High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs[*]

A. Chkifa, A. Cohen and Ch. Schwab

High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs *

Abdellah Chkifa, Albert Cohen, Christoph Schwab

August 3, 2012

Abstract

We consider the problem of Lagrange polynomial interpolation in high or countably infinite dimension, motivated by the fast computation of solution to parametric/stochastic PDE's. In such applications there is a substantial advantage in considering polynomial spaces that are sparse and anisotropic with respect to the different parametric variables. In an adaptive context, the polynomial space is enriched at different stages of the computation. In this paper, we study an interpolation technique in which the sample set is incremented as the polynomial dimension increases, leading therefore to a minimal amount of PDE solving. This construction is based on standard principle of tensorization of a one dimensional interpolation scheme and sparsification. We derive bounds on the Lebesgue constants for this interpolation process in terms of their univariate counterpart. For a class of model elliptic parametric PDE's, we have shown in [11] that certain polynomial approximations based on Taylor expansions converge in terms the polynomial dimension with an algebraic rate that is robust with respect to the parametric dimension. We show that this rate is preserved when using our interpolation algorithm. We also propose a greedy algorithm for the adaptive selection of the polynomial spaces based on our interpolation scheme, and illustrate its performance both on scalar valued functions and on parametric elliptic PDE's.

1 Introduction

In recent years, various strategies have been proposed for the numerical treatment of parametric partial differential equations [1, 3, 5, 4, 7, 11, 12, 13, 17, 18, 21, 22, 23, 24, 27, 26, 29, 30]. Such equations have the general form

$$\mathcal{D}(u, \mathbf{y}) = 0, \tag{1.1}$$

where $u \mapsto \mathcal{D}(u, \mathbf{y})$ is a partial differential operator that depends on d parameters represented by the vector $\mathbf{y} = (y_1, \ldots, y_d) \in \mathcal{P} \subset \mathbb{R}^d$. Assuming well-posedness of the problem in some Banach space V, the solution map

$$\mathbf{y} \mapsto u(\mathbf{y}), \tag{1.2}$$

is defined from the parameter domain $\mathcal{P}$ to the solution space V

Parametric problems of this type arise in stochastic and deterministic modelling, depending on the nature of the parameters y_j which may either be random or derterministic variables. In both settings, the main computational challenge is to approximate the entire solution map $\mathbf{y} \mapsto u(\mathbf{y})$ up to a prescribed accuracy, with reasonable computational cost. This task is particularly difficult when the number of involved parameters d, is large due to the *curse of dimensionality*. In certain instances, this number may even be countably infinite, meaning that $\mathbf{y} = (y_j)_{j \geq 1}$.

High order polynomial approximation methods such as studied in [5, 3, 11, 12, 13, 18, 22, 23, 26, 29, 30], build approximations of the form

$$u_\Lambda(\mathbf{y}) = \sum_{\nu \in \Lambda} u_\nu \mathbf{y}^\nu, \tag{1.3}$$

where $\Lambda \subset \mathcal{F}$ is a finite set of (multi-)indices $\nu = (\nu_j)_{j \geq 1} \in \mathcal{F}$ and $\mathbf{y}^\nu = \prod_{j \geq 1} y_j^{\nu_j}$. In the finite dimensional setting $d < \infty$, the index set $\mathcal{F}$ coincides with $\mathbb{N}_0^d$ (here and throughout, we denote by $\mathbb{N} = \{1, 2, 3, ...\}$ the set of natural numbers and by $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$). In the infinite dimensional setting, $\mathcal{F}$ denotes the (countable) set of all sequences of nonnegative integers which are *finitely supported* (i.e. those sequence for which only finitely many terms are nonzero). Note that the polynomial coefficients u_ν are functions in V, and therefore the construction of u_Λ requires in principle the computation of $\#(\Lambda)$ such functions. This means that u_Λ is picked in the space

$$V_\Lambda = \text{Span} \left\{ \sum_{\nu \in \Lambda} v_\nu \mathbf{y}^\nu : v_\nu \in V \right\} .$$

We remark that for every finite subset $\Lambda \subset \mathcal{F}$, V_Λ is a closed subspace of the Bochner space $L^2(\mathcal{P}, V)$ and that $V_\Lambda := \mathbb{P}_\Lambda \otimes V$, where

$$\mathbb{P}_\Lambda := \text{Span}\{\mathbf{y}^\nu : \nu \in \Lambda\} \tag{1.4}$$

denotes the polynomial space associated with the index set Λ and with coefficients in $\mathbb{R}$. Here, and throughout, for $1 \leq p \leq \infty$, $L^p(\mathcal{P}, V)$ denotes the Bochner space of strongly measurable mappings from $\mathcal{P}$ to V which are p-integrable (with the usual modification for $p = \infty$) in V-norm with respect to a probability measure on $\mathcal{P}$.

Polynomial approximation is well known to be effective when the solution map has some smoothness. In certain instances, it can even provably break the curse of dimensionality, in the sense that an algebraic convergence rate with respect to $\#(\Lambda)$ can be established even for functions of countably many parameters $d = \infty$. Such results were proved in [12, 13, 11] for the model parametric elliptic equation

$$- \text{div}(a \nabla u) = f \ \text{ in } \ D \subset \mathbb{R}^m, \ \ u = 0 \ \text{ on } \ \partial D, \tag{1.5}$$

where $f \in H^{-1}(D)$ and $a(x, \mathbf{y}) := \bar{a}(x) + \sum_{j \geq 1} y_j \psi_j(x)$, with the functions ψ_j and $\bar{a}$ in $L^\infty(D)$, and for the parameter domain $\mathcal{P} := [-1, 1]^{\mathbb{N}}$. More precisely, it was proved in [12, 13] that if a satisfies the *uniform ellipticity assumption*

$$0 < r \leq a(x, \mathbf{y}) \leq R < \infty, \quad x \in D, y \in \mathcal{P}, \tag{1.6}$$

2

(**UEA**(r, R) for short) and if for some $0 < p < 1$, one has $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$, then there exists a nested sequence

$$\Lambda_1 \subset \Lambda_2 \subset \cdots \subset \Lambda_n \subset \cdots \subset \mathcal{F}, \tag{1.7}$$

with $\#(\Lambda_n) = n$ which satisfies the uniform convergence bound

$$\inf_{v \in V_{\Lambda_n}} \sup_{\mathbf{y} \in \mathcal{P}} \|u(\mathbf{y}) - v(\mathbf{y})\|_V \leq Cn^{-s}, \qquad s := \frac{1}{p} - 1. \tag{1.8}$$

It was also shown in [12, 13] that there exists the sequence of sets Λ_n such that the mean square error convergence bound

$$\inf_{v \in V_{\Lambda_n}} \|u - v\|_{L^2(\mathcal{P}, V)} \leq Cn^{-\bar{s}}, \qquad \bar{s} := \frac{1}{p} - \frac{1}{2}, \tag{1.9}$$

where L^2 is defined with respect to the countable product of uniform measures on P. We point out that the results in this paper apply to a wide class of parametric problems. Problem (1.5) is chosen for the sake of illustration only; regularity estimates for the parametric solution analogous to those used in the present paper have been obtained for nonlinear, elliptic PDE in [20], for nonlinear, initial value ODEs in [21], and for parametric, parabolic evolution problems in [22], and for parametric wave equations in [23]. We also note that regularity for elliptic systems with operators depending on the parameter sequence $\mathbf{y}$ have been considered in [28]. To minimize technicalities in presenting our results, we work throughout with (1.5), but investigate in the numerical experiments in §5 the performance of the adaptive interpolation algorithm also in other settings.

The results in [12, 13] were centered on the existence of such sequences of sets Λ_n rather than on their practical construction. The construction of sequences of (not necessarily optimal) sets Λ_n which achieve the convergence rates (1.8), (1.9), and therefore of the polynomial spaces $\mathbb{P}_{\Lambda_n}$, is critical in the design of algorithms for high-dimensional approximation. Sequences of quasioptimal sets Λ_n which give the rates (1.8), (1.9) with possibly a suboptimal constant $C > 0$ can either be derived from a-priori estimates in [12, 13, 5] or by an adaptive search [18, 11]. The resulting spaces $\mathbb{P}_{\Lambda_n}$ typically differ from the standard multivariate polynomial spaces $\mathbb{P}_k$ of fixed total degree, and also from isotropic or anisotropic sparse grid polynomial spaces studied in [29, 30], although we sometimes refer to them as sparse polynomial spaces.

Given a finite index set Λ, several strategies can be used to compute $u_\Lambda \in V_\Lambda$:

(i) Taylor expansions [11] can be recursively computed in the case of problems with affine parameter dependence such as (1.5). Adaptive methods based on such expansions have been proved to converge uniformly in $u \in \mathcal{P}$ with the same rate as in (1.8).

(ii) Projection methods [1, 4, 5, 12, 18, 22, 23] produce near best approximations in V_Λ for the metric $L^2(\mathcal{P}, \rho)$ where ρ is a chosen measure in the parameter space. In addition, in the Galerkin framework, it is possible to use techniques of a-posteriori analysis in order to adaptively build the sequence of index sets $(\Lambda_n)_{n \geq 1}$. This approach was developed in [18] for the problem (1.5), and proved to converge with the same rate as in (1.9).

3

(iii) Collocation methods [3, 5, 26, 29, 30] produce a polynomial approximation in V_Λ based on the data of particular solution instances $u(\mathbf{y}_i)$ for some chosen values $\mathbf{y}_1, \ldots, \mathbf{y}_k$ of the parameter vector. One significant advantage of this approach is that it is *non intrusive*: the sampling $u(\mathbf{y}_i)$ can be computed by any given numerical solver for the problem (1.1) and the polynomial approximation is built from these solutions by numerical techniques similar to those employed for scalar valued maps such as interpolation or least-square regression. On the other hand, the theoretical analysis of collocation methods is less satisfactory in the sense that convergence rates similar to (1.8) and (1.9) do not seem to have been established for such methods. This is in part due to the difficulty to control the stability of interpolation operators in arbitrary high dimension. In addition, adaptive methods for building the sets Λ_n have not been much developped in the collocation framework.

The objective of this paper is to propose and study a collocation method based on a high dimensional interpolation process that can naturally be coupled with an adaptive selection of the polynomial spaces. We construct an interpolation operator I_Λ that maps real or complex valued functions defined on $\mathcal{P}$ into $\mathbb{P}_\Lambda$. A standard vectorization technique yields a similar operator that maps V-valued functions defined on $\mathcal{P}$ into V_Λ. We assume that the parameter domain has tensorized form

$$\mathcal{P} = P^d, \tag{1.10}$$

(with the understanding that $\mathcal{P}$ denotes the countable cartesian product when $d = \infty$), where P is a closed and bounded coordinate domain (typically, a bounded interval in $\mathbb{R}$ or disk in $\mathbb{C}$). We impose a constraint on the index sets Λ that are considered, described by the following definition.

Definition 1.1 *A set $\Lambda \subset \mathcal{F}$ is called monotone if*

$$\nu \in \Lambda \ \text{ and } \ \mu \leq \nu \Rightarrow \mu \in \Lambda, \tag{1.11}$$

where $\mu \leq \nu$ means that $\mu_j \leq \nu_j$ for all j.

Considering only polynomial spaces $\mathbb{P}_\Lambda$ associated to such sets is very natural. In particular, this allows to replace the monomials $\mathbf{y}^\nu$ in the definition of such spaces by any other tensorized basis $L_\nu(\mathbf{y}) = \prod_{j \geq 1} L_{\nu_j}(y_j)$ where $L_0 \equiv 1$ and $L_k(y_j)$ has degree exactly equal to k with respect to the coordinate y_j (for examples Legendre polynomials). Importantly, it has been shown in [11] that the sets Λ_n achieving the convergence rate (1.8) for the problem (1.5) can be chosen from the restricted class of *monotone subsets of $\mathcal{F}$*. While we develop, as in [11], the algorithms and theory for (1.5), we hasten to add that *all results and algorithms presented in the present paper apply, without any modifications, to the adaptive numerical solution of more general parametric equations*: all that is required is bounded invertibility of the parametric equation for all instances of the parameter sequene and a characterization of the parametric solution families' dependence on the parameters in the sequence. Such characterizations seem to hold for broad classes of parametric problems (we refer to [22, 23, 20, 21] for details).

4

This paper is organized as follows. In §2, we build the operator I_Λ and its associated grid Γ_Λ which is unisolvent for $\mathbb{P}_\Lambda$, for general monotone sets. This construction is based on univariate sequence of points $(z_k)_{k\geq 0}$ and a standard tensorization and sparsification technique, originally due to Smolyak [31]. The main feature of this process is the inherent nested structure the grids, which is is well adapted to an adaptive construction of the index set: the enrichment of Λ by one index is reflected by the enrichment of Γ_Λ by one point. The amount of computation is therefore minimized since all previously computed solution instances are used.

In §3, we study the stability of the interpolation operator I_Λ. In particular we establish bounds on the Lebesgue constant which only depends on the cardinality of the set Λ (not on its shape or on the parametric dimension). These bounds grow algebraically with $\#(\Lambda)$, provided that bounds that grow algebraically with k are available for the Lebesgue constants associated to the sections $\{z_0, \ldots, z_k\}$ of the sequence $(z_k)_{k\geq 0} \subset P$. Such univariate results have recently been obtained in [8, 9, 10] for particular choices of sequences.

Combining the approximation estimate (1.8) together with a bound of the form $\#(\Lambda)^b$ for the Lebesgue constant, one expects that the interpolation $I_{\Lambda_n} u$ converges towards u with a deteriorated rate $n^{-(s-b)}$. In §4, we show by a different error analysis that, under the same assumption for the elliptic equation (1.5), one can construct a sequence $(\Lambda_n)_{n\geq 0}$ such that the interpolation converges towards u with the optimal rate n^{-s}. We present in §5 numerical results that illustrate the performance of our adaptive interpolation scheme.

2 Sparse polynomial interpolation

2.1 Properties of monotone sets

For $i \in \mathbb{N}$, we denote by $e_i = (\delta_{ij})_{j\geq 1}$ the i-th Kronecker sequence. A monotone set Λ can equivalently be defined by the property

$$\nu \in \Lambda \ \text{ and } \ \nu_i \neq 0 \Rightarrow \nu - e_i \in \Lambda. \tag{2.1}$$

Any monotone set contains the null multi-index $(0, 0, \cdots)$, which we will denote by $0_\mathcal{F}$.

In addition to the partial order relation $\mu \leq \nu$ on $\mathcal{F}$, we say that $\mu < \nu$ if and only if $\mu \leq \nu$ but $\mu \neq \nu$ and $\mu \nleq \nu$ if and only if $\mu_j > \nu_j$ for some $j \geq 1$. We say that an index ν is *maximal* in a set $\Lambda \subset \mathcal{F}$ if and only if there is no $\mu \in \Lambda$ satisfying $\nu < \mu$. Any finite set has at least one maximal element. If Λ is monotone and if ν is maximal in Λ, then $\Lambda \setminus \{\nu\}$ is monotone. Conversely, if $\tilde\Lambda \subset \Lambda$ are two monotone sets that differ by one element ν, then ν is maximal in Λ.

It follows that if $(\Lambda_n)_{n\geq 1}$ is a nested sequence of monotone sets with $\#(\Lambda_n) = n$, there exists a unique sequence of indices $(\nu^n)_{n\geq 1} \in \mathcal{F}^\mathbb{N}$ with $\nu^1 = 0_\mathcal{F}$ and such that

$$\Lambda_n = \{\nu^1, \ldots, \nu^n\},$$

for all $n \geq 1$, with $\nu^1 = 0_\mathcal{F}$ and ν^n maximal in Λ_n. Particular examples of monotone sets are the

rectangles $\mathcal{R}_\nu$ defined for any $\nu \in \mathcal{F}$ by

$$\mathcal{R}_\nu := \{\mu \in \mathcal{F} : \mu \leq \nu\}. \tag{2.2}$$

The only maximal element of a rectangle $\mathcal{R}_\nu$ is ν. In general, any finite monotone set $\Lambda \subset \mathcal{F}$ is completely determined by its maximal elements according to

$$\Lambda = \bigcup_{\substack{\nu \in \Lambda \\ \nu \text{ maximal}}} \mathcal{R}_\nu. \tag{2.3}$$

Conversely, given $\mu^1, \ldots, \mu^n$, n multi-indices such that $i \neq j \Rightarrow \mu^i \nleq \mu^j$, then $\bigcup_{j=1}^n \mathcal{R}_{\mu^j}$ is the only monotone set whose maximal elements are exactly the μ^j.

2.2 Univariate interpolation and tensorization

Let $(z_k)_{k\geq 0}$ be a sequence of mutually distinct points in P. We denote by I_k the univariate polynomial interpolation operator associated with the section $\{z_0, \ldots, z_k\}$. This operator acts on a real or complex valued function g defined on P according to

$$I_k g := \sum_{i=0}^k g(z_i) l_i^k, \tag{2.4}$$

where

$$l_j^k(y) := \prod_{\substack{i=0 \\ i \neq j}}^k \frac{y - z_i}{z_j - z_i} \tag{2.5}$$

are the Lagrange polynomials associated with $\{z_0, \ldots, z_k\}$. The possible dependence of the l_j^k on the interpolation points $\{z_0, \ldots, z_k\}$ is expressed by the superscript k in (2.5). For $k \geq 0$, we introduce the difference operator

$$\Delta_k := I_k - I_{k-1}, \tag{2.6}$$

with the convention that I_{-1} is the null operator. Therefore, $\Delta_0 = I_0$ is the operator that maps g to the constant polynomial with value $g(z_0)$. With such notation, we can write for any $n \in \mathbb{N}$

$$I_n = \sum_{k=0}^n \Delta_k. \tag{2.7}$$

Introducing the hierarchical polynomials of degree k associated to the sequence $(z_k)_{k\geq 0}$

$$h_k(y) := \prod_{j=0}^{k-1} \frac{y - z_j}{z_k - z_j} , \quad k > 0 \text{ and } h_0(y) = 1, \tag{2.8}$$

one easily checks that

$$\Delta_k g = \alpha_k(g) h_k, \qquad \alpha_k(g) := g(z_k) - I_{k-1} g(z_k) , \tag{2.9}$$

6

and therefore

$$I_n g = \sum_{k=0}^{n} \alpha_k(g) h_k \,. \tag{2.10}$$

To any multi-index $\nu \in \mathcal{F}$, we associate the multivariate point

$$\mathbf{z}_\nu = (z_{\nu_j})_{j \geq 1} \in \mathcal{P}, \tag{2.11}$$

and the tensorized hierarchical function

$$\mathbf{H}_\nu(\mathbf{y}) = \prod_{j \geq 1} h_{\nu_j}(y_j), \tag{2.12}$$

and the tensorized multivariate operators

$$I_\nu := \bigotimes_{j \geq 1} I_{\nu_j} \quad \text{and} \quad \Delta_\nu := \bigotimes_{j \geq 1} \Delta_{\nu_j}, \tag{2.13}$$

The above tensorization can be defined inductively: for a real or complex valued function g defined on $\mathcal{P}$,

- If $\nu = 0_{\mathcal{F}}$, then $\Delta_\nu g = I_\nu g$ is the constant polynomial with value $g(\mathbf{z}_{0_{\mathcal{F}}})$.

- If $\nu \neq 0_{\mathcal{F}}$, then

$$I_\nu g = I_{\nu_1}(t \mapsto I_{\hat{\nu}} g_t) \quad \text{and} \quad \Delta_\nu g = \Delta_{\nu_1}(t \mapsto \Delta_{\hat{\nu}} g_t), \tag{2.14}$$

where $\hat{\nu} := (\nu_2, \nu_3, ...)$ and for $g \in \Gamma$, g_t is the function defined on $\Gamma^{\mathbb{N}}$ by $g_t(\hat{\mathbf{y}}) = g(\mathbf{y}), \mathbf{y} := (t, \hat{\mathbf{y}})$.

When $d < \infty$, the induction terminates after exactly d steps. When $d = \infty$, the induction terminates after a finite number of steps since for any $\nu \in \mathcal{F}$, applying the operation $\nu \mapsto \hat{\nu}$ sufficiently many times to ν leads to the null multi-index $0_{\mathcal{F}}$. We also observe that

$$I_\nu = \bigotimes_{j \geq 1} \left(\sum_{k=0}^{\nu_j} \Delta_k \right) = \sum_{\mu \leq \nu} \Delta_\nu \,. \tag{2.15}$$

It is easily seen that I_ν is the interpolation operator for the tensor product polynomial space $\mathbb{P}_{\mathcal{R}_\nu} = \bigotimes_{j \geq 1} \mathbb{P}_{\nu_j}$ associated to the grid $\Gamma_{\mathcal{R}_\nu}$.

2.3 The sparse interpolation operator

We are now ready to define the interpolation operator for polynomial spaces associated to monotone spaces and its corresponding unisolvent grid, following the technique of Smolyak [31]: for any monotone set $\Lambda \subset \mathcal{F}$, we define

$$I_\Lambda := \sum_{\nu \in \Lambda} \Delta_\nu, \tag{2.16}$$

7

and

$$\Gamma_\Lambda := \{ \mathbf{z}_\nu \: : \: \nu \in \Lambda \} \: . \tag{2.17}$$

The monotone sets Λ might significantly differ from the sparse grid sets which are usually considered in the literature (see, e.g., [2, 31, 6, 19]). However, the arguments showing that I_Λ is the polynomial interpolation operator on $\mathbb{P}_\Lambda$ associated to the grid Γ_Λ are very similar. For convenience of the reader, we give a precise statement of this result.

Theorem 2.1 *The grid Γ_Λ is unisolvent for $\mathbb{P}_\Lambda$ and for any function g defined on $\mathcal{P}$, the unique element in $\mathbb{P}_\Lambda$ which agrees with g on Γ_Λ is given by $I_\Lambda g$.*

Proof: It is readily seen that $I_\Lambda g$ belongs to $\mathbb{P}_\Lambda$ for any function g defined on $\mathcal{P}$. Since we have

$$\#(\Gamma_\Lambda) = \#(\Lambda) = \dim(\mathbb{P}_\Lambda), \tag{2.18}$$

it suffices to verify that $I_\Lambda g$ agrees with g on Γ_Λ. For this we use (2.15) and the fact that Λ is a monotone set to write for any $\nu \in \Lambda$,

$$I_\Lambda = I_\nu + \sum_{\mu \in \Lambda, \mu \not\leq \nu} \Delta_\mu \: . \tag{2.19}$$

Since I_ν is the interpolation operator for the space $\mathbb{P}_{\mathcal{R}_\nu}$ associated to the grid $\Gamma_{\mathcal{R}_\nu}$, we obviously have

$$\forall \mathbf{z}_\nu \in \Gamma_\Lambda : \quad I_\nu g(\mathbf{z}_\nu) = g(\mathbf{z}_\nu) \: . \tag{2.20}$$

On the other hand if $\mu \not\leq \nu$, there exists a j such that $\nu_j < \mu_j$. Therefore, the univariate operator Δ_{μ_j} returns a polynomial which vanishes at z_{ν_j}, and so $\Delta_\mu g$ vanishes at all points of j coordinate equal to z_{ν_j}. In particular, therefore $\Delta_\mu g(z_\nu) = 0$ and we have thus proved that $I_\Lambda g(\mathbf{z}_\nu) = g(\mathbf{z}_\nu)$. $\square$

Remark 2.2 *The fact that Γ_Λ is unisolvent for the polynomial space $\mathbb{P}_\Lambda$ when Λ is monotone appears to be known from early work on polynomial interpolation: see Chapter IV in the book [25], in which bivariate polynomials associated to monotone sets are referred to as "polynômes pleins". This also appears as a particular case of the theory of the "least polynomial space" for interpolation of functions on general multivariate point sets, see in particular [16]. Here, polynomials associated to monotone sets Λ are referred to as "order closed polynomials" and are proved to be the least polynomial spaces for sets of the form Γ_Λ.*

Remark 2.3 *One can generalize this construction in a straightforward way to tensorized domains of the more general form $\mathcal{P} = \prod_{j \geq 1} P_j$ with different univariate sequences $(z_k^j)_{k \geq 0}$ in each coordinate domain P_j. Another straightforward generalization is when the univariate polynomial spaces $\mathbb{P}_k$ are replaced by more general nested spaces S_k such that $\{z_0, \ldots, z_k\}$ is unisolvent for S_k. Then Γ_Λ is unisolvent for the space*

$$S_\Lambda = \bigoplus_{\nu \in \Lambda} \bigotimes_{j \geq 1} S_{\nu_j},$$

which generalizes $\mathbb{P}_\Lambda$ and the interpolation operator is defined in a similar manner as I_Λ. Sparse grid interpolation based on hierarchical finite element spaces are a particular instance of this generalization.

Note that in our construction, any sequence $(z_k)_{k\geq 0}$ of mutually distinct points can be used. However, the choice of the univariate sequence is critical for the stability of the resulting multivariate interpolation operator I_Λ, expressed by the Lebesgue constant

$$\mathbb{L}_\Lambda := \sup_{g\in B(\mathcal{P})} \frac{\|I_\Lambda g\|_{L^\infty(\mathcal{P})}}{\|g\|_{L^\infty(P)}} , \tag{2.21}$$

where $B(\mathcal{P})$ is the set of bounded functions g on $\mathcal{P}$ which are defined everywhere on $\mathcal{P}$. We are interested in choosing sequences $(z_k)_{k\geq 0}$ such that the Lebesgue constants

$$\lambda_k = \max_{g\in C(\overline{P})} \frac{\|I_k g\|_{L^\infty(P)}}{\|g\|_{L^\infty(P)}}, \tag{2.22}$$

associated with the univariate operators I_k grow moderately with respect to k, since, as shown in §3, this allows to derive estimates on $\mathbb{L}_\Lambda$.

A classical construction of such univariate sequences is by fixing $z_0 \in P$ and defining inductively

$$z_k := \mathrm{Argmax}_{z\in P} \prod_{j=0}^{k-1} |z - z_j| . \tag{2.23}$$

Such $(z_k)_{k\geq 0}$ are called *Leja sequences* on P, and moderate, algebraic growth of the Lebesgue constants λ_k can be established in certain cases as we recall in §3. In addition, the choice of a Leja sequence for $(z_k)_{k\geq 0}$ has an interesting implication on the adaptive choice of the sets Λ_n as we explain in §2.5.

2.4   Hierarchical computation of the interpolation operator

As explained in the introduction, we are interested in performing polynomial interpolation for a nested sequence of sets $(\Lambda_n)_{n\geq 1}$ with $n = \#(\Lambda_n)$. Accordingly the grids $(\Gamma_{\Lambda_n})_{n\geq 1}$ are also nested. The sets Λ_n may either be fixed in advance, or adaptively chosen based on information gained at earlier computational steps.

In this setting, we have observed that each Λ_n can be viewed as the section $\{\nu^1, \dots, \nu^n\}$ of a sequence $(\nu^k)_{k\geq 1} \in \mathcal{F}^{\mathbb{N}}$. This observation leads to an efficient algorithm for the computation of $I_{\Lambda_n} g$ from $I_{\Lambda_{n-1}} g$ and of the value of g at the new point $\mathbf{z}_{\nu^n}$. Indeed, by tensorization we observe that $\Delta_{\nu^n} g$ is a multiple of the tensorized hierarchical function $\mathbf{H}_{\nu^n}$ defined in (2.12). Since $\mathbf{H}_{\nu^n}(\mathbf{z}_{\nu^n}) = 1$, it follows that

$$\Delta_{\nu^n} g = \Delta_{\nu^n} g(\mathbf{z}_{\nu^n})\mathbf{H}_{\nu^n} = (I_{\Lambda_n} g(\mathbf{z}_{\nu^n}) - I_{\Lambda_{n-1}} g(\mathbf{z}_{\nu^n}))\mathbf{H}_{\nu^n} = (g(\mathbf{z}_{\nu^n}) - I_{\Lambda_{n-1}} g(\mathbf{z}_{\nu^n}))\mathbf{H}_{\nu^n}, \tag{2.24}$$

and therefore

$$I_{\Lambda_n} g = I_{\Lambda_{n-1}} g + (g(\mathbf{z}_{\nu^n}) - I_{\Lambda_{n-1}} g(\mathbf{z}_{\nu^n}))\mathbf{H}_{\nu^n} . \tag{2.25}$$

9

Consequently, the polynomials $I_{\Lambda_n}g$ are given by

$$I_{\Lambda_n}g = \sum_{k=0}^{n} g_{\nu^k}\mathbf{H}_{\nu^k} \qquad (2.26)$$

where the coefficients g_{ν^k} are defined recursively by

$$g_{\nu^1} = g(\mathbf{z}_0), \quad g_{\nu^{k+1}} := g(\mathbf{z}_{\nu^{k+1}}) - I_{\Lambda_k}g(\mathbf{z}_{\nu^{k+1}}) = g(\mathbf{z}_{\nu^{k+1}}) - \sum_{i=1}^{k} g_{\nu^i}\mathbf{H}_{\nu^i}(\mathbf{z}_{\nu^{k+1}}) . \qquad (2.27)$$

Remark 2.4 *In the sum that appears on the right side of (2.27), only the terms such that $\nu^i \leq \nu^{k+1}$ are non-zero. When evaluating the computational cost in the above operation, one should make the distinction betwen the cost of the evaluation of $g(\mathbf{z}_{\nu^{k+1}})$ and of computing the linear combination $\sum_{i=1}^{k} g_{\nu^i}\mathbf{H}_{\nu^i}(\mathbf{z}_{\nu^{k+1}})$. In instances where the evaluation of g requires running a heavy numerical code (for example when $g(\mathbf{y})$ is an output of the solution $u(\mathbf{y})$ to a parametric PDE), the first cost dominates the second one. It is also important to notice that only n evaluations of g have been used until the current step n.*

Remark 2.5 *The above algorithm is also efficient to construct the interpolant $I_\Lambda g$ for any given monotone set Λ. Indeed, by iteratively removing maximal elements, we see that any such set can be written as $\Lambda = \Lambda_k$ with $k := \#(\Lambda)$ and $(\Lambda_n)_{n\geq1}$ a sequence of the above type. It is also easily checked that the coefficients g_{ν^k} only depend on g and on the index ν^k and are independent on the index set Λ: these coefficients can be viewed as the unique coordinates of g in the hierarchical basis $(\mathbf{H}_\nu)_{\nu\mathcal{F}}$. One should however be cautious when writing the expansion*

$$g = \sum_{\nu\in\mathcal{F}} g_\nu\mathbf{H}_\nu, \qquad (2.28)$$

since it may fail converge for certain functions g regardless of the ordering of the summation. However, it will be proved to converge for functions that can be approximated sufficiently well by polynomials, based on the stability analysis of the interpolation operator which is the object of the next section.

2.5 Adaptive selection of polynomial spaces

We now discuss the adaptive selection of a nested sequence $(\Lambda_n)_{n\geq1}$. Let us begin with the following analogy: if $(\mathbf{H}_\nu)_{\nu\in\mathcal{F}}$ was an orthonormal basis of $L^2(\mathcal{P})$ then the choice of an index set Λ_n that minimize the L^2 error when truncating the expansion (2.28) would be the indices corresponding to the n largest $|g_\nu|$.

In our current setting however, $(\mathbf{H}_\nu)_{\nu\in\mathcal{F}}$ is not an orthonormal basis and we are rather interested in controlling the error in L^∞. A first greedy strategy is to define Λ_n as the set of indices

corresponding to the n largest contributions of (2.28) measured in the L^∞ metric, i.e. the n largest $a_\nu |g_\nu|$, where

$$a_\nu := \|\mathbf{H}_\nu\|_{L^\infty(\mathcal{P})} := \prod_{j \geq 1} \|h_{\nu_j}\|_{L^\infty(P)}. \tag{2.29}$$

This strategy obviously gives rise to a nested sequence $(\Lambda_n)_{n \geq 1}$, however the sets Λ_n are not ensured to be monotone. In addition, it is not computationally feasible since finding the n largest contributions in (2.28) hints that we should have computed all contributions.

In order to correct these defects, we define for any monotone set Λ a set of neighbours

$$\mathcal{N}(\Lambda) := \{\nu \notin \Lambda \; : \; \mathcal{R}_\nu \in \Lambda \cup \{\nu\}\}, \tag{2.30}$$

or equivalently those $\nu \notin \Lambda$ such that $\nu - e_j \in \Lambda$ for all j such that $\nu_j \neq 0$. Then a natural variant of the first strategy, that leads to a nested sequence of monotone sets, is the following greedy adaptive algorithm.

Adaptive Interpolation (AI) Algorithm:

- Start with $\Lambda_1 := \{0_{\mathcal{F}}\}$.

- Assuming that Λ_{n-1} has been computed, find

$$\nu^n := \operatorname{argmax}\{a_\nu |g_\nu| \; : \; \nu \in \mathcal{N}(\Lambda_{n-1})\}, \tag{2.31}$$

  and define $\Lambda_n = \Lambda_{n-1} \cup \{\nu^n\}$.

Let us observe that since $\mathbf{H}_\nu(\mathbf{z}_\nu) = 1$, we obviously have that $a_\nu \geq 1$. On the other hand, when $(z_k)_{k \geq 0}$ is a Leja sequence built according to (2.23), we obviously have $\max_{z \in P} |h_k(z)| = |h_k(z_k)| = 1$ and therefore

$$a_\nu = \mathbf{H}_\nu(\mathbf{z}_\nu) = 1. \tag{2.32}$$

In such a case, in view of (2.27), the greedy strategy (2.31) amounts in choosing the new index in $\mathcal{N}(\Lambda_{n-1})$ that maximizes the interpolation error at the corresponding new grid point:

$$\nu^n := \operatorname{argmax}\{|g(\mathbf{z}_\nu) - I_{\Lambda_{n-1}} g(\mathbf{z}_\nu)| \; : \; \nu \in \mathcal{N}(\Lambda_{n-1})\}. \tag{2.33}$$

This greedy strategy has several defects. The first one is that it may simply fail to converge, even if there exist sequences $(\Lambda_n)_{n \geq 0}$ such that $I_{\Lambda_n} g$ converges to g at a high rate. This is due to data oscillation that could return an artificially small interpolation error at the new grid point. Consider for example a two dimensional function of the form

$$g(\mathbf{y}) = g(y_1, y_2) = g_1(y_1) g_2(y_2), \tag{2.34}$$

where g_1 and g_2 are non-polynomial smooth functions such that g_2 takes the same values at the points z_0 and z_1. Then, the algorithm will select sets Λ_n that consist of the indices $\nu = (k, 0)$ for

$k = 0, \ldots, n-1$, since the interpolation error at the point $\mathbf{z}_{(k,1)} = (z_k, z_1)$ will vanish. Although this type of situation might be viewed as pathological, it reflects the fact that the algorithm might fail in its first steps to identify the significant variables. One way to avoid this is to impose that when all interpolation errors $|g(\mathbf{z}_\nu) - I_{\Lambda_{n-1}}g(\mathbf{z}_\nu)|$ for $\nu \in \mathcal{N}(\Lambda_n)$ are smaller than some prescribed tolerance $\varepsilon_n > 0$ (that is either fixed or tends to 0 as n grows), then the new index ν^n is chosen arbitrarily from $\mathcal{N}(\Lambda_n)$.

The second defect is that in the infinite dimensional framework $d = \infty$ the set of neighbours $\mathcal{N}(\Lambda)$ has infinite cardinality. One way to treat this defect is to modify its definition by setting

$$\mathcal{N}(\Lambda) := \{\nu \notin \Lambda \; : \; \nu_j = 0 \;\; \text{if} \;\; j > j(\Lambda) + 1 \;\; \text{and} \;\; \mathcal{R}_\nu \in \Lambda \cup \{\nu\}\}, \tag{2.35}$$

where $j(\Lambda) := \max\{j \; : \; \nu_j > 0 \text{ for some } \nu \in \Lambda\}$. This means that we can activate at most one new variable at each iteration step.

Even with such modifications, it is not clear to understand under which additional assumptions on g this adaptive, greedy selection procedure will pick sets $(\Lambda_n)_{n \geq 1}$ such that the interpolation $I_{\Lambda_n}g$ has a guaranteed convergence rate comparable to that of an optimal choice of sets such as, e.g. obtained from best n-term approximation. We give in §5 several numerical examples that illustrate the good practical behaviour of this algorithm.

3   The Lebesgue constant

The accuracy of the interpolation operator I_Λ can be related to the error of best polynomial approximation via the Lebesgue constant $\mathbb{L}_\Lambda$ defined in (2.21), according to the classical inequality

$$\|g - I_\Lambda g\|_{L^\infty(\Gamma^N)} \leq (1 + \mathbb{L}_\Lambda) \inf_{Q \in \mathbb{P}_\Lambda} \|g - Q\|_{L^\infty(\mathcal{P})}. \tag{3.1}$$

A crude, yet useful, way to estimate $\mathbb{L}_\Lambda$ is by using triangle inequality which gives

$$L_\Lambda \leq \sum_{\nu \in \Lambda} \delta_\nu, \tag{3.2}$$

where we define for $\nu \in \mathcal{F}$

$$\delta_\nu := \sup_{g \in B(\mathcal{P}) \setminus \{0\}} \frac{\|\Delta_\nu g\|_{L^\infty(\mathcal{P})}}{\|g\|_{L^\infty(\mathcal{P})}} . \tag{3.3}$$

It is readily seen that

$$\delta_\nu := \prod_{j \geq 1} \delta_{\nu_j} , \tag{3.4}$$

where

$$\delta_k := \sup_{g \in C(P)} \frac{\|\Delta_k g\|_{L^\infty(P)}}{\|g\|_{L^\infty(P)}} \leq \lambda_{k-1} + \lambda_k , \tag{3.5}$$

with λ_k the Lebesgue constant associated with the univariate operators I_k and with the convention that $\lambda_{-1} := 0$. Therefore

$$\mathbb{L}_\Lambda \leq \sum_{\nu \in \Lambda} \prod_{j \geq 1} (\lambda_{\nu_j} + \lambda_{\nu_j - 1}) \tag{3.6}$$

12

Note that $\delta_0 = \lambda_0 = 1$, regardless of the choice of the sequence $(z_k)_{k \geq 0}$.

The bound (3.6) is of course crude, since we did not take advantage of the telescoping nature in the summation of the Δ_ν. For instance, when $\Lambda = \mathcal{R}_\nu$, we have seen that $I_\Lambda = I_\nu = \bigotimes_{j \geq 1} I_{\nu_j}$, and in that case the exact value of the Lebesgue constant is given by the smaller value

$$\mathbb{L}_{\mathcal{R}_\nu} = \prod_{j \geq 1} \lambda_{\nu_j} \, . \tag{3.7}$$

Nevertheless, for general monotone sets Λ, we can use the bound (3.6) to study the behaviour of the Lebesgue constant $\mathbb{L}_\Lambda$ as the number of interpolation points $\#(\Gamma_\Lambda) = \#(\Lambda)$ grows. The following result shows that when certain algebraic bounds are available for the λ_k, then similar algebraic bounds can be derived for $\mathbb{L}_\Lambda$ in terms of $\#(\Lambda)$ regardless of the dimension d and of the shape of Λ.

Lemma 3.1 *If the Lebesgue constants λ_k satisfy*

$$\lambda_k \leq (k+1)^\theta, \quad k \geq 0 \tag{3.8}$$

for some $\theta \geq 1$, then the Lebesgue constants $\mathbb{L}_\Lambda$ satisfy

$$\mathbb{L}_\Lambda \leq (\#\Lambda)^{\theta+1} \tag{3.9}$$

for any monotone set Λ.

Proof: For any $k \geq 0$, one has $\lambda_k + \lambda_{k-1} \leq (k+1)^\theta + k^\theta \leq (2k+1)(k+1)^{\theta-1}$, therefore, for $\nu \in \Lambda$

$$
\begin{aligned}
\prod_{j \geq 1} (\lambda_{\nu_j} + \lambda_{\nu_j - 1}) & \leq \left(\prod_{j \geq 1} (\nu_j + 1) \right)^{\theta-1} \prod_{j \geq 1} (2\nu_j + 1) \\
& = (\#(\mathcal{R}_\nu))^{\theta-1} \prod_{j \geq 1} (2\nu_j + 1) \\
& \leq (\#(\Lambda))^{\theta-1} \prod_{j \geq 1} (2\nu_j + 1),
\end{aligned}
$$

where we have used $\mathcal{R}_\nu \subset \Lambda$ since Λ is monotone. To complete the proof, it remains to show that $\sigma(\Lambda) \leq (\#\Lambda)^2$, where

$$\sigma(\Lambda) := \sum_{\nu \in \Lambda} \prod_{j \geq 1} (2\nu_j + 1) \, . \tag{3.10}$$

This is done using induction on $n_\Lambda := \#(\Lambda)$. When $n_\Lambda = 1$, then $\Lambda = \{0_\mathcal{F}\}$ and it is obviously true. Let $n \geq 1$ and Λ denote a monotone set with $n_\Lambda = n + 1$. Without loss of generality, we suppose that $\mu_1 \neq 0$ for some $\mu \in \Lambda$, and denote by $K \geq 1$ the maximal value attained by the coordinate ν_1 when $\nu \in \Lambda$. For $0 \leq k \leq K$, we introduce

$$\Lambda_k := \{\hat{\nu} = (\nu_2, \nu_3, \dots) : (k, \hat{\nu}) \in \Lambda\} \tag{3.11}$$

By monotonicity of Λ, every Λ_k is monotone. Also, since $K \geq 1$ then $\#(\Lambda_k) < \#(\Lambda)$ for any k, so that the induction hypothesis implies

$$\sigma(\Lambda) = \sum_{k=0}^{K}(2k+1)\sigma(\Lambda_k) \leq \sum_{k=0}^{K}(2k+1)(\#(\Lambda_k))^2 . \tag{3.12}$$

Also, we have

$$\Lambda_K \subset \cdots \subset \Lambda_1 \subset \Lambda_0, \tag{3.13}$$

since for $k \geq 1$, $\mu \in \Lambda_k \Rightarrow (k, \mu) \in \Lambda \Rightarrow (k-1, \mu) \in \Lambda \Rightarrow \mu \in \Lambda_{k-1}$. We deduce

$$k(\#(\Lambda_k))^2 \leq \#(\Lambda_k)\#(\Lambda_0) + ... + \#(\Lambda_k)\#(\Lambda_{k-1}), \tag{3.14}$$

and consequently

$$\sigma(\Lambda) \leq \sum_{k=0}^{K}(\#(\Lambda_k))^2 + 2\sum_{k=0}^{K}(\#(\Lambda_k)\#(\Lambda_0) + ... + \#(\Lambda_k)\#(\Lambda_{k-1})) \leq \left(\sum_{k=0}^{K}\#(\Lambda_k)\right)^2 = (\#(\Lambda))^2, \tag{3.15}$$

which concludes the proof. $\qquad\square$

Remark 3.2 *In the case where $(z_k)_{k\geq 0}$ is a Leja sequence defined by (2.23) for some initial point $z_0 \in P$, the hierarchical polynomials h_k defined by (2.8) satisfy*

$$|h_k(z)| \leq |h_k(z_k)| = 1, \quad z \in P. \tag{3.16}$$

Since, according to (2.9), we have

$$\Delta_k g = (g(z_k) - I_{k-1}g(z_k))h_k, \tag{3.17}$$

it follows that

$$\delta_k \leq 1 + \lambda_{k-1}, \tag{3.18}$$

and

$$\mathbb{L}_\Lambda \leq \sum_{\nu \in \Lambda}\prod_{j\geq 1}(1 + \lambda_{\nu_j-1}) \tag{3.19}$$

which is are improvements over (3.5) and (3.19).

Let us observe that since $\lambda_{-1} = 0$ and $\lambda_0 = 1$, bounds of the form $\lambda_k \leq (k+1)^\theta$ can be established for some $\theta > 0$ provided that λ_k are bounded as $\mathcal{O}(k^b)$ for some $b > 0$. Such bounds have recently been obtained in [8, 9, 10], in the cases where P is either the complex unit disk $\{|z| \leq 1\}$ or the unit interval $[-1, 1]$.

For the complex unit disk, if $(z_k)_{k\geq 0}$ is the Leja sequence defined by (2.23) with $z_0 = 1$, it is proved in [8] that $\lambda_k = \mathcal{O}(k \log k)$ and conjectured that

$$\lambda_k \leq k + 1, \quad k \geq 0 . \tag{3.20}$$

The more precise bound

$$\lambda_k \leq 2(k+1), \quad k \geq 1, \tag{3.21}$$

is established in [10]. Using the improvement (3.19), and using the same computation as in the proof of the above Lemma, we obtain in that case that $\mathbb{L}_\Lambda \leq \#(\Lambda)^2$.

For the unit interval $[-1, 1]$, and the sequence given by the projection of the above complex Leja sequence, an asymptotic bound in $\mathcal{O}(k^3 \log k)$ has been obtained in [9], and it has been improved into $3(k+1)^2 \log(k+1)$ in [10]. Note that projections of Leja sequences on the unit disk to $[-1, 1]$ are not Leja sequences on $[-1, 1]$.

4 Application to parametric PDEs

4.1 Interpolation of Banach valued functions

We are interested in applying our interpolation process to the map $\mathbf{y} \mapsto u(\mathbf{y})$ defined by exact or approximate solving of the parametric PDE (1.1) for the given parameter $\mathbf{y}$. Therefore, we want to interpolate a function which is not real or complex valued, but instead takes its value in the solution space V. The generalization of the interpolation operator I_Λ to this setting is straighforward: $I_\Lambda u$ is the unique function in V_Λ that coincides with u at the points $\mathbf{z}_\nu$ for $\nu \in \Lambda$. As in the scalar case, it can be expanded according to

$$I_\Lambda u = \sum_{\nu \in \Lambda} u_\nu \mathbf{H}_\nu , \tag{4.1}$$

where the coefficients $u_\nu \in V$ can be computed in a recursive way similar to (2.27):

$$u_{\nu^1} = u(\mathbf{z}_0), \quad u_{\nu^{k+1}} = u(\mathbf{z}_{\nu^{k+1}}) - \sum_{i=1}^{k} u_{\nu^i} \mathbf{H}_{\nu^i}(\mathbf{z}_{\nu^{k+1}}), \tag{4.2}$$

where $\Lambda_n = \{\nu^1, \dots, \nu^n\}$, $n = 1, 2, \dots$, is a nested sequence of monotone sets. We are interested in the accuracy of the interpolant in the sense of the maximum error

$$\|u - I_\Lambda u\|_{L^\infty(\mathcal{P}, V)} := \sup_{\mathbf{y} \in \mathcal{P}} \|u(\mathbf{y}) - I_\Lambda u(\mathbf{y})\|_V. \tag{4.3}$$

The same reasoning as for interpolation of scalar valued functions shows that

$$\|u - I_\Lambda u\|_{L^\infty(\mathcal{P}, \Lambda)} \leq (1 + \mathbb{L}_\Lambda) \inf_{v \in V_\Lambda} \|u - v\|_{L^\infty(\mathcal{P}, \Lambda)}, \tag{4.4}$$

where $\mathbb{L}_\Lambda$ is the Lebesgue constant associated to the interpolation operator I_Λ which was defined and studied in the previous section.

4.2 Convergence rates for a parametric, elliptic model problem

As already explained in the introduction, for the model elliptic problem (1.5), one can establish convergence rates in $L^\infty(\mathcal{P}, V)$ and $L^2(\mathcal{P}, V)$, where $V = H_0^1(D)$, for polynomial approximation

15

that are robust with respect to the parametric dimension. Here we focus on $L^\infty(\mathcal{P}, V)$. The results in [12, 13, 11] show that if the diffusion satisfies the uniform ellipticity assumption (1.6) and if for some $0 < p < 1$, one has $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$, there exists a nested sequence $(\Lambda_n)_{n\geq 1}$ of monotone sets such that $\#(\Lambda_n) = n$ and

$$\inf_{v \in V_{\Lambda_n}} \|u - v\|_{L^\infty(\mathcal{P}, \Lambda)} \leq Cn^{-s}, \quad s := \frac{1}{p} - 1. \tag{4.5}$$

where $C > 0$ depends on $s := \frac{1}{p} - 1$ but is independent of n. In fact this rate of convergence was proved for specific approximations defined either by Taylor expansion

$$T_{\Lambda_n} u(\mathbf{y}) := \sum_{\nu \in \Lambda_n} t_\nu \mathbf{y}^\nu, \tag{4.6}$$

where $t_\nu := \frac{1}{\nu!} \partial_y^\nu u(y)|_{y=0}$ with $\nu! := \prod_{j\geq 1} \nu_j!$ or by orthogonal projections

$$P_{\Lambda_n} u(\mathbf{y}) := \sum_{\nu \in \Lambda_n} c_\nu L_\nu(\mathbf{y}), \tag{4.7}$$

where the L_ν are the tensorized Legendre polynomials in $[-1, 1]$, which are normalized with respect to the probability measure $dt/2$ and where c_ν are the corresponding Legendre coefficients. The fact that we work in the infinite dimensional case $d = \infty$ in (4.5) reveals that the convergence rate n^{-s} in (4.5) is robust with respect to the number of active variables. In this section, we work under the same assumptions on the diffusion coefficient and its expansion.

We now study the rate of convergence of $I_{\Lambda_n} u$ towards u. Combining (4.5) and (4.4), we obtain

$$\|u - I_{\Lambda_n} u\|_{L^\infty(\mathcal{P}, V)} \leq C(1 + \mathbb{L}_{\Lambda_n})n^{-s}. \tag{4.8}$$

We have seen in §4 that the Lebesgue constant can be controlled by a bound of the form

$$\mathbb{L}_{\Lambda_n} \leq n^b, \quad \text{where} \quad b = \theta + 1, \tag{4.9}$$

when the univariate sequence $(z_k)_{k\geq 0}$ is chosen so that $\lambda_k \leq (k+1)^\theta$ for some $\theta > 0$ and we gave examples of such sequences. We thus obtain a convergence estimate of the form

$$\|u - I_{\Lambda_n} u\|_{L^\infty(\mathcal{P}, V)} \leq Cn^{-(s-b)}. \tag{4.10}$$

With this simple stability (via the bound for the Lebesgue constant) plus consistency (via the n-term approximation result) analysis, the convergence rate obtained in (4.10) is reduced by $b = \theta + 1$ compared to the (benchmark) n-term approximation rate s in (4.5).

We now present a more refined argument that shows that there exists a sequence $(\Lambda_n)_{n\geq 1} \subset \mathcal{F}$ of monotone sets such that $\#(\Lambda_n) = n$ and a constant $C > 0$ such that for all $n \in \mathbb{N}$ holds

$$\|u - I_{\Lambda_n} u\|_{L^\infty(\mathcal{P}, V)} \leq Cn^{-s}. \tag{4.11}$$

This analysis relates more directly the interpolation error with the Taylor coefficients of u according to the following result.

Lemma 4.1 *Assume that $u(\mathbf{y}) = \sum_{\nu \in \mathcal{F}} t_\nu \mathbf{y}^\nu$ in the sense of unconditional convergence in $L^\infty(\mathcal{P}, V)$. If the univariate sequence $(z_k)_{k \geq 0}$ is chosen so that $\lambda_k \leq (k+1)^\theta$, then*

$$\|(I - I_\Lambda)u\|_{L^\infty(\mathcal{P},V)} \leq 2 \sum_{\nu \notin \Lambda} p_\nu(b) \|t_\nu\|_V \ , \tag{4.12}$$

for any finite monotone set Λ, where $b := \theta + 1$ and

$$p_\nu(b) := \prod_{j \geq 1} (1 + \nu_j)^b \ . \tag{4.13}$$

Proof: The unconditional convergence of the Taylor series allows us to write

$$I_\Lambda u = I_\Lambda \left(\sum_{\nu \in \mathcal{F}} t_\nu \mathbf{y}^\nu \right) = \sum_{\nu \in \mathcal{F}} t_\nu I_\Lambda \mathbf{y}^\nu = \sum_{\nu \in \Lambda} t_\nu \mathbf{y}^\nu + \sum_{\nu \notin \Lambda} t_\nu I_{\Lambda \cap \mathcal{R}_\nu} \mathbf{y}^\nu. \tag{4.14}$$

Here, we used that for monotone Λ it holds that $I_\Lambda \mathbf{y}^\nu = \mathbf{y}^\nu$ for every $\nu \in \Lambda$ and that $\Delta_\mu \mathbf{y}^\nu = 0$ for every $\mu \leq \nu$. Therefore

$$(I - I_\Lambda)u = \sum_{\nu \notin \Lambda} t_\nu (I - I_{\Lambda \cap \mathcal{R}_\nu}) \mathbf{y}^\nu, \tag{4.15}$$

where I stands for the identity operator. This implies

$$\|(I - I_\Lambda)u\|_{L^\infty(\mathcal{P},V)} \leq \sum_{\nu \notin \Lambda} \|t_\nu\|_V (1 + \mathbb{L}_{\Lambda \cap \mathcal{R}_\nu}) \leq 2 \sum_{\nu \notin \Lambda} \|t_\nu\|_V \mathbb{L}_{\Lambda \cap \mathcal{R}_\nu} \ . \tag{4.16}$$

Since $\lambda_k \leq (k+1)^\theta$, we obtain from Lemma 3.1 that

$$\mathbb{L}_{\Lambda \cap \mathcal{R}_\nu} \leq \#(\Lambda \cap \mathcal{R}_\nu)^{\theta+1} \leq \#(\mathcal{R}_\nu)^{\theta+1} = p_\nu(b), \tag{4.17}$$

which completes the proof. $\qquad\qquad\square$

Let us observe that the above lemma can be generalized to polynomial expansions other than Taylor series, for example expansions into Legendre polynomials, provided that unconditional convergence holds. We focus here on the Taylor series, which allows us to use the results in [13] that prove unconditional convergence under our assumptions on the diffusion coefficients, based on explicit bounds for the V-norms of Taylor coefficients. These bounds are obtained by application of the Cauchy formula, on the holomorphic extension of the map $\mathbf{y} \mapsto u(\mathbf{y})$, and are of the form

$$\|t_\nu\|_V \leq C_\delta \prod_{j \geq 1} \rho_j^{-\nu_j} = C_\delta \rho^{-\nu}, \tag{4.18}$$

where $C_\delta := \frac{\|f\|_{V^*}}{\delta}$ for arbitrary $\delta > 0$ and $\rho := (\rho_j)_{j \geq 1}$ is any sequence of strictly positive numbers that satisfies

$$\sum_{j \geq 1} \rho_j |\psi_j(x)| \leq \bar{a}(x) - \delta, \quad x \in D \ . \tag{4.19}$$

We call sequences ρ which satisfy (4.19) δ-admissible. We observe that δ has to be chosen smaller than the ellipticity constant r in order to ensure the existence of such sequences. In [13], (4.18) is used to prove the unconditional convergence of the Taylor series of u in $L^\infty(\mathcal{P}, V)$ and the convergence rate (4.5). In the next result, we use the same bounds to prove a summability result on the sequence $(p_\nu(b)\|t_\nu\|_V)_{\nu \in \mathcal{F}}$.

Theorem 4.2 *Assume that if the diffusion coefficient satisfies the uniform ellipticity assumption (1.6) and that for some $0 < p < 1$, one has $(\|\psi_j\|_{L^\infty(D)})_{j \geq 1} \in \ell^p(\mathbb{N})$. Then for any $b \geq 1$ and for $p_\nu(b)$ as in (4.13), the sequence $(p_\nu(b)\|t_\nu\|_V)_{\nu \in \mathcal{F}}$ belongs to $\ell^p(\mathcal{F})$.*

Proof: We fix $B = e^b$. Since $p < 1$, $(\|\psi_j\|_{L^\infty(D)})_{j \geq 1} \in \ell^1(\mathbb{N})$ and we may choose $\kappa > 1$ and an integer $J_0 \geq 1$ such that

$$(\kappa - 1)\sum_{j \geq 1} \|\psi_j\|_{L^\infty(D)} \leq \frac{r}{6} \quad \text{and} \quad \sum_{j > J_0} \|\psi_j\|_{L^\infty(D)} \leq \frac{1}{B}\frac{r}{12}, \tag{4.20}$$

and we set $E := \{j \ : \ 1 \leq j \leq J_0\}$ and $F := \{j \ : \ j > J_0\}$. For each $\nu \in \mathcal{F}$, we denote by ν_E and ν_F the restrictions of ν on E and F, and we define the sequence $\rho = \rho(\nu)$ by $\rho_j = \kappa$ if $j \leq J_0$, $\rho_j = 1$ if $j > J_0$ with $\nu_j = 0$, and

$$\rho_j = B + \frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}, \quad j > J_0, \ \nu_j \neq 0 \tag{4.21}$$

where $|\nu_F| = \sum_{j > J_0} |\nu_j|$. We claim that ρ is $\frac{r}{2}$-admissible. Indeed, for almost every $x \in D$

$$
\begin{aligned}
\sum_{j \geq 1} \rho_j |\psi_j(x)| \ &\leq \kappa \sum_{j \leq J_0} |\psi_j(x)| + \frac{r}{4} \sum_{j > J_0} \frac{\nu_j}{|\nu_F|\|\psi_j\|_{L^\infty(D)}} |\psi_j(x)| + B \sum_{j \geq J_0} |\psi_j(x)| \\
&\leq \sum_{j < J_0} |\psi_j(x)| + \frac{r}{6} + \frac{r}{4} + \frac{r}{12} \\
&\leq \bar{a}(x) - r + \frac{r}{2} \\
&= \bar{a}(x) - \frac{r}{2} \ .
\end{aligned}
$$

Therefore, by (4.18)

$$p_\nu(b)\|t_\nu\|_V \leq C_{\frac{r}{2}} \prod_{j \leq J_0} \frac{(1 + \nu_j)^b}{\kappa^{\nu_j}} \prod_{j > J_0} \frac{(1 + \nu_j)^b}{\rho_j^{\nu_j}} \ . \tag{4.22}$$

Since $\kappa > 1$, we have $(1 + n)^b \leq C_0(\frac{1+\kappa}{2})^n$ for any $n \geq 1$ provided that $C_0 = \exp\left(b / \log((1 + \kappa)/2)\right)$. Combining this with (4.22) implies

$$p_\nu(b)\|t_\nu\|_V \leq C q_\nu \quad \text{where} \quad q_\nu := \prod_{j \leq J_0} \eta^{\nu_j} \prod_{j > J_0} \frac{(1 + \nu_j)^b}{\rho_j^{\nu_j}}, \tag{4.23}$$

with $\eta := \frac{1+\kappa}{2\kappa} < 1$ and $C := C_{\frac{r}{2}} C_0^{J_0}$. Using the inequality $1 + x \leq e^x$ for $x > 0$, we deduce that for $j > J_0$ with $\nu_j \neq 0$

$$\frac{(1 + \nu_j)^b}{\rho_j^{\nu_j}} \leq \left(\frac{|\nu_F| d_j}{\nu_j}\right)^{\nu_j} \quad \text{with} \quad d_j := \frac{4B\|\psi_j\|_{L^\infty(D)}}{r}, \tag{4.24}$$

so we can write $p_\nu(b)\|t_\nu\|_V \le Cq_\nu \le C\alpha(\nu_E)\beta(\nu_F)$, where

$$\alpha(\nu_E) := \prod_{j \le J_0} \eta^{\nu_j}, \quad \beta(\nu_F) := \prod_{j > J_0} \Big(\frac{|\nu_F||d_j|}{\nu_j}\Big)^{\nu_j}. \tag{4.25}$$

By (4.20) we find that $\sum_{j>J_0} d_j \le \frac{1}{3}$. We then obtain the ℓ^p summability of $(q_\nu)_{\nu\in\mathcal{F}}$ exactly as in §3.2 in [13]. It follows that $(p_\nu(b)\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ is ℓ^p summable. $\qquad\square$

The ℓ^p summability of $(p_\nu(b)\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ implies by a standard result of Stechkin (see (3.13) in [13]) that there exists $C > 0$ and a sequence $(\Lambda_n)_{n\ge1}$ of sets $\Lambda_n \subset \mathcal{F}$ of cardinality not exceeding n such that

$$\sum_{\nu\notin\Lambda_n} p_\nu(b)\|t_\nu\|_V \le Cn^{-s}, \quad s := \frac{1}{p} - 1, \tag{4.26}$$

if each set Λ_n consists of the indices corresponding to n largest values of $p_\nu(b)\|t_\nu\|_V$. However, these sets are not necessarily monotone. In order to obtain a sequence of monotone sets, we observe that the sequence $(q_\nu)_{\nu\in\mathcal{F}}$ defined in (4.23) is monotone decreasing, in the sense that $\mu \le \nu \Rightarrow q_\nu \le q_\mu$. Indeed, on the one hand, $q_{\nu+e_j} = \eta q_\nu < q_\nu$ if $j \le J_0$. On the other hand, since $(2+n)^b \le 2^b(1+n)^b \le B(1+n)^b$ for any $n \ge 0$ and since $\frac{\nu_j+1}{|\nu_F|+1} \ge \frac{\nu_j}{|\nu_F|}$ for any $j > J_0$, then

$$\frac{(2+\nu_j)^b}{\Big(B + \frac{r(\nu_j+1)}{4(|\nu_F|+1)\|\psi_j\|_{L^\infty(D)}}\Big)^{\nu_j+1}} \le \frac{(1+\nu_j)^b}{\Big(B + \frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}\Big)^{\nu_j}}$$

which implies also that $q_{\nu+e_j} \le q_\nu$ when $j > J_0$. This shows that the sequence $(q_\nu)_{\nu\in\mathcal{F}}$ is *monotone decreasing* (i.e. monotonically decreasing with respect to the semiordering $\le$ of multiindices in $\mathcal{F}$ induced by monotonicity in the sense of Definition 1.11).

Since the sequence $(q_\nu)_{\nu\in\mathcal{F}}$ majorizes the sequence $(p_\nu\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ up to a multiplicative constant, and since its ℓ^p summability is established in the above theorem, we also obtain (4.26) with Λ_n a set of indices corresponding to the n largest values of q_ν. Since $(q_\nu)_{\nu\in\mathcal{F}}$ is monotone decreasing, such sets Λ_n can always be chosen to be monotone and nested. Combining these observations with Lemma 4.1, we thus have established the following convergence result.

Theorem 4.3 *Assume that if the diffusion coefficient satisfies the uniform ellipticity assumption* (1.6) *and one has* $(\|\psi_j\|_{L^\infty(D)})_{j\ge1} \in \ell^p(\mathbb{N})$ *for some* $0 < p < 1$. *If the univariate sequence* $(z_k)_{k\ge0}$ *is chosen so that* $\lambda_k \le (k+1)^\theta$ *for some* $\theta > 0$, *then there exists a sequence* $(\Lambda_n)_{n\ge1}$ *of monotone sets* Λ_n *such that* $\#(\Lambda_n) = n$ *and*

$$\|u - I_{\Lambda_n}u\|_{L^\infty(\mathcal{P},V)} \le Cn^{-s}, \quad s = \frac{1}{p} - 1. \tag{4.27}$$

Remark 4.4 *The above convergence result exploits the fact that the Taylor series itself converges, due to the holomorphic properties of the solution map. For other type of parametric model problems that do not have such properties, the Taylor series may fail to converge even if the parameter dependence is smooth. On the other hand, other types of polynomial approximation may still converge*

with a high rate, and therefore so may polynomial interpolation provided that the Lebesgue constant does not grow too fast. For such problems, we thus expect that our high dimensional interpolation algorithm behaves well, provided that the sets Λ_n are properly selected, either by a-priori estimates or by an adaptive strategy such as described in §2.5.

5 Numerical experiments

5.1 Scalar valued functions

We first consider the interpolation of scalar valued functions $u : \mathcal{P} \to \mathbb{R}$ where $\mathcal{P} = P^d$ with $P = [-1, 1]$. Our objective is to test the adaptive algorithm AI proposed in §2.5 in various ways:

- Ability to select good monotone index sets Λ_n, in particular when the function has anisotropic dependence on the variable.

- The effect of the choice of the univariate sequence $(z_k)_{k \geq 0}$, in particular on the robustness of the interpolation with respect to noise in the measurements.

- Robustness of the performance with respect to the variable dimension d when the function depends only of a few unkown variables, or when the depence with respect to the variables is sufficiently anisotropic.

We consider three possible choices for the univariate sequence $(z_k)_{k \geq 0}$:

- Uniform sequence (U): $z_0 = 1$, $z_1 = -1$, $z_2 = 0$ and for $k > 1$ we set $z_{2k+1} = \frac{1}{2} \sum_{j=0}^{n} \varepsilon_j 2^{-j}$ where $k = \sum_{j=0}^{n} \varepsilon_j 2^j$ is the binary expansion of k and $z_{2k+2} = -z_{2k+1}$. Such a choice produces a uniform subdivision of $[-1, 1]$ of step size 2^{-j} for the particular sections $(z_0, \ldots, z_{2^j})$, and avoids accumulation of points on a region of the interval for the intermediate sections $(z_0, \ldots, z_k)$, $2^j < k < 2^{j+1}$.

- Leja sequence (L): $z_0 = 1$ and the sequence z_k is defined recursively on $[-1, 1]$ by (2.23).

- R-Leja sequence (R): this is the projection on $[-1, 1]$ of a Leja sequence for the complex unit disk initiated at 1. The R-Leja sequence has an explicit structure which is very similar to that of the sequence U in the sense that $(z_0, \ldots, z_{2^j})$ are Clenshaw-Curtis points, that is, the projections on the real axis of a uniform subdivision of the upper half-circle with end-points at -1 and 1, see [8, 9, 10].

Our first example is the function of $d = 16$ variables

$$u_1(\mathbf{y}) = u(y_1, \ldots, y_{16}) = y_3 \sin(y_4 + y_{16}), \tag{5.1}$$

that in fact only depends on 3 variables. Figure 5.1 displays the interpolation error in terms of $n = \#(\Lambda_n)$ for the AI algorithm based on 3 possible choices U, L, R, of univariate sequence $(z_k)_{k \geq 0}$.

20

We observe that the error decays, after a (short) preasymptotic phase (required to identify the set $\{3, 4, 16\}$ of "active" dimensions) reaches machine precision when the AI algorithm is based on the choices L and R for the univariate sequence, but not with the choice U, although u_1 is analytic over $\mathbb{R}^d$. Inspection of the index sets Λ_n generated by the algorithm for the three choices reveals that for $n = 10^3$, the polynomial degree in the active variables y_4 and y_{16} reaches values above 30. This explains the bad behaviour of the error for the choice U. Indeed, it is well known that the univariate Lebesgue constant associated to k uniformly spaced points in $(-1, 1)$ grows faster than 2^k (see, e.g. [14]).

This implies that, in the range of polynomial degrees k activated by the AI algorithm based on the sequence U, the Lebesgue constants become so large, that the (inevitable) roundoff errors in the float point evaluation of the function $u_2(\mathbf{y})$ in (5.1) are amplified to $O(1)$ contributions to the interpolant. This does not occur with the choices L and R for which the Lebesgue constant has moderate growth. For these sequences, the algorithm identifies the three active variables (y_3, y_4, y_{16}) in the sense that all chosen indices ν have $\nu_j = 0$ for $j \neq 3, 4, 16$.
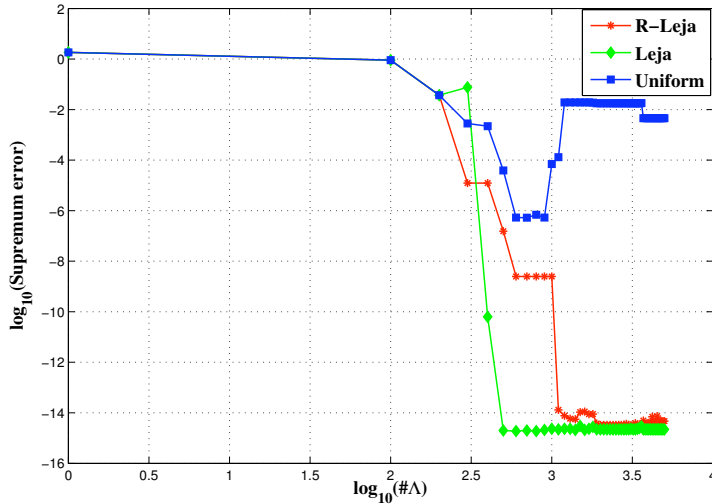
Figure 5.1: Estimated L^∞ error of the AI algorithm applied to u_1 based on the sequences R, L and U.

In our next example, we consider the parametric, scalar function

$$u_2(\mathbf{y}) = (1 + \sum_{j=1}^{d} \gamma_j y_j)^{-1}, \quad \gamma_j := \frac{3}{5j^3} . \tag{5.2}$$

This function now depends on all variables $y_1, \ldots, y_d$ but in a strongly anisotropic way due to the decay of the weights γ_j. Since $\sum_{j=1}^{\infty} \gamma_j \approx 0.72 < 1$, the function u admits an holomorphic continuation (that is, an extension which is holomorphic in each variable) in the complex polydisc

21

domain

$$\mathcal{P}_{\mathbb{C}} = (P_{\mathbb{C}})^d, \tag{5.3}$$

where $P_{\mathbb{C}}$ is the unit disk of $\mathbb{C}$, regardless of the dimension d. Moreover, an analysis completely analogous to that used in [13] to prove (1.8) for parametric PDE's, based on Cauchy's integral formula, reveals that since $(\gamma_j)_{j \geq 1} \in \ell^p(\mathbb{N})$ for $p > \frac{1}{3}$, there exists a sequence $(\Lambda_n^*)_{n \geq 1}$ with $n = \#(\Lambda_n^*)$ such that

$$\inf_{v \in \mathbb{P}_{\Lambda_n^*}} \|u - v\|_{L^\infty(\mathcal{P})} \leq C_s n^{-s}, \quad n \geq 1, \tag{5.4}$$

for all $s < 2$ where C_s is *independent* of the number d of active coordinates. Figure 5.2 reveals that this robustness with respect to d is also observed when using the AI algorithm (here based on the sequence R), since its convergence behaviour is almost unchanged for $d = 8$, 16, 32 and 64.
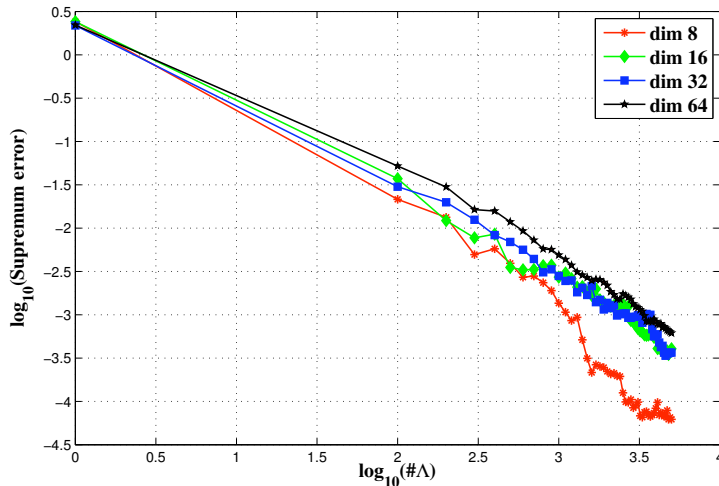
Figure 5.2: Estimated L^∞ error for the AI algorithm applied to u_2 based on the sequences R for the dimensions $d = 8, 16, 32, 64$.

We next fix $d = 16$ and compare the error of the AI algorithm applied to u_2 based on the three sequences L, R and U. Figure 5.3 reveals that, in contrast to the function u_1, the uniform sequence U gives as good results as the sequences L and R. This can be explained by inspecting more closely the index sets Λ_n, for which one finds that for $n = 10^4$ the highest polynomial degree attained on the most significant variable y_1 is 17 (due the presence of many active variables) and the amplification of roundoff in function evaluations to $O(1)$ is not yet visible.

To verify that the preceding results and findings are not a consequence of the particular functional form of the exact solutions, we repeat the previous experiment with a small amplitude added random perturbation. Specifically, we now adaptively interpolate the values $u_2(\mathbf{z}_\nu) + \varepsilon_\nu$ where ε_ν are independent realizations of a random variable with uniform law on $[-10^{-3}, 10^{-3}]$. Figure

22
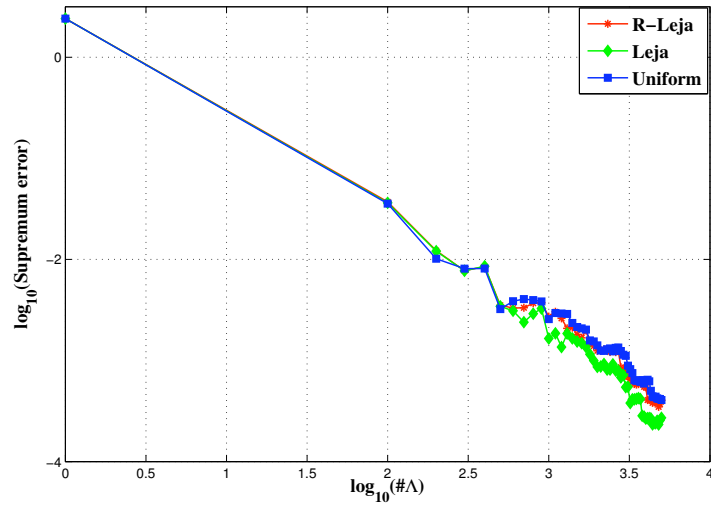
Figure 5.3: Estimated L^∞ error of the AI algorithm applied to u_2 with $d = 16$ based on the sequences R, L and U.
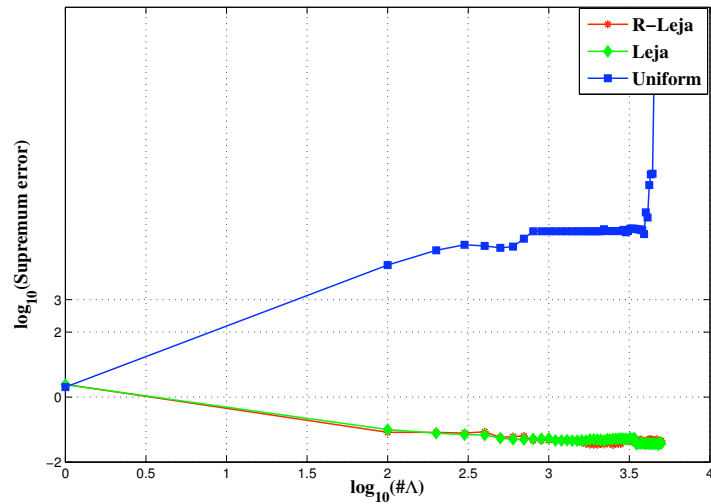

Figure 5.4: Estimated L^∞ error of the AI algorithm applied to a noisy evaluation of u_2 with $d = 16$ based on the sequences R, L and U.

5.4 reveals that the error diverges when using the U sequence, while it decays when using R or L (however not reaching arbitrarily small values due to presence of the noise).
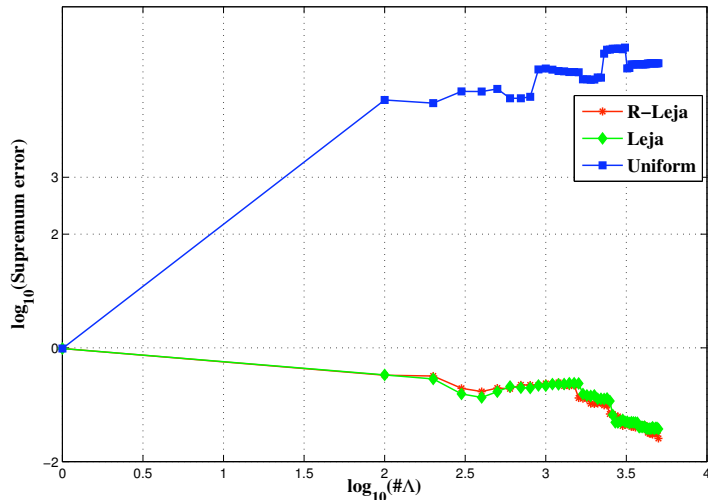
23

Figure 5.5: Estimated L^∞ error of the AI algorithm applied to u_3 with $d = 16$ based on the sequences R, L and U.

Finally, we consider with $d = 16$ the function

$$u_3(\mathbf{y}) = \big(1 + (\sum_{j=1}^{d} \gamma_j y_j)^2\big)^{-1}, \quad \gamma_j := \frac{5}{j^3}. \tag{5.5}$$

Similar to u_2, this function has an anisotropic behaviour. It also has an holomorphic continuation in a complex neighbourhood of $\mathcal{P}$, however, it has smaller domain of analyticity than the polydisc $\mathcal{P}_\mathbb{C}$ due to the fact that $\sum_{j=1}^{d} \gamma_j > 1$. As a consequence, the AI algorithm based on the uniform sequence U does not converge *even in the noiseless case*, as illustrated by Figure 5.5. This should be viewed as a manifestation of the well known Runge phenomenon.

In summary, the AI algorithm takes advantage of an anisotropic dependence on the variables, however its success is critically tied to the choice of the univariate sequence $(z_k)_{k\geq 0}$ in either one of these situations: (i) the polynomial degree reaches high values in certain variables, (ii) the measurements (ie. function evaluations) are "noisy", (iii) the function does not have sufficient smoothness in certain variables. In all cases, both sequences R and L are good choices. The convergence rates are dimension independent if the function u allow for dimensionally independent best N-term approximation rates. The practical performance of the AI algorithms are found to be robust with respect to the choice of univariate sequence $(z_j)_{j\geq 1}$ *as long as the sectional Lebesgue constants λ_k grow algebraically as k^θ with moderate exponent θ.*

5.2 Parametric PDE's

We now turn to the interpolation of functions, $u : \mathcal{P} \to V$ defined as the solution map of a parametric PDE (1.1) where V is the solution space. In practice, the PDE is solved by a numerical technique such as the finite element method applied with a certain mesh, and therefore we rather interpolate the numerical solution map

$$u_h : \mathcal{P} \to V_h, \tag{5.6}$$

where $V_h \subset V$ is finite dimensional and independent of the particular interpolation point in U.

In [11], several adaptive algorithms based on the Taylor partial sums (4.6) were proposed, analyzed and implemented for the model elliptic PDE (1.5). The most practically efficient of these algorithms acts in a very similar way as the AI algorithm, in the sense that the sets Λ_{n+1} is defined by adding to Λ_n the index ν that maximizes the V-norm of the Taylor coefficient t_ν among the set of neighbours $\mathcal{N}(\Lambda_n)$. In the sequel, we refer to this adaptive algorithm as Largest Neigbour Taylor (LNT). Note that, in contrast to AI, algorithms based on the computation of the Taylor series such as LNT are by essence intrusive and strongly benefit from the particular structure of the problem (1.5): a linear equation with affine dependence of the operator on the parameters. This struture allows us to do a recursive computation of the Taylor coefficients t_ν which are solutions to the boundary value problems

$$\int_D \bar{a} \nabla t_\nu \nabla v = - \sum_{j \, : \, \nu_j \neq 0} \int_D \psi_j \nabla t_{\nu - e_j} \nabla v, \quad v \in V. \tag{5.7}$$

These problems be derived by partial differentiation at $\mathbf{y} = 0$ of the variational formulation

$$\int_D a(\mathbf{y}) \nabla u(\mathbf{y}) \nabla v = \int_D f v, \quad v \in V, \tag{5.8}$$

and we refer to [13] for a rigourous justification.

We compare the two algorithms AI and LNT when applied to (1.5) with $D = [0,1]^2$ and diffusion coefficient $a(x, \mathbf{y})$ given by an expansion in the two dimensional Haar wavelet basis, similar to Test 2 in [11], namely

$$a(x, \mathbf{y}) := \bar{a}(x) + \sum_{l=0}^{L} \beta_l \sum_{i=1}^{3} \sum_{k \in \{0, \ldots, 2^l - 1\}^2} y_{l,k,i} h_{l,k}^i(x), \quad \bar{a} = 1. \tag{5.9}$$

In the above expansion,

$$h_{l,k}^i(x) := h^i(2^l x - k), \ \ l \in \mathbb{N}, \ \ k = (k_1, k_2) \in \{0, \ldots, 2^l - 1\}^2, \ i = 1, 2, 3, \tag{5.10}$$

where the generating wavelets h^i are defined by $h^1(x_1, x_2) := \varphi(x_1) h(x_2)$, $h^2(x_1, x_2) := h(x_1) \varphi(x_2)$ and $h^3(x_1, x_2) := h(x_1) h(x_2)$, with $\varphi := \chi_{[0,1]}$ and $h := \chi_{[0,1/2[} - \chi_{[1/2,1[}$. Using the relabelling

$$\psi_j := \beta_l h_{l,k}^i \ \text{ and } \ y_j := y_{l,k,i}, \ \text{ when } \ j = 2^{2l} + 3(2^l k_1 + k_2) + i - 1, \tag{5.11}$$

25

we may rewrite the above expansion (5.9) in the form $a(x, y) := \bar{a}(x) + \sum_{j=1}^{d} y_j \psi_j(x)$ adopted in this paper, with $d := 2^{2(L+1)} - 1$. We consider the general form

$$\beta_l := c2^{-\gamma l}, \quad c := 0.3(1 - 2^{-\gamma}), \tag{5.12}$$

which ensures that the uniform ellipticity assumption **UEA** holds with $r = 0.1$ and $R = 1.9$, regardless of the parametric dimension d. The value of the parameter $\gamma > 0$ reflects the decay of the high scale oscillation and therefore the long range correlation in the diffusion field.

In our numerical test, we use the value $\gamma = 3$, which was among those tested in [11] and we consider the maximal scale levels $L = 1$ and 2 which give parametric dimension $d = 15$ and 63. In order to refine the comparison between AI and LNT, we introduce a third process that builds the interpolants $I_{\Lambda_n} u$ by using the sets Λ_n produced by the LNT algorithm. We refer to this algorithm as LNTI (Largest Neighbour Taylor Interpolation). For both AI and LNTI we use the univariate sequence R.

Several observations may be drawn from the error curves, displayed on Figures 5.6 and 5.7. We first notice that the error curve of LNTI is close to that of LNT, with a more oscillatory behaviour. Since the sets $(\Lambda_n)_{n \geq 1}$ are the same for both algorithms, these oscillations may reflect the non-monotonic growth Lebesgue constant of the interpolation operator I_{Λ_n} that is used in LNTI. In addition, we notice that the error curve of AI is above that of LNTI, which means that AI is slightly misled in the adaptive selection of the sets Λ_n which was better performed by LNT. In all cases, we find that these algorithms are robust with respect to the growth in the dimension, and are therefore capable of capturing the anisotropic structure of the problem reflected by the decay in the weights β_l.
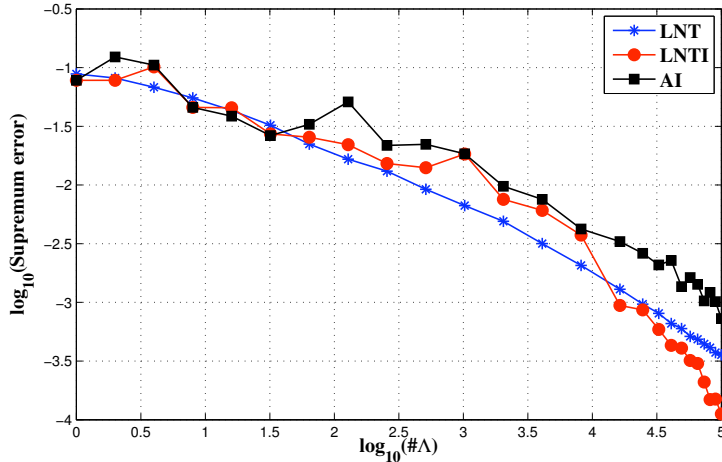

Figure 5.6: Estimated $L^\infty(\mathcal{P}, V)$ error of LNT, LNTI and AI for the model problem (1.5) with coefficients (5.9) and $d = 15$.
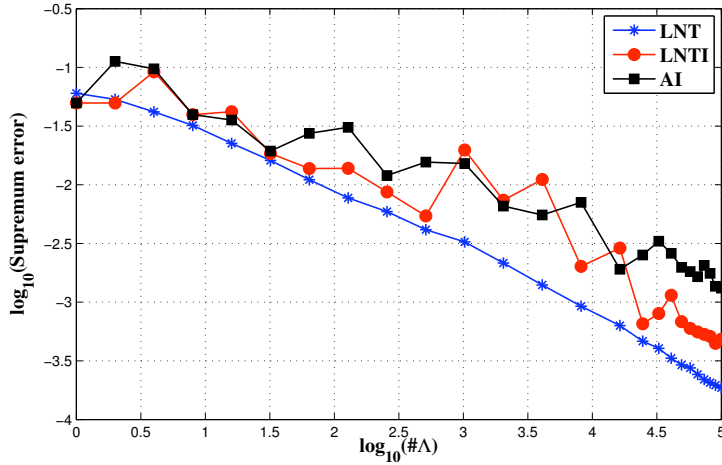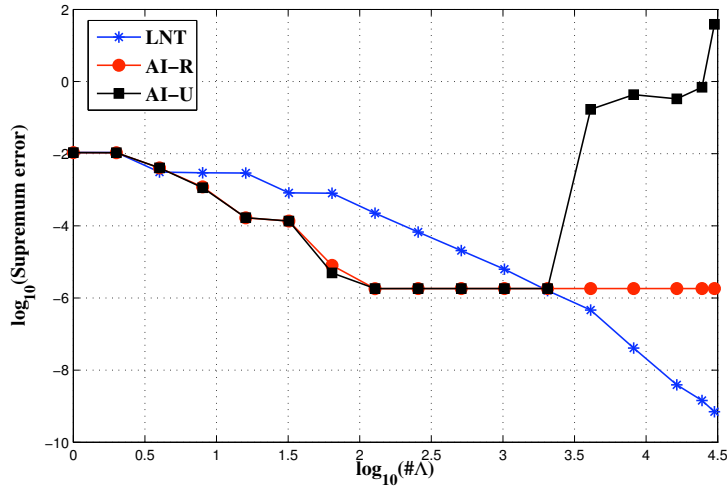
Figure 5.7: Estimated $L^\infty(\mathcal{P}, V)$ error of LNT, LNTI and AI for the model problem (1.5) with coefficients (5.9) and $d = 63$.

The effectiveness of Taylor approximations in the above problem is due to the fact that the map $\mathbf{y} \mapsto u(\mathbf{y})$ has an holomorphic continuation in the complex polydisc $\mathcal{P}_{\mathbb{C}}$, similar to the scalar valued function u_2. This approach becomes uneffective for PDE's where such continuation does not hold, even if the dependence on $\mathbf{y}$ remains analytic over $\mathcal{P}$. Here are two instances of such problems:

- Parametric nonlinear diffusion equations of the form

$$-\operatorname{div}(a\nabla u) + F(u) = f, \qquad (5.13)$$

with similar assumptions on the diffusion coefficient a as in (1.5) and F an analytic nonlinear function. In the case where f is small enough, a perturbation approach leads to a similar holomorphic extension over the polydisc $\mathcal{P}_{\mathbb{C}}$ as in the linear case, see [20]. In the case where f is arbitrary, and F is such that the problem remains coercive in $H_0^1(D)$ (for example $F(u) = u^3$ in spatial dimension 2 or 3), analyticity still holds on $\mathcal{P}$, however we are not ensured that the holomorphic extension can be carried over the whole of $\mathcal{P}_{\mathbb{C}}$.

- Linear PDE's set in parameter dependent domains. Consider for instance

$$-\Delta u = f \ \text{ in } \ D = D(\mathbf{y}) = \{(x_1, x_2) \ 0 \le x_1 \le 1, \ 0 \le x_1 \le \phi(x_1, \mathbf{y})\}, \qquad (5.14)$$

where the upper boundary has a general form $\phi(x_1, \mathbf{y}) = \bar{\phi}(x_1) + \sum_{j=1}^{d} y_j \phi_j(x_1) \ge \phi_{\min} > 0$. By a change of variable $x_2 \mapsto x_2\phi(x_1)$ mapping the reference domain $\bar{D} = [0,1]^2$ into $D(\mathbf{y})$, we are led to the elliptic PDE (1.5) over $\bar{D}$ however with a matrix coefficient $a(x, \mathbf{y})$ depending in a rational manner of $\mathbf{y}$ rather than affine. Here again, the dependence on $\mathbf{y}$ is analytic over the real domain $\mathcal{P}$, however with no guarantee that the holomorphic extension can be carried over the whole of $\mathcal{P}_{\mathbb{C}}$.

27

For such problems, we expect that Taylor based algorithms such as LNT fail, while AI still converges. We illustrate this by studying a simpler problem, namely (1.5) with diffusion coefficients of the form

$$a(x, \mathbf{y}) = \bar{a}(x) + (\sum_{j=1}^{d} y_j \psi_j(x))^2. \tag{5.15}$$

Assuming that $\bar{a}(x) \geq r > 0$, the problem is always coercive and the map $\mathbf{y} \mapsto u(\mathbf{y})$ has analytic dependence on $\mathbf{y} \in \mathcal{P}$. However, when the function ψ_j are chosen too large, the holomorphic extension does not hold over the whole of $\mathcal{P}_{\mathbb{C}}$. This example should be viewed as similar to the scalar function u_3.
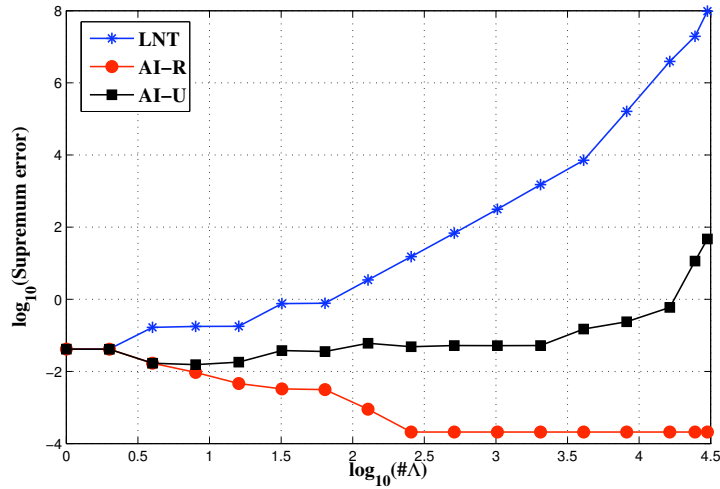

Figure 5.8: Estimated $L^\infty(\mathcal{P}, V)$ error of LNT and of AI using U and R sequences, for the model problem (1.5) with coefficients (5.15) and $\gamma = 0.5$.

Due to the polynomial form of $a(x, \mathbf{y})$ in the variable $\mathbf{y}$, it is still possible to compute the Taylor coefficients t_ν in a recursive manner that generalizes (5.7). Indeed, if $a(x, \mathbf{y})$ has the general form

$$a(x, \mathbf{y}) := \bar{a}(x) + \sum_{\nu \in \Lambda} \mathbf{y}^\nu \psi_\nu(x), \tag{5.16}$$

where Λ is a fixed set of index that does not include $0_\mathcal{F}$, then partial differentiation of (5.8) shows that

$$\int_D \bar{a} \nabla t_\nu \nabla v = - \sum_{\mu \in \Lambda \,:\, \mu \leq \nu} \int_D \psi_\mu \nabla t_{\nu - \mu} \nabla v, \quad v \in V. \tag{5.17}$$

This allows us to again perform LNT in the case where a is has the quadratic form (5.15).

As an example, we take $d = 4$, $\bar{a}(x) = 1$ and

$$\psi_j(x) = \frac{\gamma}{j^3} \phi_j(x), \tag{5.18}$$

28

Figure 5.9: Estimated $L^\infty(\mathcal{P}, V)$ error of LNT and of AI using U and R sequences, for the model problem (1.5) with coefficients (5.15) and $\gamma = 2$.

where $\gamma > 0$ and

$$\phi_1(x) = 1, \quad \phi_2(x) = \cos(2\pi x_1), \quad \phi_3(x) = \cos(2\pi x_2) \quad \text{and} \quad \phi_4(x) = \cos(2\pi(x_1 + x_2)). \qquad (5.19)$$

We compare LNT with AI, using both U and R univariate sequences for the second algorithm.

Figures 5.8 and 5.9 show the error curves, for both choices $\gamma = 0.5$ and $\gamma = 2$. When $\gamma = 0.5$, the LNT algorithm still converges due to the fact that the holomorphic extension holds over the whole of $\mathcal{P}_\mathbb{C}$, but it diverges for the larger value $\gamma = 2$. In contrast, the error still decreases when using AI based on the R univariate sequence. Similar to the scalar test case u_3, the same algorithm diverges if we use the U sequence, due to the fast growth of the univariate Lebesgue constants λ_k for the U sequence.

References

[1] R. Andreev, M. Bieri and Ch. Schwab, *Sparse Tensor Discretization of Elliptic sPDEs* SIAM J. Sci. Comput. 31, 4281-4304, 2009.

[2] R. Andreev and Ch. Schwab, Sparse Tensor Approximation of Parametric Eigenvalue Problems in Numerical Analysis of Multiscale Problems (I.G. Graham, T.Y. Hou, O. Lakkis and R. Scheichl (Eds)), Springer Lecture Notes in Comp. Sci. Engg. Vol. 83, pp. 203-242 Springer Verlag (2011).

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

[4] I. Babuska, R. Tempone and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal. 42, 800-825, 2004.

[5] J. Bäck, F. Nobile, L. Tamellini and R. Tempone, *On the optimal polynomial approximation of stochastic PDEs by Galerkin and Collocation methods*, MOX Report 23/2011, to appear in Math. Mod. Methods Appl. Sci., 2011.

[6] V. Barthelmann, E. Novak and K. Ritter, *High dimensional polynomial interpolation on sparse grids.* Adv. Comput. Math. 12-4, 273-288, 2000.

[7] A. Buffa, Y. Maday, A. T. Patera, C. Prudhomme, and G. Turinici, *A priori convergence of the greedy algorithm for the parameterized reduced basis*, M2AN 46-3, 595-603, 2012.

[8] J.P. Calvi and V.M. Phung, *On the Lebesgue constant of Leja sequences for the unit disk and its applications to multivariate interpolation*, Journal of Approximation Theory 163-5, 608-622, 2011.

[9] J.P. Calvi and V.M. Phung, *Lagrange interpolation at real projections of Leja sequences for the unit disk*, to appear in the Proceedings of the American Mathematical Society, 2012.

[10] A. Chkifa, *On the Lebesgue constant of Leja sequences for the complex unit disk and of their real projection*, preprint Laboratoire Jacques-Louis Lions, submitted to Journal of Approximation Theory, 2012.

[11] A. Chkifa, A. Cohen, R. DeVore and Ch. Schwab, *Sparse adaptive Talor approximation algorithms for parametric and stochastic elliptic PDEs*, to appear in M2AN 2012.

[12] A. Cohen, R. DeVore and Ch. Schwab, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Journ. Found. Comp. Math. 10-6, 615-646, 2010.

[13] A. Cohen, R. DeVore and Ch. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic PDE's*, Analysis and Applications 9, 1-37, 2011.

[14] D. Coppersmith and T.J. T.J. Rivlin, *The growth of polynomials bounded at equally spaced points.* SIAM J. Math. Anal, 23(4):970983, 1992.

[15] Ph. J. Davis, *Interpolation and Approximation*, Blaisdell Publishing Company, 1963.

[16] C. de Boor and A. Ron, *Computational aspects of polynomial interpolation in several variables*, Mathematics of Computation 58, 705-727, 1992.

[17] Ph. Frauenfelder, Ch. Schwab and R.A. Todor: Finite elements for elliptic problems with stochastic coefficients Comp. Meth. Appl. Mech. Engg. 194, 205-228, 2005.

[18] C.J. Gittelson, *An adaptive stochastic Galerkin method*, to appear in Math. Comp. 2012.

[19] T. Gerster and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing 71-1, 2003.

[20] M. Hansen and Ch. Schwab, *Analytic regularity and nonlinear approximation of a class of parametric semilinear elliptic PDEs* Report 2011-29, Seminar for Applied Mathematics, ETH Zürich (2011) http://www.sam.math.ethz.ch/reports/2011/29

[21] M. Hansen and Ch. Schwab, *Analytic regularity and best N-term approximation of high dimensional parametric initial value problems*, Report 2011-64, Seminar for Applied Mathematics, ETH Zürich (2011) http://www.sam.math.ethz.ch/reports/2011/64

[22] V.H. Hoang and Ch. Schwab, *Sparse tensor Galerkin discretizations for parametric and random parabolic PDEs I: Analytic regularity and gpc-approximation.* Report 2010-11, Seminar for Applied Mathematics, ETH Zürich, (in review)

[23] V.H. Hoang and Ch. Schwab, *Analytic regularity and gpc approximation for parametric and random 2nd order hyperbolic PDEs*, Analysis and Applications, 2012.

[24] M. Kleiber and T. D. Hien, *The stochastic finite element methods*, John Wiley & Sons, Chichester, 1992.

[25] J. Kuntzman, *Méthodes numériques - Interpolation, dérivées*, Dunod, Paris, 1959.

[26] G. Migliorati, F. Nobile, E. von Schwerin and R. Tempone, *Analysis of the discrete L^2 projection on polynomial spaces with random evaluations*, MOX Report 46/2011.

[27] R. Milani, A. Quarteroni and G. Rozza, *Reduced basis methods in linear elasticity with many parameters* Comp. Meth. Appl. Mech. Engg. 197, 4812-4829, 2008.

[28] V. Nistor and Ch. Schwab, *High order Galerkin approximations for parametric, second order elliptic partial differential equations* Report 2012-22, Seminar for Applied Mathematics, ETH Zürich, (in review).

[29] F. Nobile, R. Tempone and C.G. Webster, *A sparse grid stochastic collocation method for elliptic partial differential equations with random input data* , SIAM J. Num. Anal. 46, 2309-2345, 2008.

[30] F. Nobile, R. Tempone and C.G. Webster, *An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data* SIAM J. Num. Anal. 46, 2411-2442, 2008.

[31] S. Smolyak, *Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions*, Doklady Akademii Nauk SSSR 4, 240-243, 1963.

Abdellah Chkifa

UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

chkifa@ann.jussieu.fr

Albert Cohen

UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

cohen@ann.jussieu.fr

Christoph Schwab

Seminar for Applied Mathematics, ETH Zürich, CH 8092 Zürich, Switzerland

schwab@math.ethz.ch

Research Reports