Istituto di Matematica Applicata e Tecnologie Informatiche "E. Magenes" del CNR, Pavia, Italy

Abstract

We give a convergence proof for the approximation by sparse collocation of Hilbert-space-valued functions depending on countably many Gaussian random variables. Such functions appear as solutions of elliptic PDEs with lognormal diffusion coefficients. We outline a general L^2-convergence theory based on previous work by Bachmayr et al. [4] and Chen [9] and establish an algebraic convergence rate for sufficiently smooth functions assuming a mild growth bound for the univariate hierarchical surpluses of the interpolation scheme applied to Hermite polynomials. We verify specifically for Gauss-Hermite nodes that this assumption holds and also show algebraic convergence w.r.t. the resulting number of sparse grid points for this case. Numerical experiments illustrate the dimension-independent convergence rate.

Keywords: random PDEs, parameteric PDEs, lognormal diffusion coefficient, best-N-term approximation, sparse grids, stochastic collocation, high-dimensional approximation, high-dimensional interpolation, Gauss-Hermite points.

Mathematics Subject Classification: 65D05, 65D15, 65C30, 60H25.

1 Introduction

The elliptic diffusion problem

$$-\nabla\cdot(a(\omega)\,\nabla u(\omega)) = f \quad \text{in } D\subset\mathbb{R}^d, \qquad u(\omega)=0 \text{ on } \partial D, \qquad \mathbb{P}\text{-a.s.}\,, \tag{1}$$

with a random diffusion coefficient $a:\Omega\to L^\infty(D)$ with respect to an underlying probability space $(\Omega,\mathcal{A},\mathbb{P})$ has become the standard model problem for numerical methods for solving random PDEs. For modeling reasons the diffusion field is often taken to have a lognormal probability law, which complicates both the study of the well-posedness of the problem [8, 24, 19, 31] as well as the analysis of approximation methods. One of the challenges is that the most common parametrization of a Gaussian random field – the *Karhunen-Loève expansion* [2, 22] – involves a countable number of standard normal random variables

$$\log a(x,\omega) = \phi_0(x) + \sum_{m=1}^{\infty} \phi_m(x)\,\xi_m(\omega), \tag{2}$$

where $\phi_0, \phi_m \in L^\infty(D)$ and $\xi_m \sim N(0,1)$ i.i.d. for $m \in \mathbb{N}$, leading to an elliptic PDE with a countably infinite number of random parameters $\boldsymbol{\xi} = (\xi_m)_{m\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$.

Besides the stochastic Galerkin method [22, 29] the most common methods for approximating the solution $u(\boldsymbol{\xi})$ of such random or parametric elliptic PDEs are *polynomial collocation methods*. Early works on such methods for random PDEs considered a finite (if large) number of random parameters, a setting also referred to as *finite-dimensional noise* [44, 3, 37, 36]. In this case the parametric representation of $\log a$ is typically obtained by truncating a series expansion of the random field such as (2).

The analysis of the problem involving an infinite number of random variables was first discussed by Cohen, DeVore and Schwab in [14, 15] in the simpler setting in which the diffusion field a, rather than its logarithm as in (2), is expanded in a series. This results in an affine dependence of a on the random variables ξ_m, which are, moreover, assumed to have bounded support. In this framework the convergence of the best N-term approximation of the solution of the diffusion equation by Taylor and Legendre series was shown to be independent of the number of random variables; this result was further refined in the recent paper [5]. Employing the theoretical concepts stated in [14, 15], Chkifa, Cohen and Schwab analyze in [11] collocation methods based on Lagrange interpolation with Leja points for problems with diffusion coefficients depending linearly on an infinite number of bounded random variables, which are adaptive in the polynomial degree as well as the number of active dimensions or random variables, respectively. The adaptive algorithm itself is related to the earlier work [21]. Each interpolatory approximation gives rise to a quadrature scheme, and in [39] Schillings and Schwab consider sparse adaptive quadrature schemes in the same setting of [11] in connection with approximating expectations with respect to posterior measures in Bayesian inference. Extensions to the case where the diffusion coefficient a depends non-linearly on an infinite number of random variables with bounded support was discussed in [12].

Returning to the original lognormal diffusion problem, i.e., with a expanded as in (2) and depending on random variables with unbounded support, Hoang and Schwab [26] have obtained convergence results on best N-term approximation by Hermite polynomials. These were recently extended by Bachmayr et al. [4] using a different analytical approach employing a weighted ℓ^2-summability of the coefficients of the Hermite expansion of the solution and their relation to partial derivatives. The theoretical tools provided in [4] enabled a convergence analysis for adaptive sparse quadrature [9] employing, e.g., Gauss-Hermite nodes for Banach space-valued functions of countably many Gaussian random variables.

In this paper we address the convergence of sparse polynomial collocation for functions of infinitely many Gaussian random variables, such as the solution to the lognormal diffusion problem (1). Specifically, we follow the approach of [4] and [9] to prove an algebraic convergence rate with respect to the number of grid points for sparse collocation based on Gauss-Hermite interpolation nodes in the case of countably many variables. In particular, the result applies to solutions u of (1) where a is a lognormal random field. In addition, we highlight the common ideas surrounding sparse collocation found in the works mentioned above. The convergence result in terms of the number of collocation points is obtained in two steps: we first link the error to the size of the multi-index set defininig the sparse collocation and then derive a bound on the number of points in the associated sparse grid. This procedure has been followed also in all the above-mentioned work analyzing the convergence of sparse grid quadrature and collocation schemes. An alternative strategy which instead links the error directly to the number of collocation points by introducing the so-called "profits" of each component of the sparse grids, has been discussed in [34, 25], albeit only in the case of random variables with bounded support.

We remark that, besides the classical node families such as Gauss-Hermite and Genz-Keister [20] for quadrature and interpolation on $\mathbb{R}$ with respect to a Gaussian measure, Jakeman and Narayan [32] have introduced *weighted Leja points*—a generalization of the classical Leja point construction (see e.g. [30, 17] and references therein) to unbounded domains and arbitrary weight functions. Moreover, they have proved that these node sets possess the correct asymptotic distribution of interpolation nodes and illustrate their computational potential in numerical experiments. Note that such weighted Leja points provide a nested and

linearly growing sequence of interpolation nodes. The analysis of sparse collocation based on normal Leja points, i.e., weighted Leja points for a Gaussian measure, is an interesting topic for future research.

The remainder of the paper is organized as follows. In the next section we introduce the general setting and notation and construct the sparse grid collocation operator based on univariate Lagrange interpolation operators. Section 3 is devoted to the convergence analysis of such operators. First, we outline in Subsection 3.1 the general approaches to prove algebraic convergence rates as they can be found in the works mentioned above. Later, we follow in Subsection 3.2 the approach of [4, 9] and derive sufficient conditions for the underlying univariate interpolation nodes in order to obtain such rates when approximating "countably-variate" functions of certain smoothness. Finally, in Subsection 3.3 we verify these conditions for Gauss-Hermite nodes, provide bounds for the number of nodes in the resulting sparse grids, and state a convergence result with respect to this number. Section 4 comes back to our motivation and comments on the application to random elliptic PDEs before we verify our theoretical findings in Section 5 for a simple boundary value problem in one spatial dimension. We draw final conclusions in Section 6.

2 Setting and Sparse Collocation

We consider functions f defined on a parameter domain $\Gamma \subseteq \mathbb{R}^{\mathbb{N}}$ taking values in a separable real Hilbert space $\mathcal{H}$ with inner product $(\cdot,\cdot)_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$. As our interest lies in the approximation of the dependence of $f : \Gamma \to \mathcal{H}$ on $\boldsymbol{\xi} \in \Gamma$ by multivariate polynomials based on Lagrange interpolation, a minimal requirement is that point evaluation of f at any $\boldsymbol{\xi} \in \Gamma$ be well-defined. Stronger smoothness requirements on f become necessary when deriving convergence rate estimates for the approximations.

We introduce a probability measure μ on the measurable space $(\mathbb{R}^{\mathbb{N}}, \otimes_{m\geq 1}\mathcal{B}(\mathbb{R}))$ as the countable product measure of standard Gaussian measures on $\mathbb{R}$, i.e.,

$$\mu = \bigotimes_{m\geq 1} N(0,1). \tag{3}$$

and denote by $L^2_\mu(\Gamma;\mathcal{H})$ the space of all (equivalence classes of) functions with finite second moments with respect to μ in the sense that

$$\int_{\mathbb{R}^{\mathbb{N}}} \|f(\boldsymbol{\xi})\|^2_{\mathcal{H}}\, \mu(\mathrm{d}\boldsymbol{\xi}) < \infty$$

which forms a Hilbert space with inner product

$$(f,g)_{L^2_\mu} = \int_{\mathbb{R}^{\mathbb{N}}} (f(\boldsymbol{\xi}), g(\boldsymbol{\xi}))_{\mathcal{H}}\, \mu(\mathrm{d}\boldsymbol{\xi}).$$

In the following we require

Assumption A1. Let $f : \Gamma \to \mathcal{H}$ where $\mu(\Gamma) = 1$. There holds (for a measurable extension of f to $\mathbb{R}^{\mathbb{N}}$) that $f \in L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})$.

It is shown, e.g., in [40, Theorem 2.5], that the countable tensor product of Hermite polynomials forms an orthonormal basis of $L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})$. Under Assumption A1 we therefore have

$$f(\boldsymbol{\xi}) = \sum_{\boldsymbol{\nu}\in\mathcal{F}} f_{\boldsymbol{\nu}}\, H_{\boldsymbol{\nu}}(\boldsymbol{\xi}), \qquad f_{\boldsymbol{\nu}} := \int_{\mathbb{R}^{\mathbb{N}}} f(\boldsymbol{\xi})\, H_{\boldsymbol{\nu}}(\boldsymbol{\xi})\, \mu(\mathrm{d}\boldsymbol{\xi}) \in \mathcal{H}, \tag{4}$$

where $H_{\boldsymbol{\nu}}(\boldsymbol{\xi}) = \prod_{m\geq 1} H_{\nu_m}(\xi_m)$ and H_ν denotes the univariate Hermite orthonormal polynomial of degree ν as well as

$$\mathcal{F} := \left\{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} : |\boldsymbol{\nu}|_0 < \infty\right\}, \qquad |\boldsymbol{\nu}|_0 := |\{j \in \mathbb{N} : \nu_j > 0\}|.$$

2.1 Sparse Polynomial Collocation

The construction of sparse collocation operators below is based on sequences of univariate Lagrange interpolation operators U_k mapping into the set $\mathcal{P}_k$ of univariate polynomials of degree at most $k \in \mathbb{N}_0$. Thus,

$$(U_k f)(\xi) = \sum_{i=0}^{k} f(\xi_i^{(k)})\, L_i^{(k)}(\xi), \qquad f\colon \mathbb{R} \to \mathbb{R},$$

where $\{L_i^{(k)}\}_{i=0}^k$ denote the Lagrange fundamental polynomials of degree k associated with the set of $k+1$ distinct interpolation nodes $\Xi^{(k)} := \{\xi_0^{(k)}, \xi_1^{(k)}, \ldots, \xi_k^{(k)}\}$.

Remark 1. It may also be of interest to consider sequences of interpolation operators U_k with a more general degree of polynomial exactness $n(k)$ where $n : \mathbb{N}_0 \to \mathbb{N}_0$ is indecreasing and $n(0) = 0$, see for instance [44, 3, 37, 36, 35, 34]. However, we restrict ourselves to $n(k) = k$ for simplicity.

We also introduce the *detail operators*

$$\Delta_k := U_k - U_{k-1}, \quad k \geq 0,$$

where we set $U_{-1} :\equiv 0$, and observe that

$$U_k = U_{k-1} + \Delta_k = \Delta_0 + \Delta_1 + \cdots + \Delta_k\,.$$

Tensorization For any multi-index $\boldsymbol{k} = (k_m)_{m\in\mathbb{N}} \in \mathcal{F}$ the *(full) tensor product interpolation operator* $U_{\boldsymbol{k}} := \bigotimes_{m\in\mathbb{N}} U_{k_m}$ is defined by

$$(U_{\boldsymbol{k}} f)(\boldsymbol{\xi}) = \left(\bigotimes_{m\in\mathbb{N}} U_{k_m} f\right)(\boldsymbol{\xi}) = \sum_{\boldsymbol{i}\leq\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{i}}^{(\boldsymbol{k})}) L_{\boldsymbol{i}}^{(\boldsymbol{k})}(\boldsymbol{\xi}), \qquad f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}, \tag{5}$$

where $\boldsymbol{\xi}_{\boldsymbol{i}}^{(\boldsymbol{k})} \in \mathbb{R}^{\mathbb{N}}$ ranges over all points in the Cartesian product

$$\Xi^{(\boldsymbol{k})} := \bigtimes_{m\in\mathbb{N}} \Xi^{(k_m)}, \quad \text{with} \quad |\Xi^{(\boldsymbol{k})}| = \prod_{m\in\mathbb{N}} (1+k_m), \tag{6}$$

and where

$$L_{\boldsymbol{i}}^{(\boldsymbol{k})}(\boldsymbol{\xi}) := \prod_{m\in\mathbb{N}} L_{i_m}^{(k_m)}(\xi_m) \tag{7}$$

is a multivariate polynomial of (total) degree $|\boldsymbol{k}|_1 = \sum_m k_m$. Note that $L_0^{(0)}(\xi) \equiv 1$; in particular, since $\boldsymbol{k} \in \mathcal{F}$ all but a finite number of factors in (6) and (7) are equal to one so that the corresponding products can be regarded as finite. The tensor product interpolation operator $U_{\boldsymbol{k}}$ maps into the multivariate (tensor product) polynomial space

$$\mathcal{Q}_{\boldsymbol{k}} := \operatorname{span}\{\boldsymbol{\xi}^{\boldsymbol{i}} : 0 \leq i_m \leq k_m, m \in \mathbb{N}\}, \qquad \boldsymbol{k} \in \mathcal{F}. \tag{8}$$

Note that, since both the univariate polynomial sets of Lagrange fundamental polynomials $\{L_i^{(k)}\}_{i=0}^k$ and the Hermite orthonormal polynomials $\{H_i\}_{i=0}^k$ form a basis of $\mathcal{P}_k$, equivalent characterizations are

$$\begin{aligned}\mathcal{Q}_{\boldsymbol{k}} &= \operatorname{span}\{L_{\boldsymbol{i}}^{(\boldsymbol{k})} : 0 \leq i_m \leq k_m, m \in \mathbb{N}\} \\ &= \operatorname{span}\{H_{\boldsymbol{i}} : 0 \leq i_m \leq k_m, m \in \mathbb{N}\}, \qquad \boldsymbol{k} \in \mathcal{F}.\end{aligned}$$

In order for the tensor product interpolation operator $U_{\boldsymbol{k}}$ to be applicable also to functions defined only on a subset $\Gamma \subset \mathbb{R}^{\mathbb{N}}$, we assume the interpolation nodes to all lie in Γ:

Assumption A2. Let $\Gamma \subset \mathbb{R}^{\mathbb{N}}$ denote the domain from Assumption A1. For all $\boldsymbol{k} \in \mathcal{F}$ the Cartesian products of nodal sets $\Xi^{(\boldsymbol{k})}$ given in (6) satisfy $\Xi^{(\boldsymbol{k})} \subset \Gamma$.

In the following we denote by $\mathbb{R}^{\Gamma}$ the set of all mappings from Γ to $\mathbb{R}$. In analogy to (5) we define for any multi-index $\boldsymbol{k} \in \mathcal{F}$ the tensorized detail operator

$$\Delta_{\boldsymbol{k}} := \bigotimes_{m \in \mathbb{N}} \Delta_{k_m} : \mathbb{R}^{\Gamma} \to \mathcal{Q}_{\boldsymbol{k}}.$$

Finally, we associate with a finite subset $\Lambda \subset \mathcal{F}$ the multivariate polynomial space

$$\mathcal{P}_{\Lambda} := \sum_{\boldsymbol{i} \in \Lambda} \mathcal{Q}_{\boldsymbol{i}} \tag{9}$$

and define the associated *sparse (polynomial) collocation operator* $U_{\Lambda} : \mathbb{R}^{\Gamma} \to \mathcal{P}_{\Lambda}$ by

$$U_{\Lambda} := \sum_{\boldsymbol{i} \in \Lambda} \Delta_{\boldsymbol{i}}. \tag{10}$$

We will see that U_{Λ} is exact on $\mathcal{P}_{\Lambda}$ under some natural assumptions on the multi-index set Λ, for which we first recall some basic definitions given in [13, 11, 12].

Partial orderings and monotone sets of multi-indices We define a partial ordering on $\mathcal{F}$ by

$$\tilde{\boldsymbol{\nu}} \leq \boldsymbol{\nu} \quad :\Leftrightarrow \quad \tilde{\nu}_m \leq \nu_m \ \forall m \in \mathbb{N}$$

as well as

$$\tilde{\boldsymbol{\nu}} < \boldsymbol{\nu} \quad :\Leftrightarrow \quad \tilde{\boldsymbol{\nu}} \leq \boldsymbol{\nu} \text{ and } \tilde{\nu}_m < \nu_m \text{ for at least one } m \in \mathbb{N}$$

and introduce the relation

$$\tilde{\boldsymbol{\nu}} \not\leq \boldsymbol{\nu} \quad :\Leftrightarrow \quad \tilde{\nu}_m > \nu_m \text{ for at least one } m \in \mathbb{N}.$$

We shall call a set of multi-indices $\Lambda \subset \mathcal{F}$ *monotone* if $\boldsymbol{\nu} \in \Lambda$ and $\tilde{\boldsymbol{\nu}} \leq \boldsymbol{\nu}$ together imply that also $\tilde{\boldsymbol{\nu}} \in \Lambda$. Finally, for a multi-index $\boldsymbol{\nu} \in \mathcal{F}$ we define its *rectangular envelope* $\mathcal{R}_{\boldsymbol{\nu}}$ by

$$\mathcal{R}_{\boldsymbol{\nu}} := \{\tilde{\boldsymbol{\nu}} \in \mathcal{F} : \tilde{\boldsymbol{\nu}} \leq \boldsymbol{\nu}\}.$$

Note that $\mathcal{R}_{\boldsymbol{\nu}}$ for $\boldsymbol{\nu} \in \mathcal{F}$ is a finite (and monotone) set with cardinality

$$|\mathcal{R}_{\boldsymbol{\nu}}| = \prod_{m \in \mathbb{N}} (1 + \nu_m) < \infty. \tag{11}$$

2.2 Polynomial Exactness of Sparse Collocation

The introduction of the rectangular envelope $\mathcal{R}_{\boldsymbol{\nu}}$ of a multi-index $\boldsymbol{\nu} \in \mathcal{F}$ permits a convenient characterization of monotone multi-index sets Λ and the associated polynomial space $\mathcal{P}_{\Lambda}$ introduced in (9).

Proposition 2. If $\Lambda \subset \mathcal{F}$ is monotone, then

$$\Lambda = \bigcup_{\boldsymbol{\nu} \in \Lambda} \mathcal{R}_{\boldsymbol{\nu}} \quad \text{and} \quad \mathcal{P}_{\Lambda} = \operatorname{span}\{\boldsymbol{\xi}^{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda\} = \operatorname{span}\{H_{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda\}.$$

Proof. Since $\boldsymbol{\nu} \in \mathcal{R}_{\boldsymbol{\nu}}$ for all $\boldsymbol{\nu} \in \Lambda$ the set on the left is obviously a subset of that on the right. Conversely, given $\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}$ for some $\boldsymbol{\nu} \in \Lambda$, the definition of $\mathcal{R}_{\boldsymbol{\nu}}$ implies $\boldsymbol{i} \leq \boldsymbol{\nu}$, which in turn implies $\boldsymbol{i} \in \Lambda$ by the monotonicity of Λ. Moreover, monontonicity also implies

$$\mathcal{P}_\Lambda = \sum_{\boldsymbol{k}\in\Lambda} \mathcal{Q}_{\boldsymbol{k}} = \operatorname{span}\{\boldsymbol{\xi}^{\boldsymbol{i}} : \boldsymbol{i} \leq \boldsymbol{k}, \boldsymbol{k} \in \Lambda\} = \operatorname{span}\{\boldsymbol{\xi}^{\boldsymbol{i}} : \boldsymbol{i} \in \Lambda\} = \operatorname{span}\{H_{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda\},$$

where monontonicity is required for the two last equalities. □

In view of Proposition 2, $\mathcal{P}_\Lambda$ for a multi-index set $\Lambda \subset \mathcal{R}_{\boldsymbol{k}}$ represents a sparsification of $\mathcal{Q}_{\boldsymbol{k}}$. In particular, the full tensor product polynomial space $\mathcal{Q}_{\boldsymbol{k}}$ coincides with $\mathcal{P}_\Lambda$ for $\Lambda = \mathcal{R}_{\boldsymbol{k}}$. Similarly, the full tensor approximation operator $U_{\boldsymbol{k}}$ defined in (5) can be expressed as $U_{\boldsymbol{k}} = \sum_{\boldsymbol{i}\in\mathcal{R}_{\boldsymbol{k}}} \Delta_{\boldsymbol{i}}$.

Proposition 3. Let $\Lambda \subset \mathcal{F}$ be a finite and monotone set. Then $U_\Lambda p = p$ for all $p \in \mathcal{P}_\Lambda$. In particular, for all $p \in \mathcal{P}_\Lambda$ we have $\Delta_{\boldsymbol{i}} p = 0$ for $\boldsymbol{i} \notin \Lambda$.

Proof. Observe first that, for any $\boldsymbol{\nu}, \boldsymbol{i} \in \mathcal{F}$ such that $\boldsymbol{i} \not\leq \boldsymbol{\nu}$ we have

$$\Delta_{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{\nu}} = \prod_{m\in\mathbb{N}} \Delta_{i_m}\xi_m^{\nu_m} = \prod_{m\in\mathbb{N}} \underbrace{(U_{i_m} - U_{i_m-1})\xi_m^{\nu_m}}_{=\xi_m^{\nu_m}-\xi_m^{\nu_m}\equiv 0 \text{ for at least one } m} = 0.$$

It suffices to prove the assertions for all monomials $\boldsymbol{\xi}^{\boldsymbol{\nu}}$ in $\mathcal{P}_\Lambda$. For $\boldsymbol{\nu} \in \Lambda$ any $\boldsymbol{i} \in \mathcal{F} \setminus \Lambda$ must satisfy $\boldsymbol{i} \not\leq \boldsymbol{\nu}$ and therefore $\Delta_{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{\nu}} = 0$, proving the second assertion. We conclude that

$$U_\Lambda\boldsymbol{\xi}^{\boldsymbol{\nu}} = \sum_{\boldsymbol{i}\in\Lambda} \Delta_{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{\nu}} = \sum_{\boldsymbol{i}\in\Lambda\cap\mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{\nu}} = \sum_{\boldsymbol{i}\in\mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{\nu}},$$

where the third equality follows from the fact that $\mathcal{R}_{\boldsymbol{\nu}} \subseteq \Lambda$ for all $\boldsymbol{\nu} \in \Lambda$ due to the monotonicity of Λ. The proof concludes with

$$\begin{aligned} U_\Lambda\boldsymbol{\xi}^{\boldsymbol{\nu}} = \sum_{\boldsymbol{i}\in\mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}}\boldsymbol{\xi}^{\boldsymbol{\nu}} = \sum_{\boldsymbol{i}\in\mathcal{R}_{\boldsymbol{\nu}}} \left(\prod_{m\in\mathbb{N}} \Delta_{i_m}\xi_m^{\nu_m}\right) = \prod_{m\in\mathbb{N}} \left(\sum_{i_m=0}^{\nu_m} \Delta_{i_m}\xi_m^{\nu_m}\right) = \prod_{m\in\mathbb{N}} U_{\nu_m}\xi_m^{\nu_m} \\ = \prod_{m\in\mathbb{N}} \xi_m^{\nu_m} = \boldsymbol{\xi}^{\boldsymbol{\nu}}. \end{aligned}$$

Note that the third equality is obtained by rewriting a (finite) product of sums: since $\boldsymbol{\nu} \in \mathcal{F}$ there exists an $M \in \mathbb{N}$ such that $\nu_m = 0$ for $m > M$. For such m we have $\Delta_{i_m}^{\nu_m}\xi_m^{\nu_m} = \Delta_0\xi_m^0 \equiv 1$ and therefore

$$\begin{aligned} \prod_{m\in\mathbb{N}} \left(\sum_{i_m=0}^{\nu_m} \Delta_{i_m}\xi_m^{\nu_m}\right) &= (\Delta_0\xi_1^{\nu_1} + \cdots + \Delta_{\nu_1}\xi_1^{\nu_1}) \cdots \left(\Delta_0\xi_M^{\nu_M} + \cdots + \Delta_{\nu_M}\xi_M^{\nu_M}\right) \\ &= \sum_{\substack{\boldsymbol{i}\in\mathbb{N}_0^M \\ i_m\leq\nu_m}} \Delta_{i_1}\xi_1^{\nu_1} \cdots \Delta_{i_M}\xi_M^{\nu_M} = \sum_{\boldsymbol{i}\in\mathcal{R}_{\boldsymbol{\nu}}} \left(\prod_{m\in\mathbb{N}} \Delta_{i_m}\xi_m^{\nu_m}\right). \end{aligned}$$

□

Proposition 3 can be seen as an extension of [6, Proposition 1] to general monotone multi-index sets as well as an extension of [13, Theorem 6.1] and [11, Theorem 2.1] to interpolation operators U_i with non-nested node sets. As mentioned in [13, p. 89], if the set Λ is not monotone then U_Λ will not be exact on $\mathcal{P}_\Lambda$ in general. However, the exactness on $\mathcal{P}_\Lambda$ is a crucial property in the subsequent convergence analysis and we therefore choose to work exclusively with monotone sets Λ.

2.3 Sparse Grid Associated with U_Λ

The construction of $U_\Lambda f$ for $f\colon \Gamma \to \mathbb{R}$ consists of a linear combination of tensor product interpolation operators requiring the evaluation of f at certain multivariate nodes. We shall refer to the collection of these nodes as the *sparse grid* $\Xi_\Lambda \subset \Gamma$ associated with Λ. For a monotone and finite set $\Lambda \subset \mathcal{F}$ there holds

$$\Xi_\Lambda = \bigcup_{\boldsymbol{i}\in\Lambda} \Xi^{(\boldsymbol{i})}, \tag{12}$$

because for $\boldsymbol{i} \in \mathcal{F}$ we have

$$\Delta_{\boldsymbol{i}} f = \Big[\bigotimes_{m\geq 1}(U_{i_m} - U_{i_m-1})\Big] f = \sum_{\boldsymbol{i}-\mathbf{1}\leq\boldsymbol{k}\leq\boldsymbol{i}} (-1)^{|\boldsymbol{i}-\boldsymbol{k}|_1}\Big[\bigotimes_{m\geq 1} U_{k_m}\Big] f,$$

i.e., for computing $\Delta_{\boldsymbol{i}} f$ we need to evaluate f at

$$\Xi^{(\boldsymbol{i}),\Delta} := \bigcup_{\boldsymbol{i}-\mathbf{1}\leq\boldsymbol{k}\leq\boldsymbol{i}} \Xi^{(\boldsymbol{k})}.$$

Since Λ is a monotone set, the resulting sparse grid for $U_\Lambda = \sum_{\boldsymbol{i}\in\Lambda} \Delta_{\boldsymbol{i}}$ is

$$\Xi_\Lambda = \bigcup_{\boldsymbol{i}\in\Lambda} \Xi^{(\boldsymbol{i}),\Delta} = \bigcup_{\boldsymbol{i}\in\Lambda} \bigcup_{\boldsymbol{i}-\mathbf{1}\leq\boldsymbol{k}\leq\boldsymbol{i}} \Xi^{(\boldsymbol{k})} = \bigcup_{\boldsymbol{i}\in\Lambda} \Xi^{(\boldsymbol{i})}.$$

We remark that the unisolvence on $\mathcal{P}_\Lambda$ of point evaluation on Ξ_Λ is discussed in [13, Theorem 6.1].

3 Convergence Analysis

In this section we analyze the error

$$\|f - U_\Lambda f\|_{L^2_\mu}, \qquad f\colon \Gamma \to \mathcal{H},$$

where $\|\cdot\|_{L^2_\mu}$ denotes the norm in $L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})$, f is assumed to satisfy Assumption A1 and $\Lambda \subset \mathcal{F}$ is required to be monotone and finite. Our first goal here is to establish a convergence rate $s > 0$ for the error of U_{Λ_N} for a nested sequence Λ_N of monotone subsets of $\mathcal{F}$ with $|\Lambda_N| = N$, i.e.,

$$\|f - U_{\Lambda_N} f\|_{L^2_\mu} \leq C N^{-s}, \qquad f\colon \Gamma \to \mathcal{H}, \tag{13}$$

where $C < \infty$ may depend on f as well as the univariate nodal sets. The line of proof we present here follows and builds upon the works [9, 26, 4]. We complement this convergence rate with a bound on the number of collocation points associated with a given multi-index set.

3.1 General Convergence Results

The subsequent error analysis for the sparse collocation operator U_Λ is based on the representation of multivariate functions $f \in L^2_\mu(\mathbb{R}^{\mathbb{N}};\mathcal{H})$ in the orthonormal basis of multivariate Hermite polynomials $H_{\boldsymbol{\nu}}$. We shall therefore examine the worst-case approximation error of any U_Λ applied to a given multivariate Hermite basis polynomial $H_{\boldsymbol{\nu}}$. To this end we define

$$c_{\boldsymbol{\nu}} := \sup_{\Lambda\subset\mathcal{F},|\Lambda|<\infty} \|(I - U_\Lambda) H_{\boldsymbol{\nu}}\|_{L^2_\mu}, \qquad \boldsymbol{\nu} \in \mathcal{F}. \tag{14}$$

This quantity is finite since $\Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}} = 0$ for $\boldsymbol{i} \not\leq \boldsymbol{\nu}$ and hence

$$c_{\boldsymbol{\nu}} = \max_{\Lambda \subseteq \mathcal{R}_{\boldsymbol{\nu}}} \|(I - U_\Lambda) H_{\boldsymbol{\nu}}\|_{L^2_\mu},$$

where the maximum is taken over a finite set. The quantities $c_{\boldsymbol{\nu}}$ also measure the deviation of the error of oblique projection U_Λ from that of orthogonal projection, as these numbers would all be zero or one if U_Λ is replaced with the L^2_μ-orthogonal projection onto $\mathcal{P}_\Lambda$. Moreover, we obtain the following bound:

Proposition 4. For all $\boldsymbol{\nu} \in \mathcal{F}$ the quantity $c_{\boldsymbol{\nu}}$ defined in (14) satisfies

$$c_{\boldsymbol{\nu}} \leq \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \|\Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}}\|_{L^2_\mu}.$$

In particular, if for the univariate Hermite polynomials there exists $\theta \geq 0$ and $K \geq 1$ such that

$$\|\Delta_i H_\nu\|_{L^2_\mu} \leq (1 + K\nu)^\theta \qquad \text{for all } i \in \mathbb{N}_0, \tag{15}$$

where we have denoted the univariate Gaussian measure again by μ, then

$$c_{\boldsymbol{\nu}} \leq \prod_{m \in \mathbb{N}} (1 + K\nu_m)^{\theta+1}, \qquad \boldsymbol{\nu} \in \mathcal{F}. \tag{16}$$

Proof. In view of Proposition 3 we have $H_{\boldsymbol{\nu}} = U_{\boldsymbol{\nu}} H_{\boldsymbol{\nu}} = \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}}$ and, particularly, $\Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}} = 0$ for $\boldsymbol{i} \notin \mathcal{R}_{\boldsymbol{\nu}}$, since $H_{\boldsymbol{\nu}} \in \mathcal{P}_{\mathcal{R}_{\boldsymbol{\nu}}}$. Therefore

$$\begin{aligned}(I - U_\Lambda) H_{\boldsymbol{\nu}} &= \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}} - \sum_{\boldsymbol{i} \in \Lambda} \Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}} = \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}} - \sum_{\boldsymbol{i} \in \Lambda \cap \mathcal{R}_{\boldsymbol{\nu}}} \Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}} \\ &= \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}} \setminus \Lambda} \Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}},\end{aligned}$$

giving

$$c_{\boldsymbol{\nu}} = \max_{\Lambda \subseteq \mathcal{R}_{\boldsymbol{\nu}}} \|(I - U_\Lambda) H_{\boldsymbol{\nu}}\|_{L^2_\mu} \leq \max_{\Lambda \subseteq \mathcal{R}_{\boldsymbol{\nu}}} \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}} \setminus \Lambda} \|\Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}}\|_{L^2_\mu} \leq \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \|\Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}}\|_{L^2_\mu}.$$

Moreover, if (15) holds, then

$$\begin{aligned}c_{\boldsymbol{\nu}} &\leq \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \|\Delta_{\boldsymbol{i}} H_{\boldsymbol{\nu}}\|_{L^2_\mu} = \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \prod_{m \in \mathbb{N}} \|\Delta_{i_m} H_{\nu_m}\|_{L^2_\mu} \leq \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}} \prod_{m \in \mathbb{N}} (1 + K\nu_m)^\theta \\ &= |\mathcal{R}_{\boldsymbol{\nu}}| \prod_{m \in \mathbb{N}} (1 + K\nu_m)^\theta \leq \prod_{m \in \mathbb{N}} (1 + K\nu_m)^{\theta+1}.\end{aligned}$$

where we have used (11) and $K \geq 1$ in the last inequality. □

Remark 5. Bounds such as (15) can often be found in the sparse collocation or sparse quadrature literature, e.g., for quadrature operators applied to Hermite polynomials [9], norms of quadrature operators on bounded domains [39] or Lebesgue constants for Leja points [12]. Numerical estimates for the specific case of Genz-Keister points have been provided in [7].

The following lemma provides a natural starting point for bounding the approximation error of $U_\Lambda f$ for monotone subsets Λ. The proof follows the same line of argument as the proof of [9, Lemma 3.2].

Lemma 6 (cf. [9, Lemma 3.2])**.** For a finite and monotone subset $\Lambda \subset \mathcal{F}$ there holds

$$\|f - U_\Lambda f\|_{L^2_\mu} \leq \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda} c_{\boldsymbol{\nu}} \|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}. \tag{17}$$

Proof. Due to the monotonicity of Λ we can apply Proposition 3 and obtain

$$\begin{aligned}\|f - U_\Lambda f\|_{L^2_\mu} &= \Big\| \sum_{\boldsymbol{\nu}\in\mathcal{F}} f_{\boldsymbol{\nu}} \, (I - U_\Lambda) H_{\boldsymbol{\nu}}(\boldsymbol{\xi}) \Big\|_{L^2_\mu} = \Big\| \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda} f_{\boldsymbol{\nu}} \, (I - U_\Lambda) H_{\boldsymbol{\nu}}(\boldsymbol{\xi}) \Big\|_{L^2_\mu} \\ &\leq \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda} \|f_{\boldsymbol{\nu}}\|_{\mathcal{H}} \|(I - U_\Lambda) H_{\boldsymbol{\nu}}\|_{L^2_\mu} \leq \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda} c_{\boldsymbol{\nu}} \|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}.\end{aligned}$$

□

Building on Lemma 6 the approximation error $\|f - U_\Lambda f\|_{L^2_\mu}$ may be further bounded given summability results for the sequence $(c_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}})_{\boldsymbol{\nu}\in\mathcal{F}}$. The key result here is known as *Stechkin's lemma* which provides a decay rate for the ℓ^q-tail of an p-summable sequence for $q > p$ and is due to Stechkin [41] for $q = 2$ (cf. also [13, Lemma 3.6]).

Lemma 7 (Stechkin)**.** Let $0 < p < q < \infty$ and let

$$(a_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F}) := \left\{ (b_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \colon \sum_{\boldsymbol{\nu}\in\mathcal{F}} |b_{\boldsymbol{\nu}}|^p < \infty \right\}$$

be a sequence of nonnegative numbers. Then for Λ_N denoting the set of multi-indices $\boldsymbol{\nu}$ corresponding to the N largest elements $a_{\boldsymbol{\nu}}$, there holds

$$\left(\sum_{\boldsymbol{\nu}\notin\Lambda_N} a_{\boldsymbol{\nu}}^q \right)^{1/q} \leq \|(a_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^p} (N+1)^{-s}, \qquad s = \frac{1}{p} - \frac{1}{q}. \tag{18}$$

The index sets Λ_N in Stechkin's lemma associated with the N largest sequence elements are not necessarily monotone and, therefore Lemma 6 and Lemma 7 can not be combined to bound the error without additional assumptions. An obvious way to ensure monotonicity of the sets Λ_N in Stechkin's lemma is to assume the sequence $(a_{\boldsymbol{\nu}})$ to be *nonincreasing*, i.e.,

$$\boldsymbol{\nu} \leq \tilde{\boldsymbol{\nu}} \quad \Rightarrow \quad a_{\boldsymbol{\nu}} \geq a_{\tilde{\boldsymbol{\nu}}}.$$

This leads to

Theorem 8. Let Assumptions A1 and A2 be satisfied and let there exist a nonincreasing sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$ with $p \in (0,1)$ such that

$$c_{\boldsymbol{\nu}} \|f_{\boldsymbol{\nu}}\|_{\mathcal{H}} \leq \hat{c}_{\boldsymbol{\nu}} \qquad \forall \boldsymbol{\nu} \in \mathcal{F}.$$

Then there exists a nested sequence $(\Lambda_N)_{N\in\mathbb{N}}$ of finite and monotone subsets $\Lambda_N \subset \mathcal{F}$ with $|\Lambda_N| = N$ such that (13) holds with rate $s = 1/p - 1$.

We will provide a proof below. The convergence analysis in [12, 39] for sparse quadrature and interpolation in case of bounded Γ follows Theorem 8, although sometimes hidden in the details. There the authors employ explicit bounds on the norms of the Legendre or Taylor coefficients of $f : \Gamma \to \mathcal{H}$ to construct a dominating and nonincreasing sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$, $p \in (0,1)$.

In our setting it is, however, not always possible to derive explicit bounds on the norm of the Hermite coefficients $\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}$. In [4] a technique was developed which relies on somewhat implicit bounds on $\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}$ via a weighted ℓ^2-summability property. We adapt this approach to the current setting in

Theorem 9. Let Assumptions A1 and A2 be satisfied and let there exist a sequence $(b_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ of positive numbers such that

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}} b_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}^2 < \infty \tag{19}$$

as well as another nonincreasing sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$, $p \in (0,2)$, for which

$$\frac{c_{\boldsymbol{\nu}}}{b_{\boldsymbol{\nu}}^{1/2}} \le \hat{c}_{\boldsymbol{\nu}} \qquad \forall \boldsymbol{\nu} \in \mathcal{F}.$$

Then there exists a nested sequence $(\Lambda_N)_{N\in\mathbb{N}}$ of finite and monotone subsets $\Lambda_N \subset \mathcal{F}$ with $|\Lambda_N| = N$ such that (13) holds with rate $s = 1/p - 1/2$.

Proof of Theorem 8 and Theorem 9. Let Λ_N be the set of multi-indices $\boldsymbol{\nu}$ corresponding to the N largest elements of $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$. Then each Λ_N is monotone and the sequence $(\Lambda_N)_{N\in\mathbb{N}}$ can be chosen to be nested.

If the assumption of Theorem 8 hold, we can apply Lemma 6 and Stechkin's lemma with $q = 1 > p$ to obtain

$$\|f - U_{\Lambda_N} f\|_{L^2_\mu} \le \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} c_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}} \le \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} \hat{c}_{\boldsymbol{\nu}} \le C(N+1)^{-(1/p-1)}$$

where $C = \|(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^p}$.

If the assumptions of Theorem 9 hold, Lemma 6 combined with the Cauchy-Schwarz inequality and Stechkin's lemma for $q = 2 > p$ give

$$\begin{aligned}
\|f - U_{\Lambda_N} f\|_{L^2_\mu} &\le \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} c_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}} = \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} \left(\frac{c_{\boldsymbol{\nu}}}{b_{\boldsymbol{\nu}}^{1/2}}\right)\left(b_{\boldsymbol{\nu}}^{1/2}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}\right) \\
&\le \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} b_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}^2\right)^{1/2} \cdot \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} \frac{c_{\boldsymbol{\nu}}^2}{b_{\boldsymbol{\nu}}}\right)^{1/2} \\
&\le \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}} b_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}^2\right)^{1/2} \cdot \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} \hat{c}_{\boldsymbol{\nu}}^2\right)^{1/2} \\
&\le C(N+1)^{-(1/p-1/2)},
\end{aligned}$$

where now $C = \|(b_{\boldsymbol{\nu}}^{1/2}\|f_{\boldsymbol{\nu}}\|)_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^2} \cdot \|(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\|_{\ell^p}$, respectively. □

Remark 10. Another application of the weighted ℓ^2-summability property (19) is the analysis of sparse quadrature given in [9], where the author employs the slightly different estimate

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} c_{\boldsymbol{\nu}}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}} \le \sup_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} b_{\boldsymbol{\nu}}^{q-1/2} \sum_{\boldsymbol{\nu}\in\mathcal{F}\setminus\Lambda_N} \frac{c_{\boldsymbol{\nu}}}{b_{\boldsymbol{\nu}}^{-q}}\, b_{\boldsymbol{\nu}}^{1/2}\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}.$$

After showing that the series on the right is bounded and applying Stechkin's lemma to $(b_{\boldsymbol{\nu}}^{q-1/2})_{\boldsymbol{\nu}\in\mathcal{F}}$, this yields the same convergence rate as stated in Theorem 9.

Remark 11. We mention that sparse collocation attains a smaller convergence rate than best N-term approximation in case the assumptions of Theorem 9 hold. Namely, under these assumptions the best N-term rate is $s = \frac{1}{p}$, see [4, Theorem 1.2]. This reduced convergence rate is not caused by the additional factors $c_{\boldsymbol{\nu}}$ in the error analysis of sparse collocation. The reason for the slower rate is missing orthogonality: in the proof of Lemma 6 we could not apply Parseval's identity and had to use the triangle inequality to bound the error. This led to bounds in terms of $\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}$ rather than $\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}^2$ as in the case of orthogonal projections, e.g., best N-term approximations.

We emphasize that the construction of such a nonincreasing, p-summable dominating sequence is by no means trivial. Without the first property we can not conclude that the multi-index sets Λ_N occurring in Stechkin's lemma are monotone, which in turn is needed to use Lemma 6 as the starting point of our error analysis. Of course, we could consider monotone envelopes $\Lambda_N \subset \tilde{\Lambda}_N$ of Λ_N, but their size can grow quite rapidly with N (e.g., polynomially or even faster, see counterexample below). Moreover, it is not at all obvious that for a sequence $(a_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$ there exists a dominating and nonincreasing $(\hat{a}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$. In particular, we provide the following counterexample: let $\mathcal{F} = \mathbb{N}$ and define a_n, $n \in \mathbb{N}$ by

$$a_n = \begin{cases} \frac{1}{m^2}, & n = \sum_{k=1}^m k, \\ 0, & \text{otherwise}, \end{cases}$$

i.e., $a_1 = 1, a_2 = 0, a_3 = \frac{1}{4}, a_4 = 0, a_5 = 0, a_6 = \frac{1}{9}, a_7 = 0, \ldots, a_9 = 0, a_{10} = \frac{1}{16}, a_{11} = 0, \ldots$. Then $(a_n)_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$. The smallest positive nonincreasing sequence $(\hat{a}_n)_{n\in\mathbb{N}}$ dominating $(a_n)_{n\in\mathbb{N}}$ is given by $\hat{a}_n := \sup_{m\geq n} |a_m|$, see [13, Section 3.8]. In our case, we get

$$\hat{a}_n = \frac{1}{m^2} \qquad \text{for each } n \text{ such that} \qquad 1 + \sum_{k=1}^{m-1} k \leq n \leq \sum_{k=1}^{m} k$$

and, thus,

$$\sum_{n=1}^{\infty} |\hat{a}_n| = \sum_{m=1}^{\infty} m \frac{1}{m^2} = \infty.$$

Although the example is somewhat pathological, it illustrates that for $(a_{\boldsymbol{\nu}}) \in \ell^p(\mathcal{F})$ a p-summable nonincreasing dominating sequence need not exist.

3.2 Sufficient Conditions for Weighted Summability and Majorization

We now follow the strategy of Theorem 9 and study under which requirements the assumptions of Theorem 9 hold. To this end we recall a result from [4] for weighted ℓ^2-summability of Hermite coefficients $\|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}$ given the following smoothness conditions on f:

Assumption A3. Let f satisfy Assumption A1. There exists an integer $r \in \mathbb{N}_0$ and a sequence of positive numbers $(\tau_m^{-1})_{m\in\mathbb{N}} \in \ell^p(\mathbb{N})$, $p \in (0, 2)$, such that

(a) for any $\boldsymbol{\alpha} \in \mathcal{F}$ with $|\boldsymbol{\alpha}|_\infty \leq r$ the (weak) partial derivative $\partial^{\boldsymbol{\alpha}} f$ exists and satisfies $\partial^{\boldsymbol{\alpha}} f \in L^2_\mu(\mathbb{R}^{\mathbb{N}}; \mathcal{H})$,

(b) there holds

$$\sum_{|\boldsymbol{\alpha}|_\infty \leq r} \frac{\boldsymbol{\tau}^{2\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \|\partial^{\boldsymbol{\alpha}} f\|^2_{L^2_\mu} < \infty, \tag{20}$$

where $\boldsymbol{\tau}^{\boldsymbol{\alpha}} = \prod_{m=1}^{\infty} \tau_m^{\alpha_m}$ and $\boldsymbol{\alpha}! = \prod_{m=1}^{\infty} \alpha_m!$.

Observe that the sum in (20) is actually a series, because $\boldsymbol{\alpha}$ has infinitly many components and therefore there are countably many vectors such that $|\boldsymbol{\alpha}|_\infty \leq r$. Assumption A3(a) states that we require a *finite* order of partial differentiability of f, i.e., up to order r with respect to each variable ξ_m, and, maybe more importantly, Assumption A3(b) asks for a *weighted square-summability* of the L^2_μ-norms of the corresponding partial derivatives. The latter, in particular, implies bounds of the form

$$\|\partial^{\boldsymbol{\alpha}} f\|_{L^2_\mu} \leq K \sqrt{\boldsymbol{\alpha}!}\, \boldsymbol{\tau}^{-\boldsymbol{\alpha}}, \qquad |\boldsymbol{\alpha}|_0 \leq r,$$

since otherwise the summability requirement (20) would not hold. Recalling that $(\tau_m^{-1})_{m\in\mathbb{N}} \in \ell^p(\mathbb{N})$ this bound implies that, e.g., the L^2_μ-norm of the derivative $\partial^\alpha_{\xi_m} f$, $\alpha \leq r$, decays if $m \to \infty$.

The following result shows that the smoothness condition of Assumption A3 implies the first condition (19) of Theorem 9:

Theorem 12 (cf. [4, Theorem 3.1])**.** Let Assumption A3 be satisfied. Then, with the weights

$$b_{\boldsymbol{\nu}} = b_{\boldsymbol{\nu}}(\boldsymbol{\tau}, r) = \sum_{|\boldsymbol{\alpha}|_\infty \le r} \binom{\boldsymbol{\nu}}{\boldsymbol{\alpha}} \boldsymbol{\tau}^{2\boldsymbol{\alpha}} = \prod_{m\ge 1} \left(\sum_{l=0}^{r} \binom{\nu_m}{l} \tau_m^{2l} \right), \qquad \boldsymbol{\nu} \in \mathcal{F}, \tag{21}$$

where

$$\binom{\boldsymbol{\nu}}{\boldsymbol{\alpha}} := \prod_{m\ge 1} \binom{\nu_m}{\alpha_m} \quad \text{and} \quad \binom{\nu_m}{\alpha_m} := 0 \ \text{ if } \ \alpha_m > \nu_m,$$

there holds

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}} b_{\boldsymbol{\nu}} \|f_{\boldsymbol{\nu}}\|_{\mathcal{H}}^2 = \sum_{|\boldsymbol{\alpha}|_\infty \le r} \frac{\boldsymbol{\tau}^{2\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \|\partial^{\boldsymbol{\alpha}} f\|_{L^2_\mu}^2 < \infty. \tag{22}$$

(We mention in passing that in [4] the assertion of Theorem 12 was actually proven without requiring that both series in (22) be finite.) To apply Theorem 9 it remains to prove the existence of a nonincreasing and p-summable sequence which dominates $c_{\boldsymbol{\nu}}/b_{\boldsymbol{\nu}}^{1/2}$, $\boldsymbol{\nu} \in \mathcal{F}$. Since the $b_{\boldsymbol{\nu}}$ are explicitly given in (21), this boils down to the question, how fast the projection errors $c_{\boldsymbol{\nu}}$ are allowed to grow. As it turns out, a polynomial growth w.r.t. $\boldsymbol{\nu}$ as given in (16) in Proposition 4 is sufficient. We therefore state the following lemma, which is strongly based on the techniques developed in the proofs of [4, Lemma 5.1] and [9, Lemma 3.4].

Lemma 13. Let there exists a $\theta \ge 0$ and a $K \ge 1$ such that

$$c_{\boldsymbol{\nu}} \le \prod_{m\ge 1}^{\infty} (1 + K\nu_m)^{\theta+1}, \qquad \boldsymbol{\nu} \in \mathcal{F}.$$

Then for any increasing sequence $(\tau_m)_{m\in\mathbb{N}}$ such that $\sum_{m\ge 1} \tau_m^{-p} < \infty$ for a $p > 0$ and for any $r > 2(\theta + 1) + \frac{2}{p}$ there exists a nonincreasing sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$ such that

$$\frac{c_{\boldsymbol{\nu}}}{b_{\boldsymbol{\nu}}^{1/2}} \le \hat{c}_{\boldsymbol{\nu}} \qquad \forall \boldsymbol{\nu} \in \mathcal{F},$$

where $b_{\boldsymbol{\nu}} = b_{\boldsymbol{\nu}}(\tau, r)$ is as in (21).

Proof. We start with constructing the dominating sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ and show afterwards that it belongs to $\ell^p(\mathcal{F})$ and is nonincreasing. In the following we use the notation $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$.

Step 1: Constructing $\hat{c}_{\boldsymbol{\nu}}$ We get due to

$$\binom{\nu_m}{\nu_m \wedge r} \tau_m^{2(\nu_m \wedge r)} \le \binom{\nu_m}{r} \tau_m^{2r} \le \sum_{l=0}^{r} \binom{\nu_m}{l} \tau_m^{2l}$$

that

$$\frac{c_{\boldsymbol{\nu}}^2}{b_{\boldsymbol{\nu}}} \le \prod_{m\ge 1} \frac{(1 + K\nu_m)^{2(\theta+1)}}{\sum_{l=0}^{r} \binom{\nu_m}{l} \tau_m^{2l}} \le \prod_{m\ge 1} \frac{(1 + K\nu_m)^{2\theta+2}}{\binom{\nu_m}{\nu_m\wedge r} \tau_m^{2(\nu_m\wedge r)}} = \prod_{m\ge 1} \tau_m^{-2(\nu_m\wedge r)} h(\nu_m) \tag{23}$$

where we defined the auxiliary function $h(n) := \frac{(1+Kn)^{2\theta+2}}{\binom{n}{n\wedge r}}$, $n \in \mathbb{N}$. We will now derive bounds for $h(n)$ as well as for $\tau_m^{-2(\nu_m\wedge r)}$ in order to construct a dominating sequence $\hat{c}_{\boldsymbol{\nu}}$.

For $n \le r$ we get $h(n) = (1+Kn)^{2\theta+2}$, but for $n > r$ holds

$$h(n) = \frac{(1+Kn)^{2\theta+2}}{\binom{n}{r}} = \frac{r!\,(1+Kn)^{2\theta+2}}{(n+1)\cdots(n+r)}.$$

Thus, we have $h \in \mathcal{O}(n^{2\theta+2-r})$, i.e., there exists a $C_h \in [1,\infty)$ such that

$$h(n) \le C_h n^{2\theta+2-r} =: \hat{h}(n) \qquad \forall n \in \mathbb{N}.$$

By setting $\hat{h}(0) := 1 = h(0)$, we get $h(n) \le \hat{h}(n)$ for all $n \in \mathbb{N}_0$.

Furthermore, since $(\tau_m^{-1})_{m\in\mathbb{N}} \in \ell^p(\mathbb{N})$ we have $\tau_m \to \infty$ as $m \to \infty$. Thus, there exists an $M \in \mathbb{N}$ such that $\tau_m \ge \sqrt{C_h}$ for $m \ge M$ and $\tau_m \le \sqrt{C_h}$ for $m < M$. We define

$$\hat{\tau}_m := \sqrt{C_h} \vee \tau_m, \qquad m \in \mathbb{N},$$

and notice that $\hat{\tau}_m \ge 1$ as well as $(\hat{\tau}_m^{-1})_{m\in\mathbb{N}} \in \ell^p(\mathbb{N})$ by assumption. Moreover, we obtain for $m \ge M$

$$\tau_m^{2(\nu_m \wedge r)} = \hat{\tau}_m^{2(\nu_m\wedge r)} \ge \hat{\tau}_m^{2(\nu_m \wedge 1)}, \qquad \forall \nu_m \in \mathbb{N}_0,$$

since $\tau_m = \hat{\tau}_m \ge \sqrt{C_h} \ge 1$ in this case. Further, let us define

$$C_\tau := \min_{m\ge 1} \min_{n=0,\dots,r} \frac{\tau_m^{2n}}{C_h^{n\wedge 1}} > 0$$

which then yields for $1 \le m < M$

$$\tau_m^{2(\nu_m\wedge r)} \ge C_\tau\, C_h^{\nu_m\wedge 1} = C_\tau\, \hat{\tau}_m^{2(\nu_m\wedge 1)}, \qquad \forall \nu_m \in \mathbb{N}_0$$

since $\hat{\tau}_m = \sqrt{C_h}$ for $m < M$. We now define

$$\hat{c}_{\boldsymbol{\nu}}^2 := C_\tau^{-M} \prod_{m\ge 1} \hat{\tau}_m^{-2(\nu_m\wedge 1)} \hat{h}(\nu_m). \tag{24}$$

and notice that $\hat{c}_{\boldsymbol{\nu}}^2$ dominates $\frac{c_{\boldsymbol{\nu}}^2}{b_{\boldsymbol{\nu}}}$ by (23).

Step 2: Show that $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$ As for the p-summability, there holds

$$\begin{aligned}\sum_{\boldsymbol{\nu}\in\mathcal{F}} \hat{c}_{\boldsymbol{\nu}}^p &= C_\tau^{-pM/2} \sum_{\boldsymbol{\nu}\in\mathcal{F}} \prod_{m\ge 1} \hat{\tau}_m^{-p(\nu_m\wedge 1)} \hat{h}^{p/2}(\nu_m) \\ &= C_\tau^{-pM/2} \prod_{m\ge 1} \sum_{n\ge 0} \hat{\tau}_m^{-p(n\wedge 1)} \hat{h}^{p/2}(n).\end{aligned}$$

We get

$$\sum_{n\ge 0} \hat{\tau}_m^{-p(n\wedge 1)} \hat{h}^{p/2}(n) = 1 + C_h^{p/2} \hat{\tau}_m^{-p} \underbrace{\sum_{n\ge 1} n^{-p(r-2\theta-2)/2}}_{=:S}$$

where the sum S is finite due to the assumption $\frac{p}{2}(r-2\theta-2) = \frac{p}{2}(r-2\theta-2) > 1$. The rest follows by using $\log(1+x) \le x$ for x positive in order to get

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}} \hat{c}_{\boldsymbol{\nu}}^p = C_\tau^{-pM/2} \prod_{m\ge 1} (1 + C_h^{p/2} S \hat{\tau}_m^{-p}) \le C_\tau^{-pM/2} \exp\left(C_h^{p/2} S \sum_{m\ge 1} \hat{\tau}_m^{-p} \right) < \infty$$

since $(\hat{\tau}_m^{-1})_{m\in\mathbb{N}}$ is in $\ell^p(\mathbb{N})$ by construction.

Step 3: Show that $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is nonincreasing Let $\boldsymbol{\nu}\in\mathcal{F}$ be arbitrary. If $m\in\operatorname{supp}\boldsymbol{\nu}=\{m\in\mathbb{N}:\nu_m>0\}$, then we get

$$\hat{c}^2_{\boldsymbol{\nu}+\boldsymbol{e}_m}=\hat{c}^2_{\boldsymbol{\nu}}\cdot\frac{\hat{h}(\nu_m+1)}{\hat{h}(\nu_m)}\le\hat{c}^2_{\boldsymbol{\nu}},$$

since $\hat{h}(n)$ is nonincreasing for $n\ge 1$. Let now $m\notin\operatorname{supp}\boldsymbol{\nu}$. Then

$$\hat{c}^2_{\boldsymbol{\nu}+\boldsymbol{e}_m}=\hat{c}^2_{\boldsymbol{\nu}}\cdot\hat{\tau}_m^{-2}\cdot\hat{h}(1)=\hat{c}^2_{\boldsymbol{\nu}}\cdot C_h\hat{\tau}_m^{-2}\le\hat{c}^2_{\boldsymbol{\nu}}\cdot C_h(\sqrt{C_h})^{-2}\le\hat{c}^2_{\boldsymbol{\nu}}.$$

In summary, we obtain

$$\hat{c}_{\boldsymbol{\nu}+\boldsymbol{e}_m}\le\hat{c}_{\boldsymbol{\nu}}\qquad\forall m\in\mathbb{N},$$

hence, $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ is nonincreasing. □

We can now state our main convergence result for sparse collocation.

Theorem 14 (Convergence of sparse collocation)**.** Assume that for $\theta\ge 0$ and $K\ge 1$ there holds

$$\|\Delta_i H_\nu\|_{L^2_\mu}\le(1+K\nu)^\theta,\qquad i\in\mathbb{N}_0. \tag{25}$$

Then, for any function f which satisfies Assumption A3 with $r>2(\theta+1)+\frac{2}{p}$ and Assumption A2, there exists a nested sequence of monotone finite subsets $\Lambda_N\subset\mathcal{F}$ with $|\Lambda_N|=N$ such that for the sparse collocation error holds

$$\|f-U_{\Lambda_N}f\|_{L^2_\mu}\le C(1+N)^{-\left(\frac{1}{p}-\frac{1}{2}\right)}.$$

Proof. We prove the assertion by verifying the assumptions of Theorem 9. Since f satisfies Assumption A3 with $r>2(\theta+1)+\frac{2}{p}$, condition (19) of Theorem 9 holds due to Theorem 12. Moreover, we can apply Lemma 13 to verify the remaining assumption of Theorem 9 about a nonincreasing dominating sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}\in\ell^p(\mathcal{F})$, $p\in(0,2)$: due to Proposition 4) the bound (25) implies

$$c_{\boldsymbol{\nu}}\le\prod_{m\ge1}^{\infty}(1+K\nu_m)^{\theta+1},\qquad\boldsymbol{\nu}\in\mathcal{F},$$

and the sequence $(\tau_m)_{m\in\mathbb{N}}$ appearing in Assumption A3 can w.l.o.g. be assumed to be increasing (otherwise we can permute the dimension accordingly). □

3.3 Convergence of Sparse Collocation Using Gauss-Hermite Nodes

In the following, we will verify the assumption (25) in Theorem 14 for the interpolation operators U_i based on Gauss-Hermite nodes. Moreover, we bound the number of sparse grid points $|\Xi_{\Lambda_N}|$ associated with a multi-index set Λ_N allowing us to relate the convergence rate previously derived to a quantity which reflects the computational effort of the collocation approximation. For nested univariate node sets, i.e., when $\Xi_{i+i}=\Xi_i\cup\{\xi_{i+1}^{(i+1)}\}$, we have $|\Xi_{\Lambda_N}|=|\Lambda_N|$. This simple relation, however, fails to hold for non-nested interpolation sequences such as those based on Gauss-Hermite nodes.

Lemma 15. For U_i being the interpolation operator based on the zeros of the $(i+1)$th Hermite polynomial we have for each $\nu\in\mathbb{N}$ that

$$\|U_iH_\nu\|^2_{L^2_\mu}\le c^2\,\mathrm{e}\,\sqrt{2\nu-1}\qquad\forall i\in\mathbb{N}_0$$

where $c=1.086435$ is the constant appearing in *Cramér's inequality* for Hermite functions. In particular, there holds

$$\|\Delta_iH_\nu\|_{L^2_\mu}\le(1+K\nu)$$

with $K=2c\sqrt{\mathrm{e}}>1$.

Proof. We start by recalling the L^2_μ-orthogonality (μ refers here to the univariate standard Gaussian measure $N(0,1)$) of Lagrange basis polynomials $L_k^{(i)}$ constructed from the zeros $\{\xi_k^{(i)}\}_{k=0}^i$ of the Hermite polynomial of degree $i+1$ ([42, Theorem 14.2.1]). This orthogonality yields

$$\begin{aligned}\|U_i H_\nu\|^2_{L^2_\mu} &= \int_{\mathbb{R}} \left(\sum_{k=0}^{i} H_\nu(\xi_k^{(i)}) L_k^{(i)}(\xi)\right)^2 \mu(\mathrm{d}\xi) = \sum_{k=0}^{i} H_\nu^2(\xi_k^{(i)}) \int_{\mathbb{R}} \left(L_k^{(i)}(\xi)\right)^2 \mu(\mathrm{d}\xi) \\ &= \sum_{k=0}^{i} H_\nu^2(\xi_k^{(i)}) w_k^{(i)}\end{aligned}$$

where $\{w_k^{(i)}\}_{k=0}^i$ denotes the weights of the Gauss quadrature formulae based on the zeros of the $(i+1)$th Hermite polynomial, see also [42, Theorem 14.2.1].

Next, we recall Cramér's inequality for the Hermite polynomials $\tilde{H}_\nu$ taken w.r.t. the weight function $\tilde{\rho}(\xi) = \exp(-\xi^2)$, i.e.,

$$|\tilde{H}_n(\xi)| \le c\pi^{-1/4} \exp(\xi^2/2),$$

see, e.g., [1, Chapter 22, p.787]. Since there holds $\tilde{H}_n(\xi) = \pi^{-1/4} H_n(\xi\sqrt{2})$ [1, Chapter 22, p.778], we get

$$|H_n(\xi)| \le c \exp(\xi^2/4)$$

and, thus,

$$\|U_i H_\nu\|^2_{L^2_\mu} \le c^2 \sum_{k=0}^{i} \exp(\xi_{ki}^2/2) w_{ki},$$

where we switched notation to $\xi_{ki} := \xi_k^{(i)}$ and $w_{ki} := w_k^{(i)}$ for convenience. Furthermore, we use a consequence of [33, Lemma 4]. The latter states for $\tilde{\xi}_{kn}$ being the zeros of $\tilde{H}_n$ and $\tilde{w}_{kn}$ the Christoffel numbers of corresponding Gauss-Hermite quadrature (i.e. Gauss-Hermite weights for $\tilde{\rho}$) that

$$\sum_{k=1}^{n} \tilde{w}_{kn} \exp(\tilde{\xi}_{kn}^2) \le \mathrm{e}\sqrt{\pi(2n+1)}.$$

It can be easily verified that

$$\xi_{kn} = \sqrt{2}\tilde{\xi}_{kn} \quad \text{and} \quad w_{kn} = \pi^{-1/2}\tilde{w}_{kn}.$$

Hence, we get

$$\sum_{k=0}^{i} \exp(\xi_{ki}^2/2) w_{ki} \le \mathrm{e}\sqrt{2(i+1)+1}$$

and by noticing that for $i \ge \nu$ we have $U_i H_\nu = H_\nu$ and, thus, $\|U_i H_\nu\|^2_{L^2_\mu} = 1$, and for $i = \nu - 1$ we get $U_i H_\nu \equiv 0$ the first assertion is shown.

For the second statement we notice

$$\|U_i H_\nu\|^2_{L^2_\mu} \le c^2\, \mathrm{e}\, \nu, \qquad \forall i \in \mathbb{N}_0\, \forall \nu \ge 1$$

since $\nu \ge \sqrt{2\nu - 1}$ for $\nu \ge 1$. And, because of $\Delta_i H_0 \equiv 0$ for $i \ge 1$ and $\Delta_0 H_0 \equiv H_0$, we get

$$\|\Delta_i H_\nu\|_{L^2_\mu} \le 1 + K\nu, \qquad \forall i, \nu \in \mathbb{N}_0.$$

□

Thus, interpolation on Gauss-Hermite points satisfies the assumptions of Theorem 14 with $\theta = 1$ and we obtain

Theorem 16 (Convergence of sparse collocation, Gauss–Hermite nodes)**.** For any function f which satisfies Assumption A3 with $r > 4 + \frac{2}{p}$ and Assumption A2, there exists a nested sequence of mononote finite subsets $\Lambda_N \subset \mathcal{F}$ with $|\Lambda_N| = N$ such that for the error of the sparse collocation operator U_{Λ_N} based on Gauss-Hermite nodes holds

$$\|f - U_{\Lambda_N} f\|_{L^2_\mu} \leq C(1+N)^{-\left(\frac{1}{p}-\frac{1}{2}\right)}.$$

Remark 17. In numerical experiments we actually observed for $\nu = 0, \ldots, 39$ that

$$\|U_i H_\nu\|_{L^2_\mu} \leq 1, \qquad \forall i \in \mathbb{N}_0,$$

see Figure 1. This would imply

$$\|\Delta_i H_\nu\|_{L^2_\mu} \leq \begin{cases} 1 & \text{if } \nu = 0, \\ 2 & \text{otherwise,} \end{cases} \qquad \forall i, \nu \in \mathbb{N}_0.$$

Again, we even observed a smaller bound numerically, see the right plot in Figure 1. However, we have not been able to prove $\|U_i H_\nu\|_{L^2_\mu} \leq 1$ and the improvement in the statement of Theorem 16 would have been minor, i.e., the assertion would also hold with the same rate for functions $f : \Gamma \to \mathcal{H}$ satisfying Assumption A3 with $r > 2 + \frac{2}{p}$. Note that similar numerical evidence was presented in [9] for quadrature operators applied to Hermite polynomials. See also [7] for analogous numerical bounds in the case of Genz-Keister points.

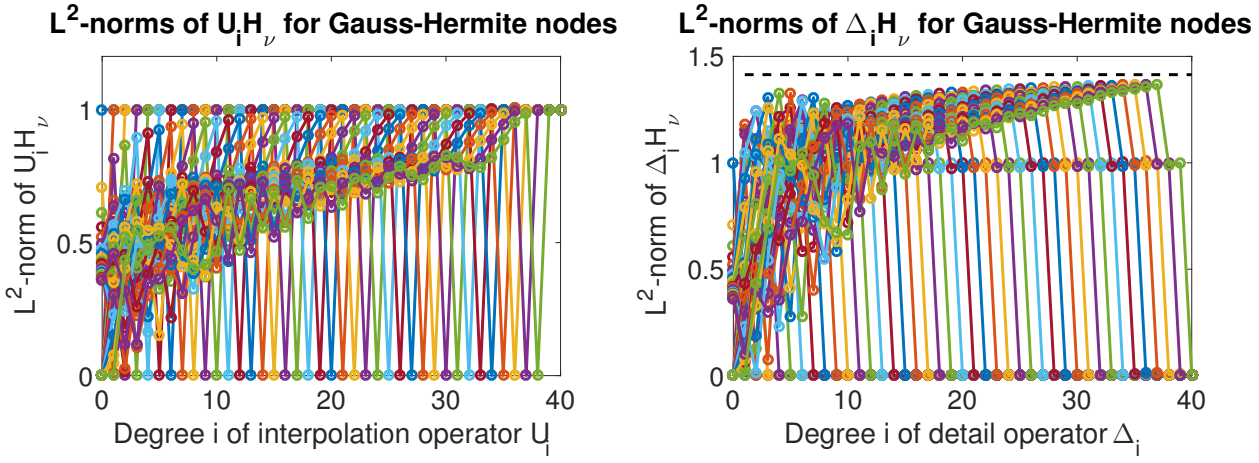

Figure 1: Computed values of $\|U_i H_\nu\|_{L^2_\mu}$ (left) and $\|\Delta_i H_\nu\|_{L^2_\mu}$ (right) for Gauss-Hermite nodes. The dashed, black line in the right plot indicates the value $\sqrt{2}$.

3.4 Convergence Rate With Respect to Number of Collocation Nodes

We now derive bounds for the number of nodes in the sparse grid Ξ_Λ associated with U_Λ. Consider first the following simple monotone index of cardinality N: $\Lambda_N = \{0\boldsymbol{e}_j, \ldots, (N-1)\boldsymbol{e}_j\}$. Then due to $|\Xi_{\{k\boldsymbol{e}_j\}}| = (k+1)$ we get for this Λ_N that

$$|\Xi_{\Lambda_N}| \leq \sum_{k=0}^{N-1} (k+1) = \frac{N(N+1)}{2} \in \mathcal{O}(N^2).$$

The quadratic complexity is indeed sharp, since 0 is the only reappearing Gauss-Hermite node. We will show in the subsequent two propositions that this complexity holds also for arbitrary monotone multi-index sets. We start with rectangular envelopes $\Lambda = \mathcal{R}_{\boldsymbol{\nu}}$ and provide also a rather technical ordering result which we will require later on. Recall that $|\Xi^{(\boldsymbol{i})}| = \prod_{m\geq 1}(1+i_m)$.

Proposition 18. Let $\boldsymbol{\nu} \in \mathcal{F}$. Then there exists an ordering n of $\mathcal{R}_{\boldsymbol{\nu}}$, i.e., a bijective mapping $n : \mathcal{R}_{\boldsymbol{\nu}} \to \{1, \dots, |\mathcal{R}_{\boldsymbol{\nu}}|\}$ such that

$$|\Xi^{(i)}| \leq n(\boldsymbol{i}) \qquad \forall \boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}},$$

which implies, in particular,

$$|\Xi_{\mathcal{R}_{\boldsymbol{\nu}}}| \leq \frac{|\mathcal{R}_{\boldsymbol{\nu}}|\,(|\mathcal{R}_{\boldsymbol{\nu}}|+1)}{2}.$$

Proof. The second assertion follows easily by the first one since

$$|\Xi_{\mathcal{R}_{\boldsymbol{\nu}}}| \leq \sum_{\boldsymbol{i}\leq\boldsymbol{\nu}} |\Xi^{(\boldsymbol{i})}| \leq \sum_{n=1}^{|\mathcal{R}_{\boldsymbol{\nu}}|} n = \frac{|\mathcal{R}_{\boldsymbol{\nu}}|\,(|\mathcal{R}_{\boldsymbol{\nu}}|+1)}{2}.$$

We prove the first assertion by induction. Since $\boldsymbol{\nu} \in \mathcal{F}$, there exist only finitely many $m \in \mathbb{N}$ such that $\nu_m > 0$. Without loss of generality we assume that $\nu_m = 0$ for $m > M$ where $M \in \mathbb{N}$. We now perform an induction over the number M of non-zero entries in ν.

- **base case** $M = 1$: The only possible multi-indices $\boldsymbol{\nu} \in \mathcal{F}$ are $\boldsymbol{\nu} = k\boldsymbol{e}_1$, $k \in \mathbb{N}_0$, and we have $R_{\boldsymbol{\nu}} = \{0\boldsymbol{e}_1, 1\boldsymbol{e}_1, \dots, \nu_1\boldsymbol{e}_1\}$. The ordering is then simply $n(\boldsymbol{i}) = i_1 + 1$. Then
$$|\Xi^{((i_1,0,\dots))}| = 1 + i_1 = n(\boldsymbol{i}).$$
- **Induction step:** the assertion holds for $M \geq 1$. Let $\boldsymbol{\nu} \in \mathcal{F}$ be such that $\nu_m = 0$ for $m \geq M+2$. Moreover, let n_M denote the ordering for $\mathcal{R}_{\boldsymbol{\nu}-\nu_{M+1}\boldsymbol{e}_{M+1}} = \{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}} : i_{M+1} = 0\}$, i.e., it holds
$$|\Xi^{(\boldsymbol{i})}| = \prod_{m=1}^{M}(1+i_m) \leq n_M(\boldsymbol{i}) \qquad \forall \boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}-\nu_{M+1}\boldsymbol{e}_{M+1}}.$$
For notational convenience, we set $\boldsymbol{i}_M := (i_1, \dots, i_M, 0, \dots)$ for each $\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}$ and observe that $\boldsymbol{i}_M \in \mathcal{R}_{\boldsymbol{\nu}-\nu_{M+1}\boldsymbol{e}_{M+1}}$. We define the ordering
$$n(\boldsymbol{i}) := i_{M+1}\left(\prod_{m=1}^{M}(1+\nu_m)\right) + n_M(\boldsymbol{i}_M), \qquad \boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}.$$
It is easy to check that $n : \mathcal{R}_{\boldsymbol{\nu}} \to \{1, \dots, |\mathcal{R}_{\boldsymbol{\nu}}|\}$ is again bijective. Furthermore, we get for each $\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}$
$$\begin{aligned}
|\Xi^{(\boldsymbol{i})}| &= \prod_{m=1}^{M+1}(1+i_m) = (i_{M+1}+1)\prod_{m=1}^{M}(1+i_m) \\
&= i_{M+1}\left(\prod_{m=1}^{M}(1+i_m)\right) + \prod_{m=1}^{M}(1+i_m) \\
&= i_{M+1}\left(\prod_{m=1}^{M}(1+i_m)\right) + |\Xi_{\boldsymbol{i}_m}| \\
&\leq i_{M+1}\left(\prod_{m=1}^{M}(1+\nu_m)\right) + n_M(\boldsymbol{i}_M) = n(\boldsymbol{i})
\end{aligned}$$
where the last line follows by $i_m \leq \nu_m$ for all $m \geq 1$ and the fact that $\boldsymbol{i}_M \in \mathcal{R}_{\boldsymbol{\nu}-\nu_{M+1}\boldsymbol{e}_{M+1}}$ for $\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}}$.

□

We extend the estimate for $\Xi_{\mathcal{R}_{\boldsymbol{\nu}}}$ in the above proposition now to arbitrary finite and monotone index sets Λ:

Proposition 19. Let $\Lambda \subset \mathcal{F}$ be a finite and monotone, then there holds

$$|\Xi_\Lambda| \leq \frac{|\Lambda|\,(|\Lambda|+1)}{2}. \tag{26}$$

Proof. Since Λ is supposed to be monotone, it is a union of rectangular envelopes, see Proposition 2. Thus, there exist n indices $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n \in \mathcal{F}$ such that

$$\Lambda = \bigcup_{k=1}^{n} \mathcal{R}_{\boldsymbol{\nu}_k} \quad \text{and} \quad \Xi_\Lambda = \bigcup_{k=1}^{n} \Xi_{\mathcal{R}_{\boldsymbol{\nu}_k}}.$$

We prove the assertion by induction over n:

- **base case** $n = 1$: The assertion follows by Proposition 18.
- **Induction step:** the assertion holds for $n \geq 1$. With a slight abuse of notation we set $\Lambda_n := \bigcup_{k=1}^{n} \mathcal{R}_{\boldsymbol{\nu}_k}$ and obtain

$$\sum_{\boldsymbol{i} \in \Lambda_n \cup \mathcal{R}_{\boldsymbol{\nu}_{n+1}}} |\Xi^{(\boldsymbol{i})}| = \sum_{\boldsymbol{i} \in \Lambda_n} |\Xi^{(\boldsymbol{i})}| + \sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n} |\Xi^{(\boldsymbol{i})}|.$$

  Let $m := |\mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n|$. The first statement of Proposition 18 now implies

$$\sum_{\boldsymbol{i} \in \mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n} |\Xi^{(\boldsymbol{i})}| \leq \sum_{k=1+|\mathcal{R}_{\boldsymbol{\nu}_{n+1}}|-|\mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n|}^{|\mathcal{R}_{\boldsymbol{\nu}_{n+1}}|} k \leq \sum_{k=1+|\Lambda_n|}^{|\Lambda_n|+|\mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n|} k$$

  where the last inequality is due to $|\mathcal{R}_{\boldsymbol{\nu}_{n+1}}| \leq |\mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n| + |\Lambda_n|$. Thus, we get by the induction hypothesis

$$\sum_{\boldsymbol{i} \in \Lambda_n \cup \mathcal{R}_{\boldsymbol{\nu}_{n+1}}} |\Xi^{(\boldsymbol{i})}| \leq \sum_{k=1}^{|\Lambda_n|} k + \sum_{k=1+|\Lambda_n|}^{|\Lambda_n|+|\mathcal{R}_{\boldsymbol{\nu}_{n+1}} \setminus \Lambda_n|} k = \sum_{k=1}^{|\Lambda_n \cup \mathcal{R}_{\boldsymbol{\nu}_{n+1}}|} k.$$

□

Thus, employing non-nested points such as Gauss-Hermite points, yields at most a quadratic growth of the number of sparse grid points

$$|\Xi_\Lambda| \in \mathcal{O}(|\Lambda|^2)$$

whereas in the nested case one has $|\Xi_\Lambda| = |\Lambda|$.

Remark 20. We provide some numerical validation of the bound (26). More precisely, we consider the following two families of multi-index sets Λ (cf. [6]):

Total Degree (TD): $\Lambda = \Lambda(w, M) = \{\boldsymbol{\nu} \in \mathcal{F} : \sum_{m=1}^{M} \nu_m \leq w,\ \nu_m = 0 \text{ for } m > M\}$

Hyperbolic Cross (HC): $\Lambda = \Lambda(w, M) = \{\boldsymbol{\nu} \in \mathcal{F} : \prod_{m=1}^{M} (\nu_m + 1) \leq w,\ \nu_m = 0 \text{ for } m > M\}$,

In Figure 2 we fix the number of (active) dimensions M and display the cardinality of $\Xi_{\Lambda(w,M)}$ for both choices of $\Lambda(w, M)$ and increasing values of $w \in \mathbb{N}$. The plot shows that estimate (26) is valid but slightly pessimistic for the two specific examples considered here.

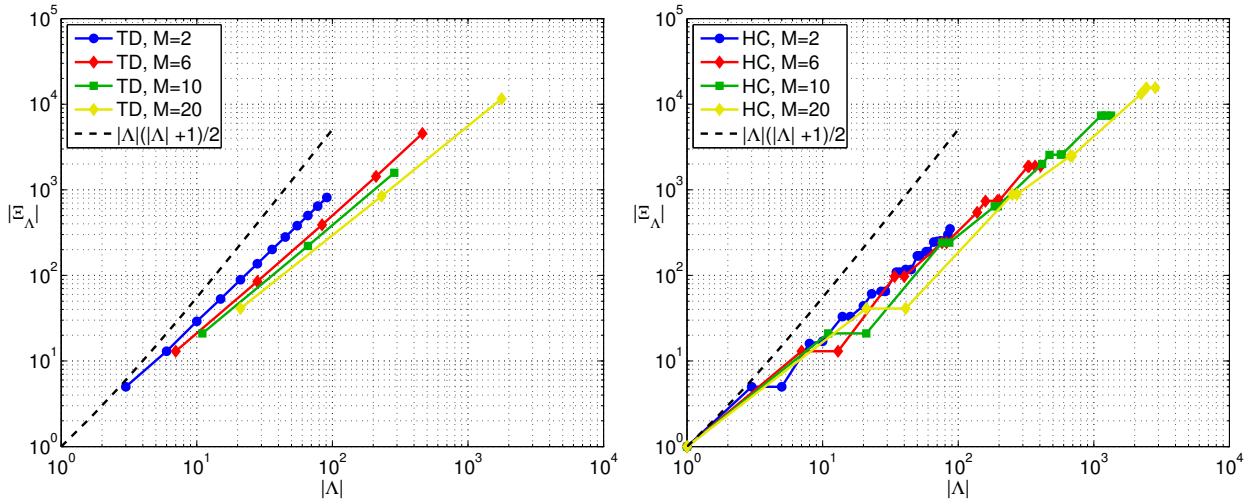

Figure 2: Numerical verification of estimate (26) for "Total Degree" sparse grids (left) and "Hyperbolic Cross" sparse grids (right).

We finally arrive at the resulting error-cost theorem:

Theorem 21 (Convergence rate of Gauss-Hermite sparse grid collocation in terms of nodes)**.** For any function f which satisfies Assumption A3 with $r > 4 + \frac{2}{p}$ and Assumption A2, there exists a nested sequence of mononote finite subsets $\Lambda_N \subset \mathcal{F}$ with $|\Lambda_N| = N$ such that for the error of the sparse collocation operator U_{Λ_N} based on Gauss-Hermite nodes holds

$$\|f - U_{\Lambda_N} f\|_{L^2_\mu} \leq C |\Xi_{\Lambda_N}|^{-\left(\frac{1}{2p} - \frac{1}{4}\right)}$$

where C depends on f.

Hence, assume we require an approximation error $\|f - U_{\Lambda_N} f\|_{L^2_\mu} \leq \varepsilon$, then we can achieve this accuracy with

$$\text{cost}(\varepsilon) \in \mathcal{O}\left(\varepsilon^{\frac{1}{2p} - \frac{1}{4}}\right) \tag{27}$$

number of function evalutions of f. In this cost complexity (27) we neglected of course the computational work which is necessary to find the resulting multi-index sets Λ_N. This is a very important issue. Typically, they are constructed employing adaptive algorithms, see [11, 39, 35] and also Section 5. Our result makes no statement about the actual computational work of those.

Remark 22 (On sparse collocation employing weighted Leja points)**.** As mentioned in the introduction weighted Leja points [32] seem to be a promising node family for interpolation and sparse collocation. So far we are, however, unable to prove bounds like (25) for them. Possibly a more suitable approach for analyzing convergence in case of weighted Leja nodes is to measure the approximation error in the L^∞_μ-norm instead of the L^2_μ-norm and to estimate the corresponding Lebesgue constant. See [27] for first results on the latter – which does not yet imply an analogous estimate to (25) – and [12, 11] for the convergence analysis of sparse collocation using Leja points on $[-1, 1]$ via estimates of the associated Lebesgue constant [10].

4 Application to Elliptic PDEs

We recall our motivation from the introduction: approximating the weak solution u of an elliptic boundary value problem with lognormal diffusion coefficient as in (1) where $f \in L^2(D)$ and $a(\boldsymbol{\xi}) \in L^\infty(D)$ is given

as (2). We will discuss now under which conditions the map $\boldsymbol{\xi} \mapsto u(\boldsymbol{\xi}) \in H_0^1(D)$ satisfies Assumptions A1, A2 and A3 and can therefore be approximated by sparse grid collocation methods based on Gauss-Hermite nodes as outlined in the previous section. We will mainly cite results from [4] but try to emphasize those details which are sometimes omitted in the literature.

Verifying Assumption A1 First of all, we have to investigate the domain Γ of the mapping $\boldsymbol{\xi} \mapsto u(\boldsymbol{\xi})$. There holds $\Gamma \neq \mathbb{R}^{\mathbb{N}}$ since for arbitrary $\boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}}$ the expansion (2) need not converge. A natural domain Γ for the mapping $\boldsymbol{\xi} \mapsto u(\boldsymbol{\xi})$ is

$$\Gamma := \{\boldsymbol{\xi} \in \mathbb{R}^{\mathbb{N}} : \| \sum_{m=1}^{\infty} \phi_m \xi_m \|_{L^\infty(D)} < \infty\}. \tag{28}$$

Further, a natural condition on the decay of the ϕ_m is

$$\sum_{m=1}^{\infty} \|\phi_m\|_{L^\infty(D)}^2 < \infty \tag{29}$$

since (29) implies that the series (2) converges in $L^2_\mu(\mathbb{R}^{\mathbb{N}}; L^\infty(D))$:

$$\begin{aligned}\mathbf{E}\left[\left\|\sum_{m=1}^{\infty} \phi_m\, \xi_m\right\|_{L^\infty(D)}^2\right] &\leq \mathbf{E}\left[\left(\sum_{m=1}^{\infty} \|\phi_m(x)\|_{L^\infty(D)}\, |\xi_m|\right)^2\right] \\ &= \sum_{m=1}^{\infty} \|\phi_m(x)\|_{L^\infty(D)}^2\, \mathbf{E}\left[|\xi_m|^2\right] = \sum_{m=1}^{\infty} \|\phi_m(x)\|_{L^\infty(D)}^2\end{aligned}$$

due to $\mathbf{E}\left[\xi_m \xi_n\right] = \delta_{mn}$. Moreover, by a classical result [28, Lemma 4.16] from probability theory this implies that the series converges also μ-a.e. in $L^\infty(D)$. Thus, if (29) holds, then we get $\mu(\Gamma) = 1$. It remains to state conditions under which we can ensure that $\boldsymbol{\xi} \mapsto u(\boldsymbol{\xi})$ belongs to $L^2_\mu(\Gamma; H_0^1(D))$. Measurability follows from the continuous dependence of the weak solution $u \in H_0^1(D)$ on $\exp(a) \in L^\infty(D)$, see [23]. Moreover, if we can ensure that for $\underline{a}(\boldsymbol{\xi}) := \operatorname{essinf}_{x \in D} \exp(a(x, \boldsymbol{\xi}))$ we have $\underline{a}^{-1} \in L^2_\mu(\Gamma; \mathbb{R})$ (e.g., by Fernique's lemma, as shown in [8]) then the $\boldsymbol{\xi}$-pointwise application of the Lax-Milgram lemma [23] yields for the random solution $u \in L^2_\mu(\Gamma; H_0^1(D))$. The latter can be guaranteed by an even weaker assumption than (29)

Assumption A4 ([4, Assumption A])**.** There exists a strictly positive sequence $(\tau_m)_{m \in \mathbb{N}}$ such that

$$\sup_{x \in D} \sum_{m=1}^{\infty} \tau_m |\phi_m(x)| < \infty, \qquad \sum_{m=1}^{\infty} \exp(-\tau_m^2) < \infty.$$

Under Assumption A4, it is shown in [4, Corollary 2.1] that $u \in L^2_\mu(\Gamma; H_0^1(D))$ with $\mu(\Gamma) = 1$, hence $u : \Gamma \to H_0^1(D)$ satisfies Assumption A1.

Verifying Assumption A2 It is obvious that for Gauss-Hermite nodes there holds $\Xi^{(\boldsymbol{i})} \subset \Gamma$, $\boldsymbol{i} \in \mathcal{F}$, with Γ as in (28), because due to $\boldsymbol{i} \in \mathcal{F}$ there exists an $M \in \mathbb{N}$ such that for $\boldsymbol{\xi} \in \Xi^{(\boldsymbol{i})}$ we have $\xi_m = \xi_0^{(0)}$ for any $m \geq M$ and $\xi_0^{(0)} = 0$. Actually, by Assumption A4 there holds for any $\boldsymbol{\xi} \in \ell^\infty(\mathbb{N})$ that $\boldsymbol{\xi} \in \Gamma$:

$$\left\|\sum_{m=1}^{\infty} \phi_m \xi_m\right\|_{L^\infty(D)} \leq \|\boldsymbol{\xi}\|_{\ell^\infty} \sup_{x \in D} \sum_{m=1}^{\infty} |\phi_m(x)| \leq \frac{\|\boldsymbol{\xi}\|_{\ell^\infty}}{\min_m \tau_m} \sup_{x \in D} \sum_{m=1}^{\infty} \tau_m\, |\phi_m(x)| < \infty,$$

where $\min_m \tau_m > 0$, because Assumption A4 implies $\tau_m \to \infty$ as $m \to \infty$.

Verifying Assumption A3 Again, we refer to results from [4], namely, [4, Theorem 4.2] where the authors show that the (weak) solution u of (1) satisfies Assumption A3 for any $r \in \mathbb{N}_0$ given

Assumption A5. There exists a strictly positive sequence $(\tau_m^{-1})_{m\in\mathbb{N}} \in \ell^p(\mathbb{N})$ such that

$$\sup_{x\in D} \sum_{m=1}^{\infty} \tau_m |\phi_m(x)| < \infty.$$

Please note that Assumption A5 implies Assumption A4, see [4, Remark 2.2]. Hence, we obtain

Theorem 23. Let a be given as in (2) and satisfy Assumption A5. Then there exists a nested sequence of monotone finite subsets $\Lambda_N \subset \mathcal{F}$ with $|\Lambda_N| = N$ such that for the sparse collocation operator U_{Λ_N} based on Gauss-Hermite nodes applied to the solution u of (1) holds

$$\|u - U_{\Lambda_N} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}};H^1_0(D))} \;\le\; C_1 N^{-\left(\frac{1}{p}-\frac{1}{2}\right)} \;\le\; C_2 |\Xi_{\Lambda_N}|^{-\left(\frac{1}{2p}-\frac{1}{4}\right)}.$$

5 Numerical Experiments

We apply the sparse collocation outlined and analyzed in the previous sections to approximate the solution u of a simple boundary value problem taken from [4, Section 7]. In particular, we verify numerically the statement of Theorem 23 and provide some comments on algorithms for constructing sparse grid approximations.

5.1 Problem Setting

We consider the following boundary value problem on the unit interval $D = [0, 1]$:

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x,\boldsymbol{\xi})\,\frac{\mathrm{d}}{\mathrm{d}x}u(x,\boldsymbol{\xi})\right) = f(x), \quad u(0,\boldsymbol{\xi}) = u(1,\boldsymbol{\xi}) = 0, \qquad \mu\text{-a.e.} \tag{30}$$

where we choose $f(x) = 0.03\sin(2\pi x)$ and employ for $\log a$ the following expansion

$$\log a(x,\boldsymbol{\xi}) = 0.1 \sum_{m=1}^{\infty} \frac{\sqrt{2}}{(\pi m)^q} \sin(m\pi x)\xi_m, \qquad \xi_m \sim N(0,1) \text{ i.i.d.}, \; q \ge 1. \tag{31}$$

For $q = 1$ the random field $\log a$ is a Brownian bridge, cf. [4, Section 7], and for $q > 1$ it is a smoother random field. In particular, we get with $\phi_m(x) := \frac{\sqrt{2}}{(\pi m)^q}\sin(m\pi x)$ that for $k = q - 1 - \varepsilon$ with $\varepsilon > 0$

$$\sup_{x\in D} \sum_{m\ge 1} m^k |\phi_m(x)| \le \frac{\sqrt{2}}{\pi^q} \sum_{m\ge 1} m^{-(q-k)} \propto \sum_{m\ge 1} m^{-(1+\varepsilon)} < \infty.$$

Thus, given $q > 1$ the expansion (31) satisfies Assumption A5 for each $p > \frac{1}{q-1}$ and according to Theorem 23 there exists, if $q > 1.5$, a nested sequence of monotone finite subsets $\Lambda_N \subset \mathcal{F}$, $|\Lambda_N| = N$, such that for the sparse collocation operator U_{Λ_N} based on Gauss-Hermite nodes holds

$$\|u - U_{\Lambda_N} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}};H^1_0(D))} \;\le\; C N^{-(q-1.5)} \;\le\; C |\Xi_{\Lambda_N}|^{-\left(\frac{q-1.5}{2}\right)}. \tag{32}$$

In the following we will verify these rates numerically for various values of q.

5.2 Numerical Algorithms

The multi-index sets Λ_N appearing in Theorem 23 and (32) correspond to the largest entries in a p-summable decreasing sequence $(\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ which dominates $\left(c_{\boldsymbol{\nu}}/b_{\boldsymbol{\nu}}^{1/2}\right)_{\boldsymbol{\nu}\in\mathcal{F}}$. Such a dominating sequence was constructed in Lemma 13. However, the resulting multi-index sets Λ_N are in general not available as closed form expressions and need to be constructed by numerical algorithms.

A-priori algorithm The following greedy algorithm is based on [21] and appears in a similar form in the recent work [9]. It successively adds to the set of multi-indices Λ a new multi-index $\boldsymbol{\nu}$ from the set of neighbors $\mathcal{N}(\Lambda)$ which maximizes $|\hat{c}_{\boldsymbol{\nu}}|$. A constraint m_{buffer} restricts the index of dimensions considered for admissible neighbors:

1. Initialize $N = 1$ and $\tilde{\Lambda}_N := \{\mathbf{0}\}$, choose $m_{\text{buffer}} \in \mathbb{N}$ and $N_{\max} \in \mathbb{N}$.
2. For $N = 2, \ldots, N_{\max}$ set

$$\tilde{\Lambda}_N := \tilde{\Lambda}_{N-1} \cup \{\boldsymbol{\nu}_N^*\}, \qquad \boldsymbol{\nu}_N^* := \operatorname*{argmax}_{\boldsymbol{\nu}\in\mathcal{N}(\tilde{\Lambda}_{N-1})} |\hat{c}_{\boldsymbol{\nu}}| \tag{33}$$

where with $\operatorname{supp}(\boldsymbol{\nu}) := \{m \in \mathbb{N} : \nu_m > 0\}$ and $\operatorname{supp}(\Lambda) := \bigcup_{\boldsymbol{\nu}\in\Lambda} \operatorname{supp}(\boldsymbol{\nu})$

$$\mathcal{N}(\Lambda) := \{\boldsymbol{\nu} \in \mathcal{F} \setminus \Lambda : \boldsymbol{\nu} - \boldsymbol{e}_m \in \Lambda \ \forall m \in \operatorname{supp}(\boldsymbol{\nu}) \text{ and } \nu_m = 0 \text{ for } m > \max(\operatorname{supp}(\Lambda)) + m_{\text{buffer}}\}.$$

The set of admissible neighbors $\mathcal{N}(\Lambda)$ of Λ is defined such that adding any $\boldsymbol{\nu} \in \mathcal{N}(\Lambda)$ to Λ maintains monotonicity. The restriction in the definition of $\mathcal{N}(\Lambda)$ above is that we do not allow the *activation* of any dimension $m \in \mathbb{N}$, i.e., including $\boldsymbol{\nu} = \boldsymbol{e}_m$ for arbitrarily (large) $m \in \mathbb{N}$, but restrict the selection to the "next" m_{buffer} higher dimensions. Moreover, for our numerical simulations, we have chosen

$$\hat{c}_{\boldsymbol{\nu}} := \prod_{m\geq 1} (\nu_m)^{2\theta+2-r} \tau_m^{-2(1\wedge\nu_m)}$$

with $\tau_m = m^{q-1}$, $\theta = 1$ and a suitable value[1] for $r > 2(\theta+1) + \frac{2}{p}$, cf. the proof of Lemma 13.

A-posteriori algorithm Beside this *a-priori* construction which is usually cheap to run, we also apply a more costly a-posteriori algorithm for generating monotone multi-index sets Λ_N. Such an algorithm already appeared in [39, 11, 12, 9, 35] and is motivated by using *a-posteriori* heuristics for estimating the improvement of including $\Delta_{\boldsymbol{\nu}} u$ in the sparse collocation approximation. In particular, the a-posteriori algorithm works exactly as the a-priori algorithm except for substituting the choice (33) by

$$\boldsymbol{\nu}_n^* := \operatorname*{argmax}_{\boldsymbol{\nu}\in\mathcal{N}(\tilde{\Lambda}_{n-1})} \frac{\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty_\mu}}{|\Xi^{(\boldsymbol{\nu})}|}. \tag{34}$$

The ratio $\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty_\mu}/|\Xi^{(\boldsymbol{\nu})}|$ represents the *profitability* or *profit* of the multi-index $\boldsymbol{\nu} \in \mathcal{F}$, i.e., the associated gain in approximation $\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty_\mu}$ in relation to the associated computational cost $|\Xi^{(\boldsymbol{\nu})}|$. By choosing the most profitable multi-index in the neighborhood of $\tilde{\Lambda}_{n-1}$ we may obtain a better sparse collocation approximation

[1] We used $r = 2\,(2(\theta+1) + 2/p + 1) = 10 + 4(q-1)$ in the numerical simulations.

Table 1: Statistics for numerical results in Tests - Part I. See equation (32) for the theoretical rates.

q	**var. perc.**	**rate w.r.t.** $\lvert\Lambda_N\rvert$			**rate w.r.t.** $\lvert\Xi_{\Lambda_N}\rvert$		
		theory	a-post.	a-priori	theory	a-post.	a-priori
1	99.91%	N.A.	0.5	0.4	N.A.	0.5	0.5
1.5	99.9999%	0	0.8	0.7	0	0.9	0.8
2	99.9999999%	0.5	1.1	1.0	0.25	1.2	1.1
3	100%	1.5	1.7	1.7	0.75	2	2

than when applying the a-priori construction (33), although the theory developed above does not apply to the multi-indices generated in this way. Here, we estimated $\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty_\mu}$ as in [35] by

$$\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty_\mu} \approx \max_{\boldsymbol{\xi_k} \in \Xi^{(\boldsymbol{\nu})}} \|\rho(\boldsymbol{\xi_k})\, [\Delta_{\boldsymbol{\nu}} u]\,(\boldsymbol{\xi_k})\|_{H^1_0(D)},$$

where $\rho(\boldsymbol{\xi}) = \exp(-\frac{1}{2}\sum_{m\geq 1} \xi_m^2)$ represents the (unnormalized) product density funtion of $\mu = \bigotimes_{m\geq 1} N(0,1)$. Thus, for the a-posteriori algorithm we have to evaluate u on a much larger grid than just $\Xi_{\tilde{\Lambda}_N}$, namely, $\Xi_{\tilde{\Lambda}_N} \cup \Xi_{\mathcal{N}(\tilde{\Lambda}_{N-1})}$, $\Xi_{\mathcal{N}(\tilde{\Lambda}_{N-1})} := \bigcup_{\boldsymbol{\nu}\in\mathcal{N}(\tilde{\Lambda}_{N-1})} \Xi^{(\boldsymbol{\nu})}$. We will refer to $\Xi_{\tilde{\Lambda}_N}$ as the a-posteriori grid (associated with Λ_N) and to $\Xi_{\tilde{\Lambda}_N} \cup \Xi_{\mathcal{N}(\tilde{\Lambda}_{N-1})}$ as the extended grid (associated with Λ_N). The latter represents the "true" computational cost of the sparse collocation approximation generated by the a-posteriori algorithm.

Remark 24. For our numerical simulations we choose a maximal number of parameter dimensions M, which may be arbitrarily large, to construct the refererence solution. Then, for a given $\boldsymbol{\xi} \in \mathbb{R}^M$ we approximate the solution $u(x, \boldsymbol{\xi})$ to (30) by evaluating its exact representation

$$u(x,\boldsymbol{\xi}) = \int_0^x \frac{K(\boldsymbol{\xi}) - F(y)}{a(y,\boldsymbol{\xi})}\,\mathrm{d}y, \qquad F(x) := \int_0^x f(y)\,\mathrm{d}y, \quad K(\boldsymbol{\xi}) := \frac{\int_0^1 \frac{F(y)}{a(y,\boldsymbol{\xi})}\,\mathrm{d}y}{\int_0^1 \frac{1}{a(y,\boldsymbol{\xi})}\,\mathrm{d}y},$$

by numerical quadrature, particularly the trapezoidal rule based on an equidistant spatial grid with spacing $\Delta x = 2^{-10}$.

5.3 Results

We are now ready to discuss the details and the results of the numerical tests we performed. The tests are divided into two parts: in the first set of experiments we aim at validating the sharpness of our analysis, i.e., whether we can actually observe numerically the rate predicted by Theorem 23 for the case of countably many random variables; in the second set of experiments, we will instead gradually increase the number of random variables and see if the observed rate of convergence is actually dimension-independent. Concerning the first set of experiments, we recall that the convergence results in Theorem 23 strictly apply only to the sparse collocation constructed by the a-priori index selection algorithm. However, we will assess whether the set of indices proposed by the a-posteriori construction, i.e., the a-posteriori grid, achieves the same rate and also examine the convergence rate w.r.t. number of points in the extended grid.

Tests - Part I In this section, we will compare the numerical convergence rate of both the a-priori and the a-posteriori versions of the proposed algorithm against the theoretical convergence rate for $q = 1, 1.5, 2, 3$ to verify the sharpness of our theoretical analysis. For each tested value of q, the errors will be computed against a reference solution u_{ref} based on the first 640 random variables which captures more than 99% of the log-diffusion variability for every value of q (see Table 1 for the precise value). The error is computed

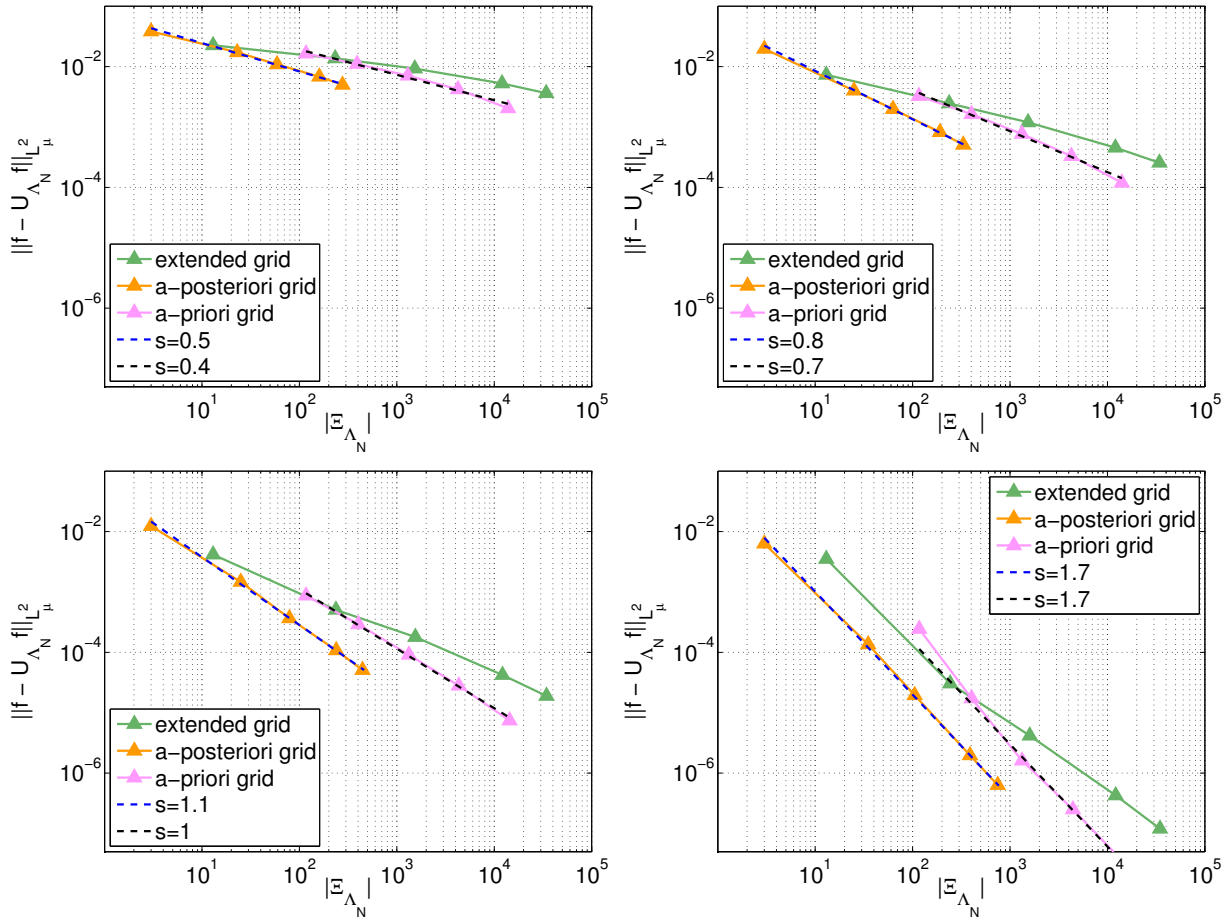

Figure 3: From top-left to bottom-right: convergence with respect to the number of points in the sparse grid for $q = 1, 1.5, 2, 3$.

with a Monte Carlo sampling over $N_{MC} = 1000$ random samples:

$$\|u - U_{\Lambda_N} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}};H^1_0(D))} \approx \|u_{ref} - U_{\Lambda_N} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}};H^1_0(D))} \approx \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \|u_{ref}(\boldsymbol{\xi}_k) - U_{\Lambda_N} u(\boldsymbol{\xi}_k)\|_{H^1_0(D)}, \tag{35}$$

where $\boldsymbol{\xi}_k$ are samples drawn from $\bigotimes_{m=1}^{640} N(0,1)$. We remark that we have verified that N_{MC} is large enough for our purposes.[2]

We begin by reporting in Figure 3 the convergence of the error measure (35) with respect to the number of collocation points needed to construct the sparse grid approximation for each value of q. The convergence plots in Figure 3 show a monotone, well-established decreasing trend for the error for all the variations of the sparse grid considered. As expected, the errors get larger in size and the convergence rate gets worse as q decreases for all the reported sparse grids (a-posteriori grid, extended grid, a-priori). In particular, the convergence rate appears to be similar for the a-priori and the a-posteriori algorithm, with the rate of the latter being actually slightly larger, thus validating the a-posteriori construction. On top of this, the error of the a-posteriori algorithm appears to be smaller in size than the a-priori construction. We also remark that

[2] i.e., repeating the same analysis with $N_{MC} = 5000$ produced identical results.

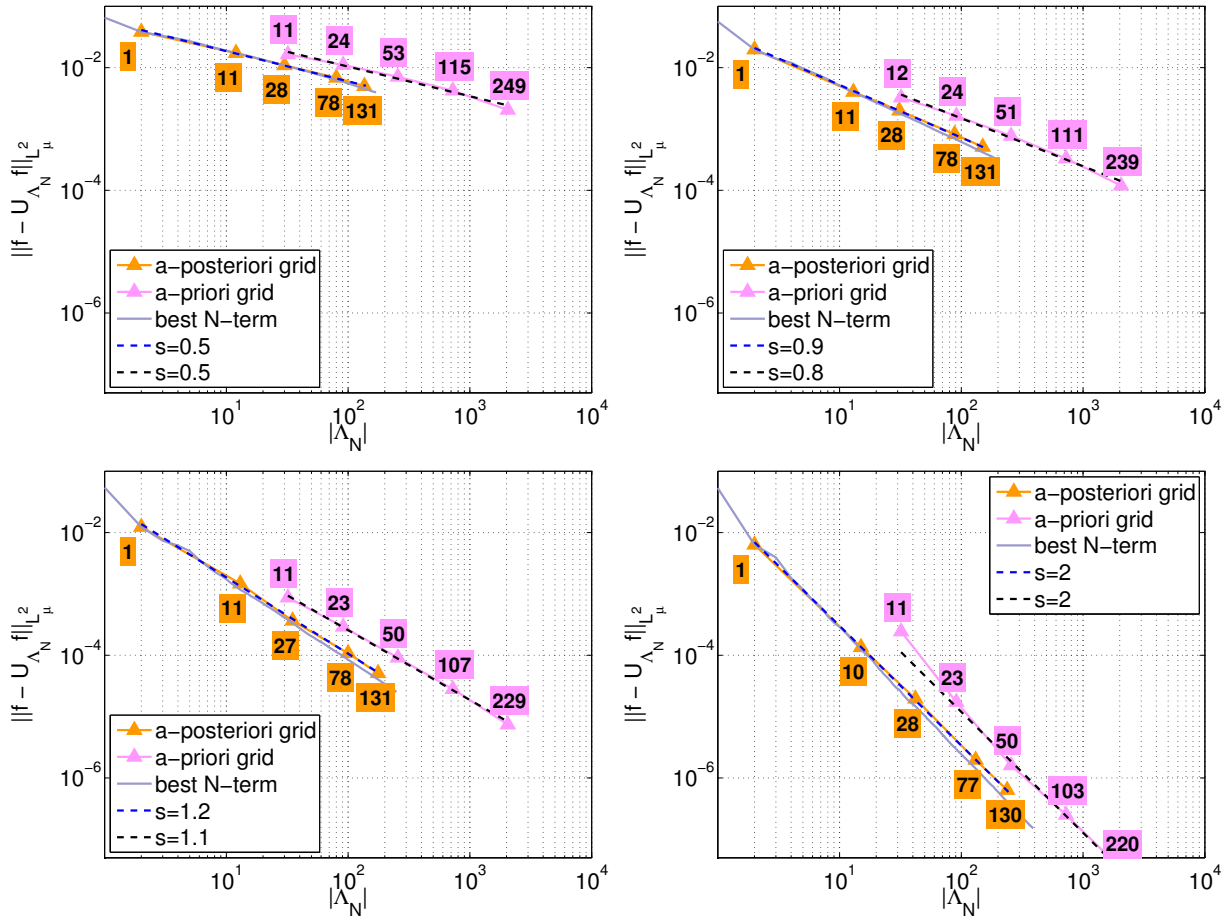

Figure 4: From top-left to bottom-right: convergence with respect to the number of indices in the set Λ_N for $q = 1, 1.5, 2, 3$.

the rate that we measure numerically is better than the one predicted by our theory, cf. Table 1. The quite significant difference between the rate of convergence of the a-posteriori grid and the extended grid is also to be expected. These results are consistent with the ones detailed in [9], although there the a-priori construction is a bit different from the one we propose. At this junction, two factors can explain the suboptimality of our theoretical result: a conservative estimate of the growth of the number of points in the sparse grid with respect to the number of indices in the set Λ_N and a conservative link between the summability of the log-diffusion field representation and the convergence of the sparse grid. As will be clearer later, both issues turn out to be actually affecting our analysis.

The numbers in the plot show the number of activated random variables in the a-posteriori grid and in the a-priori grid, i.e., in how many random variables these grids allocate at least one non-trivial point (observe that by construction the numbers for the extended grid are the ones of a-posteriori grid plus the buffer m_{buffer}). It can be seen that this number steadily increases. The numerical results we show were obtained by $m_{\text{buffer}} = 5$. [3]

We then report in Figure 4 the convergence of the error (35) with respect to the number of indices in the set Λ_N. In this Figure, we show the convergence of both the a-priori and the a-posteriori algorithm, as well as an estimate of the convergence of the best N-term approximation of u (we will detail in a moment how we

[3] We report (not shown) that we have also run the same simulations with a larger buffer $m_{\text{buffer}} = 20$ and the results were identical (i.e., same a-posteriori grid and same number of activated random variables).

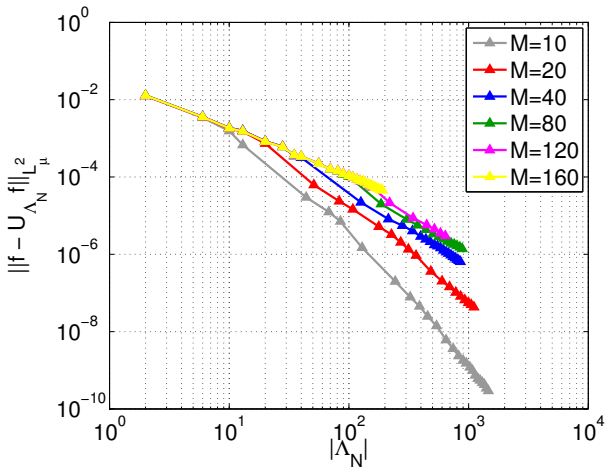

Figure 5: Convergence of the sparse grid approximation with increasingly larger number of dimensions: the asymptotic rate is not constant with respect to M.

computed this approximation). Also in this case, the convergence plots show a monotone, well-established decreasing trend for the error. The results are similar to the previous case: a) the convergence rate of the sparse grid gets worse as q decreases; b) the convergence rate seems to be identical for both the a-priori and the a-posteriori constructions, and again quite larger than the theoretical estimate, cf. Table 1; c) the error of the a-posteriori algorithm is substantially smaller than the one of the a-priori algorithm. It is also relevant to notice that the measured convergence rate here is essentially identical to the one observed with respect to the number of sparse grid points. This is in agreement with the results in [9] and implies that for the sparse grids constructed here the growth of number of points w.r.t. the number of indices is essentially linear, and therefore our Lemma 19 is quite conservative.

We now turn the attention to the best N-term approximation presented in Figure 4. To compute this approximation, we follow [18, 38, 43] and convert the extended grid first into its *combination technique* form, i.e., as a linear combination of Lagrange polynomials, and then we further convert this expression into the equivalent linear combination of Hermite polynomials; see also [16]. By sorting in decreasing order the coefficients of the Hermite expansion thus computed and picking them one at a time, we obtain an approximation of the sequence of best N-term approximations. [4] The comparison of the best N-term and the a-posteriori grid in Figure 4 reveals that the two approximations are actually very close a-posteriori grid for every value of q, which suggests that the a-posteriori algorithm is producing an excellent approximation.

Tests - Part II In this set of experiments, with fix $q = 2, \sigma = 0.1$, and we consider log-diffusion coefficients with $M = 10, 20, 40, 80, 120, 160$ random variables. For each M, the reference solution uses M random variables as well, contrary to the previous experiment, where the reference solution was based on 640 random variables. In this way, we aim at assessing the behavior of the convergence rate as M increases: indeed, the previous experiment was only verifying that we get *a* rate for $M \to \infty$. We report our results in Figure 5, where we show the convergence with respect to the cardinality of the index set Λ_N. It is clearly visible that the convergence curves are all superposed at the beginning of the convergence and then they depart from each other: the point of departure is actually the point where all M variables have been activated. The result seems to suggest that the convergence rate with respect to the cardinality of Λ_N for finite M is actually depending on M, and decreases as M increases, until reaching the asymptotic rate for $M \to \infty$.

[4] Of course, this approximation is as good as the original extended grid; however, we found the results to be stable as the number of points in the extended grid grows, and therefore we deemed this approximation to be sufficient for our purposes.

6 Conclusions

We have presented a general convergence analysis of sparse grid collocation based on Lagrange interpolation for functions of countably many Gaussian variables. In particular, we have stated sufficient conditions on the underlying univariate interpolation nodes such that for functions of a certain smoothness we obtain an algebraic rate of convergence for the sparse collocation approximation with respect to the number of multi-indices. Moreover, we verified these assumptions for the classical Gauss-Hermite nodes and were able to state also a convergence result in terms of the resulting number of collocation points. We finally discussed in detail that these methods can be applied to weak solutions of lognormal diffusion problems and illustrated our theory with numerical tests, which show that the convergence rate achieved by a-priori sparse grid constructions is actually higher than predicted, both with respect to the number of multi-indices and the number of collocation points. The classical adaptive a-posteriori sparse grid construction is also seen to achieve such rates, although not covered by our theory.

Acknowledgments

The authors are grateful to Hans-Jörg Starkloff for pointing out the original reference to Stechkin's lemma.

References

[1] M. ABRAMOWITZ AND I. A. STEGUN, *Handbook of Mathematical Functions*, vol. 55 of Applied Mathematics Series, National Bureau of Standards, Washington, 10th ed., 1972.

[2] R. J. ADLER, *The Geometry of Random Fields*, John Wiley & Sons, New York, 1981.

[3] I. BABUŠKA, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM Review, 52 (2010), pp. 317–355.

[4] M. BACHMAYR, A. COHEN, R. DEVORE, AND G. MIGLIORATI, *Sparse polynomial approximation of parametric elliptic PDEs. part II: lognormal coefficients*, ESAIM Math. Model. Numer. Anal., (2016).

[5] M. BACHMAYR, A. COHEN, AND G. MIGLIORATI, *Sparse polynomial approximation of parametric elliptic PDEs. part I: affine coefficients*, ESAIM Math. Model. Numer. Anal., (2016).

[6] J. BÄCK, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison*, in Spectral and High Order Methods for Partial Differential Equations, vol. 76 of Lecture Notes in Computational Science and Engineering, Springer, 2011, pp. 43–62.

[7] J. BECK, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *A Quasi-optimal Sparse Grids Procedure for Groundwater Flows*, in Spectral and High Order Methods for Partial Differential Equations - ICOSAHOM 2012, vol. 95 of Lecture Notes in Computational Science and Engineering, Springer, 2014, pp. 1–16.

[8] J. CHARRIER, *Strong and weak error estimates for elliptic partial differential equations with random coefficients*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 216–246.

[9] P. CHEN, *Convergence analysis of an adaptive sparse quadrature for high-dimensional integration with Gaussian random variables*. arXiv:1604.08466, 2016.

[10] A. CHKIFA, *On the Lebesgue constant of Leja sequences for the complex unit disk and of their real projection*, Journal of Approximation Theory, 166 (2013), pp. 176 – 200.

[11] A. CHKIFA, A. COHEN, AND C. SCHWAB, *High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs*, Foundations of Computational Mathematics, 14 (2014), pp. 601–633.

[12] ——, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs*, J. Math. Pures Appl., 103 (2015), pp. 400–428.

[13] A. COHEN AND R. DEVORE, *Approximation of high-dimensional parametric PDEs*, Acta Numerica, 24 (2015), pp. 1–159.

[14] A. COHEN, R. DEVORE, AND C. SCHWAB, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Foundations of Computational Mathematics, 10 (2010), pp. 615–646.

[15] A. COHEN, R. DEVORE, AND C. SCHWAB, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs*, Analysis and Applications, 9 (2011), pp. 11–47.

[16] P. CONSTANTINE, M. S. ELDRED, AND E. T. PHIPPS, *Sparse pseudospectral approximation method*, Comput. Methods Appl. Mech. Engrg., 229/232 (2012), pp. 1–12.

[17] S. DE MARCHI, *On Leja sequences: Some results and applications*, Applied Mathematics and Computation, 152 (2004), pp. 621–647.

[18] L. FORMAGGIA, A. GUADAGNINI, I. IMPERIALI, V. LEVER, G. PORTA, M. RIVA, A. SCOTTI, AND L. TAMELLINI, *Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model*, Computational Geosciences, 17(1) (2013), pp. 25–42.

[19] J. GALVIS AND M. SARKIS, *Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity*, SIAM J. Numer. Anal., 47 (2009), pp. 3624–3651.

[20] A. GENZ AND B. D. KEISTER, *Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight*, Journal of Computational and Applied Mathematics, 71 (1996), pp. 299–309.

[21] T. GERSTNER AND M. GRIEBEL, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.

[22] R. GHANEM AND P. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.

[23] D. GILBARG AND N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, Berlin Heidelberg, 2001.

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

[25] A.-L. HAJI-ALI, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *Multi-index Stochastic Collocation convergence rates for random PDEs with parametric regularity*, Foundations of Computational Mathematics, (2016).

[26] 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, 24 (2014), pp. 797–826.

[27] P. JANTSCH, C. G. WEBSTER, AND G. ZHANG, *On the Lebesgue constant of weighted Leja points for Lagrange interpolation on unbounded domains*. arXiv:1606.07093, 2016.

[28] O. KALLENBERG, *Foundations of Modern Probability*, Springer, New York, 2002.

[29] O. P. LE MAITRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, Scientific Computation, Springer, New York, 2010.

[30] F. LEJA, *Une méthode élémentaire de résolution du problème de Dirichlet dans le plan*, Ann. Soc. Math. Polon., 23 (1950), pp. 230–245.

[31] A. MUGLER AND H.-J. STARKLOFF, *On the convergence of the stochastic Galerkin method for random elliptic partial differential equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 47 (2013), pp. 1237–1263.

[32] A. NARAYAN AND J. D. JAKEMAN, *Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation*, SIAM Journal on Scientific Computing, 36 (2014), pp. A2952–A2983.

[33] P. G. NEVAI, *Mean convergence of Lagrange interpolation, II*, Journal of Approximation Theory, 30 (1980), pp. 263–276.

[34] F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs*, Numerische Mathematik, 134 (2016), pp. 343–388.

[35] F. NOBILE, L. TAMELLINI, F. TESEI, AND R. TEMPONE, *An adaptive sparse grid algorithm for elliptic PDEs with lognormal diffusion coefficient*, in Sparse Grids and Applications – Stuttgart 2014, Springer-Verlag, 2016.

[36] F. NOBILE, R. TEMPONE, AND C. WEBSTER, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2411–2442.

[37] ——, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2309–2345.

[38] G. PORTA, L. TAMELLINI, V. LEVER, AND M. RIVA, *Inverse modeling of geochemical and mechanical compaction in sedimentary basins through polynomial chaos expansion*, Water Resources Research, 50 (2014), pp. 9414–9431.

[39] C. SCHILLINGS AND C. SCHWAB, *Sparse, adaptive Smolyak quadratures for Bayesian inverse problems*, Inverse Problems, 29 (2013). doi:10.1088/0266-5611/29/6/065011.

[40] C. SCHWAB AND C. GITTELSON, *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*, Acta Numerica, 20 (2011), pp. 291–467.

[41] S. B. STECHKIN, *On the absolute convergence of orthogonal series*, Doklady Akademii Nauk SSSR, 102 (1955), pp. 37–40.

[42] G. SZEGŐ, *Orthogonal Polynomials*, American Mathematical Society, New York, fourth ed., 1975.

[43] L. TAMELLINI, *Polynomial Approximation of PDEs with Stochastic Coefficients*, PhD thesis, Politecnico di Milano, 2012.

[44] D. XIU AND J. HESTHAVEN, *High-order collocation methods differential equations with random inputs*, SIAM Journal on Scientific Computing, 37 (2005), pp. 1118–1139.