Numerische
Mathematik

High-dimensional approximation with kernel-based multilevel methods on sparse grids

Rüdiger Kempf[1] · Holger Wendland[1]

Abstract

Moderately high-dimensional approximation problems can successfully be solved by combining univariate approximation processes using an intelligent combination technique. While this has so far predominantly been done with either polynomials or splines, we suggest to employ a multilevel kernel-based approximation scheme. In contrast to those schemes built upon polynomials and splines, this new method is capable of combining arbitrary low-dimensional domains instead of just intervals and arbitrarily distributed points in these low-dimensional domains. We introduce the method and analyse its convergence in the so-called isotropic and anisotropic cases.

Mathematics Subject Classification 65D12 · 65D15 · 46E22 · 41A63

1 Introduction

High-dimensional problems occur in various applications from physics, engineering and even business science. Typical examples comprise solutions to the Schroedinger equation, the Fokker-Planck equation, the Black-Scholes equation or the Hamilton-Jacobi-Bellmann equation, to name a few. Another frequent application is in the context of uncertainty quantification, where a partial differential equation with coefficients that depend on random variables or high-dimensional parameters has to be solved over a low-dimensional domain. The solution or more generally the quantity of interest is then also a high-dimensional function of the parameters.

For moderately high-dimensional problems, sparse grids, see for example [1] for an overview, have proven to be an efficient and reliable tool. The underlying procedure for building the high-dimensional approximation goes back to Smolyak [2] but has

---

✉ Holger Wendland
holger.wendland@uni-bayreuth.de

[1] Applied and Numerical Analysis, Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany

since then been frequently used and studied, see for example [3–8]. Another term which comes up in this context is the term *combination technique*, see for example [9–20].

The idea behind the Smolyak construction is to combine sequences of low-dimensional approximation schemes in an orderly fashion to obtain a high-dimensional approximation method which exhibits nearly the same approximation features as the low-dimensional schemes. So far, however, Smolyak's procedure has predominantly been used within the following setting, though there are notable exception, see for example [21] for other functions and [22] for other point sets. The low-dimensional problems were usually be defined on univariate domains and the approximation schemes were either splines or polynomials. While polynomials become more and more expensive if larger and larger data sets are used, they provide easily even spectral convergence. However, the possible point sets are usually restricted to Chebyshev or Clenshaw-Curtis points. In contrast, splines, are computationally more efficient but usually only produce low approximation orders. Both have in common that a generalisation to other low-dimensional domains are not-straight forward. This, however, might be desirable. For example, if the solution of a time-dependent partial differential equation is studied then the time interval is usually univariate while the spatial domain is higher-dimensional. Also, the restriction to only one-dimensional domains limits the structure of the resulting high-dimensional domains.

In this paper, we suggest to combine Smolyak's technique with that of multilevel radial basis functions. Radial basis functions are a well-established tool to tackle multivariate interpolation and approximation problems, see for example [23, 24]. They allow the user to build approximations and interpolations using arbitrarily distributed data sites in low-dimensional domains. Moreover, if compactly supported radial basis functions are combined with a multilevel strategy, see for example [25–27], they provide a stable and fast approximation method. The combination of radial basis functions with Smolyak's construction is not entirely new. For example, in [28, 29] Gaussians are used in this context. However, these papers are purely numerical. We propose to use multiscale compactly supported radial basis functions. We derive the method, provide several representations of the multivariate approximation operator and give rigorous error estimates. We do this for the classical Smolyak method over isotropic index sets, as well, as for the generalisation over anisotropic index sets.

The paper is organised as follows. In the next section we collect all required material on tensor product spaces to describe Smolyak's method in its most general form. In the Sect. 3 we collect the required material on radial basis functions, or more generally, on kernel-based approximation methods. Here, we review the multilevel method and provide first new representations of it. In the Sect. 4 we introduce the kernel-based tensor product multilevel method, derive different types of representation and provide a thorough error analysis. In the Sect. 5 we give some numerical examples.

2 Smolyak's construction

For higher dimensional approximation problems, the method of Smolyak, introduced in [2], is one of the standard approximation methods. The basic idea of Smolyak's

construction is to combine univariate approximation processes in an orderly fashion, resulting in a multivariate approximation process, which comes close to achieving the approximation power of the fully resolved high-dimensional product rule, using significantly less information. Even if the number of required information still depends exponentially on the space dimension, the dependence is significantly better than it is for classical product rules.

2.1 Tensor product spaces and operators

We start by giving a short introduction to tensor products of spaces, functions and operators. For the convenience of the reader we collect in this subsection all the relevant material. This will allow us to describe our methods in the most general form. For more details and proofs, we refer, for example, to [30–32].

Tensor product spaces are usually first defined algebraically by the so-called universal property. The definition is then extended to normed spaces.

Definition 2.1 Let $V^{(1)}, \ldots, V^{(d)}$ be linear spaces. The linear space T is called the *algebraic tensor product space* of $V^{(1)}, \ldots, V^{(d)}$, denoted by $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$, if there is a multilinear mapping $\phi : V^{(1)} \times \cdots \times V^{(d)} \to T$ having the following properties:

1. The mapping ϕ generates T, i.e. $T = \text{span}\{\phi(u^{(1)}, \ldots, u^{(d)}) : u^{(1)} \in V^{(1)}, \ldots, u^{(d)} \in V^{(d)}\}$.
2. For every multilinear mapping $\psi : V^{(1)} \times \cdots \times V^{(d)} \to H$ with a linear space H, there is a unique linear map $f : T \to H$ such that $f \circ \phi = \psi$.

The algebraic tensor product space always exists and is uniquely determined up to isomorphisms. From this, it particularly follows that it is essentially independent of the ordering of the building spaces $V^{(j)}$ and also associative. By definition, the algebraic tensor product space is the span of the *elementary* tensors $u^{(1)} \otimes \cdots \otimes u^{(d)}$ with $u^{(j)} \in V^{(j)}$, i.e.

$$T = \text{span}\left\{u^{(1)} \otimes \cdots \otimes u^{(d)} : u^{(j)} \in V^{(j)}, 1 \leq j \leq d\right\}.$$

A direct consequence of this is that a basis of T is given by the tensor products of bases of the $V^{(j)}$.

Before discussing normed tensor product spaces, we will shortly recall the concept of operators between algebraic tensor product spaces.

Definition 2.2 Let $U^{(j)}, V^{(j)}, 1 \leq j \leq d$, be linear spaces with algebraic tensor products $S := U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $A^{(j)} : U^{(j)} \to V^{(j)}$, $1 \leq j \leq d$, be linear operators. Then, the *tensor product operator* $A^{(1)} \otimes \cdots \otimes A^{(d)} : S \to T$ is defined by

$$[A^{(1)} \otimes \cdots \otimes A^{(d)}](u) = \sum_{i=1}^{n} A^{(1)} u_i^{(1)} \otimes \cdots \otimes A^{(d)} u_i^{(d)},$$

$$u = \sum_{i=1}^{n} u_i^{(1)} \otimes \cdots \otimes u_i^{(d)}, \quad n \in \mathbb{N}.$$

It can be shown that this is well-defined, i.e. independent of the actual representation of u. Moreover, this is obviously a linear map. A specific case of this is given if the $V^{(j)} = \mathbb{R}$, i.e. if the $A^{(j)} : U^{(j)} \to \mathbb{R}$ are linear functionals. Using the fact that the tensor product of d copies of $\mathbb{R}$ is $\mathbb{R}$, the above definition yields for linear $\lambda^{(j)} : U^{(j)} \to \mathbb{R}$ the linear functional $\lambda^{(1)} \otimes \cdots \otimes \lambda^{(d)} : S \to \mathbb{R}$ defined by

$$[\lambda^{(1)} \otimes \cdots \otimes \lambda^{(d)}](u^{(1)} \otimes \cdots \otimes u^{(d)}) := \prod_{j=1}^{d} \lambda^{(j)}(u^{(j)}).$$

Remark 2.3 We will particularly be interested in spaces of functions. Hence, if $U^{(j)}$ and $V^{(j)}$ contain certain functions $u^{(j)} : \Omega^{(j)} \to \mathbb{R}$ and $v^{(j)} : \Omega^{(j)} \to \mathbb{R}$, then using the notations $u = u^{(1)} \otimes \cdots \otimes u^{(d)}$, $\mathbf{x} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(d)}) \in \Omega^{(1)} \times \cdots \times \Omega^{(d)}$ and $A = A^{(1)} \otimes \cdots \otimes A^{(d)}$, the elementary tensor products are given by

$$u(\mathbf{x}) = \prod_{j=1}^{d} u^{(j)}(\mathbf{x}^{(j)}),$$

$$Au(\mathbf{x}) = \prod_{j=1}^{d} [A^{(j)} u^{(j)}](\mathbf{x}^{(j)}).$$

This mainly follows from the uniqueness of the tensor product space and the fact that the universal property is easy to verify for the mapping $\phi : U^{(1)} \times \cdots \times U^{(d)} \to \mathrm{span}\{u^{(1)} \otimes \cdots \otimes u^{(d)}\}$ defined by $\phi(u^{(1)}, \ldots, u^{(d)}) = u^{(1)} \otimes \cdots \otimes u^{(d)}$.

We now turn to normed spaces. Unfortunately, the norms on the building blocks do not automatically lead to a norm on the algebraic tensor product space, unless the norms are given by inner products. Hence, it is usual to make the following assumptions and requirements.

Definition 2.4 Let $U^{(j)}, V^{(j)}$, $1 \leq j \leq d$, be normed spaces with algebraic tensor products $S := U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$.

1. A norm $\| \cdot \|_T : T \to \mathbb{R}$ is called a *crossnorm* if

$$\|v^{(1)} \otimes \cdots \otimes v^{(d)}\|_T = \prod_{j=1}^{d} \|v^{(j)}\|_{V^{(j)}}, \quad v^{(j)} \in V^{(j)}, 1 \leq j \leq d. \quad (1)$$

2. A norm $\| \cdot \|_T : T \to \mathbb{R}$ is called *reasonable* if

$$\|\lambda^{(1)} \otimes \cdots \otimes \lambda^{(d)}\|_{T^*} = \prod_{j=1}^{d} \|\lambda^{(j)}\|_{(V^{(j)})^*}, \quad \lambda^{(j)} \in (V^{(j)})^*, 1 \leq j \leq d. \quad (2)$$

Here, T^* denotes the dual space of T.

3. Two norms $\|\cdot\|_S : S \to \mathbb{R}$ and $\|\cdot\|_T : T \to \mathbb{R}$ are called *compatible* if

$$\|A^{(1)} \otimes \cdots \otimes A^{(d)}\|_{S \to T} = \prod_{j=1}^{d} \|A^{(j)}\|_{U^{(j)} \to V^{(j)}} \tag{3}$$

for all linear and bounded operators $A^{(j)} : U^{(j)} \to V^{(j)}$, $1 \le j \le d$.

If $\|\cdot\|_S$ and $\|\cdot\|_T$ are crossnorms, it is not too difficult to verify, see for example the comments after Definition 4.77 in [32], that (3) is equivalent to the easier verifiable

$$\|A^{(1)} \otimes \cdots \otimes A^{(d)}\|_{S \to T} \le \prod_{j=1}^{d} \|A^{(j)}\|_{U^{(j)} \to V^{(j)}}.$$

Having a norm on an algebraic tensor product space, we can, if necessary, complete the space with respect to this norm.

Definition 2.5 Let $V^{(1)}, \ldots, V^{(d)}$ be normed spaces with algebraic tensor product space $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $\|\cdot\|_T : T \to \mathbb{R}$ be a norm on T. Then, the *tensor product* space $\left(\bigotimes_{j=1}^{d} V^{(j)}, \|\cdot\|_T \right)$ is the completion of T under the norm $\|\cdot\|_T$.

Remark 2.6 As the tensor product spaces are the closure of the algebraic tensor product spaces under given norms, it is always possible to extend an operator $A^{(1)} \otimes \cdots \otimes A^{(d)} : S \to T$ to a bounded operator between the tensor product spaces $\bigotimes U^{(j)}$ and $\bigotimes V^{(j)}$ having the same norm, provided that the operator is bounded from $S \to T$, which is the case if the norms on S and T are compatible.

We are mainly interested in Hilbert spaces, in particular in Sobolev spaces.

Definition 2.7 Let $V^{(1)}, \ldots, V^{(d)}$ be (pre-)Hilbert spaces with algebraic tensor product $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$ and inner products $\langle \cdot, \cdot \rangle_{V^{(j)}}$, $1 \le j \le d$. Then, the *induced inner product* on T is defined by

$$\langle u, v \rangle = \prod_{j=1}^{d} \langle u^{(j)}, v^{(j)} \rangle_{V^{(j)}}, \qquad u = u^{(1)} \otimes \cdots \otimes u^{(d)}, \quad v = v^{(1)} \otimes \cdots \otimes v^{(d)},$$

and linear extension.

This indeed defines an inner product on the algebraic tensor product space. Moreover, the norm induced by this tensor product has all the required properties.

Proposition 2.8 *The induced norm on the algebraic tensor product space of (pre)-Hilbert spaces is a reasonable crossnorm. Induced norms are always compatible.*

We are particularly interested in Sobolev spaces. However, not necessarily standard Sobolev spaces $H^m(\Omega)$, which consists of all functions $u \in L_2(\Omega)$ having weak derivatives $D^\alpha u \in L_2(\Omega)$ for all $|\alpha| \leq m$ but in so-called mixed-order Sobolev spaces.

To introduce them, we will employ the following notation. Let $\Omega^{(j)} \subseteq \mathbb{R}^{n_j}$, $1 \leq j \leq d$, be given and $\Omega := \Omega^{(1)} \times \cdots \times \Omega^{(d)}$. For a function $u : \Omega \to \mathbb{R}$ and multi-indices $\boldsymbol{\alpha}^{(j)} = (\alpha_1^{(j)}, \ldots, \alpha_{n_j}^{(j)})^{\mathrm{T}} \in \mathbb{N}_0^{n_j}$, $1 \leq j \leq d$, we will write

$$
\begin{aligned}
D^{\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(d)}} u : &= \frac{\partial^{|\boldsymbol{\alpha}^{(1)}|+\cdots+|\boldsymbol{\alpha}^{(d)}|} u}{(\partial \mathbf{x}^{(1)})^{\boldsymbol{\alpha}^{(1)}} \cdots (\partial \mathbf{x}^{(d)})^{\boldsymbol{\alpha}^{(d)}}} \\
&= \frac{\partial^{|\boldsymbol{\alpha}^{(d)}|+\cdots+|\boldsymbol{\alpha}^{(d)}|} u}{(\partial x_1^{(1)})^{\alpha_1^{(1)}} \cdots (\partial x_{n_1}^{(1)})^{\alpha_{n_1}^{(1)}} \cdots (\partial x_1^{(d)})^{\alpha_1^{(d)}} \cdots (\partial x_{n_d}^{(d)})^{\alpha_{n_d}^{(d)}}}
\end{aligned}
$$

for both weak and strong derivatives, using also the notation $\mathbf{x}^{(j)} = (x_1^{(j)}, \ldots, x_{n_j}^{(j)})^{\mathrm{T}} \in \mathbb{R}^{n_j}$.

Definition 2.9 Let $\Omega^{(j)} \subseteq \mathbb{R}^{n_j}$, $1 \leq j \leq d$, be measurable. Let $\Omega = \Omega^{(1)} \times \cdots \times \Omega^{(d)}$. Then, the *Sobolev space of mixed regularity* $\mathbf{m} \in \mathbb{N}_0^d$ is defined as

$$
H_{\mathrm{mix}}^{\mathbf{m}}(\Omega) := \left\{ u \in L_2(\Omega) : D^{\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(d)}} u \in L_2(\Omega), |\boldsymbol{\alpha}^{(j)}| \leq m_j, 1 \leq j \leq d \right\}
$$

with inner product given by

$$
\langle u, v \rangle_{H_{\mathrm{mix}}^{\mathbf{m}}(\Omega)} = \sum_{|\boldsymbol{\alpha}^{(1)}| \leq m_1} \cdots \sum_{|\boldsymbol{\alpha}^{(d)}| \leq m_d} \langle D^{\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(d)}} u, D^{\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(d)}} v \rangle_{L_2(\Omega)}
$$

and norm given by

$$
\|u\|_{H_{\mathrm{mix}}^{\mathbf{m}}(\Omega)} := \left(\sum_{|\boldsymbol{\alpha}^{(1)}| \leq m_1} \cdots \sum_{|\boldsymbol{\alpha}^{(d)}| \leq m_d} \|D^{\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(d)}} u\|_{L_2(\Omega)}^2 \right)^{1/2} .
$$

Following the ideas of the proof for standard Sobolev spaces, it is easy to see that this is indeed a Hilbert space. Moreover, it is the tensor product space of the Sobolev spaces $H^{m_j}(\Omega_j)$ under the induced norm.

Lemma 2.10 *Let $\Omega^{(j)} \subseteq \mathbb{R}^{n_j}$, $1 \leq j \leq d$, be measurable. Let $\Omega = \Omega^{(1)} \times \cdots \times \Omega^{(d)}$. Then, with the induced norm on the tensor product space,*

$$
\left(\bigotimes_{j=1}^{d} H_j^{m_j}(\Omega^{(j)}), \|\cdot\| \right) = \left(H_{mix}^{\mathbf{m}}(\Omega), \|\cdot\|_{H_{mix}^{\mathbf{m}}(\Omega)} \right) .
$$

Proof This is a standard result for $\mathbf{m} = (0, \ldots, 0)^{\mathrm{T}}$, i.e. for the tensor product of $L_2(\Omega^{(j)})$, see for example [30]. However, the proof easily carries over to this more general situation. □

2.2 Smolyak's algorithm

We will describe the procedure in its most general form. To this end, we need the following concept. We will write $\boldsymbol{\mu} \leq \boldsymbol{\lambda}$ for two vectors $\boldsymbol{\mu}, \boldsymbol{\lambda} \in \mathbb{N}^d$ if $\mu_j \leq \lambda_j$ for $1 \leq j \leq d$. Similarly we will use the notation $\boldsymbol{\mu} < \boldsymbol{\lambda}$, $\boldsymbol{\mu} \geq \boldsymbol{\lambda}$ and $\boldsymbol{\mu} > \boldsymbol{\lambda}$. Moreover, we will employ the specific vectors $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{N}^d$ and $\mathbf{0} = (0, \ldots, 0)^{\mathrm{T}}$.

Definition 2.11 A set $\Lambda \subseteq \mathbb{N}^d$ is called *monotone* or *downwards closed* if $\boldsymbol{\lambda} \in \Lambda$ and $\boldsymbol{\mu} \in \mathbb{N}^d$ with $\boldsymbol{\mu} \leq \boldsymbol{\lambda}$ implies $\boldsymbol{\mu} \in \Lambda$.

In this paper we will consider two types of particular monotone index sets, which have been studied frequently in the literature. The first type is the one that was used originally by Smolyak [2] and in subsequent papers like [3, 7, 33] and is for $q, d \in \mathbb{N}$ given by

$$\Lambda = \Lambda(q, d) = \{\mathbf{i} \in \mathbb{N}^d : |\mathbf{i}| \leq q\}. \tag{4}$$

Obviously, this set makes only sense if $q \geq d$. It then consists of $\binom{q}{d}$ elements. To also treat functions which have a different regularity in each direction, this set was subsequently generalised, see for example [8, 34–36], to

$$\Lambda = \Lambda_{\boldsymbol{\omega}}(\ell, d) = \left\{ \mathbf{i} \in \mathbb{N}^d : \sum_{j=1}^d (i_j - 1)\omega_j \leq \ell \min_{1 \leq j \leq d} \omega_j \right\}, \tag{5}$$

using a positive, real weight vector $\boldsymbol{\omega} > \mathbf{0}$ and a threshold $\ell \in \mathbb{N}$. This is indeed a generalisation, as we obviously have for any constant vector $\boldsymbol{\omega} = (\omega, \ldots, \omega)^{\mathrm{T}} \neq \mathbf{0}$ the relation

$$\Lambda_{\boldsymbol{\omega}}(\ell, d) = \Lambda(\ell + d, d). \tag{6}$$

In contrast to the isotropic index set (4), it is not so easy to determine the exact number of elements in the index set (5) due to the real weight vector $\boldsymbol{\omega}$. However, we will need a bound on the number of elements later on, which we quote from [35, Lemma 5.3].

Lemma 2.12 *If the weight vector* $\boldsymbol{\omega} \in \mathbb{R}_+^d$ *is non-decreasingly ordered, i.e. if* $\omega_1 \leq \omega_2 \leq \cdots \leq \omega_d$ *then the cardinality of the set* $\Lambda_{\boldsymbol{\omega}}(\ell, d)$ *from* (5) *is bounded by*

$$\#\Lambda_{\boldsymbol{\omega}}(\ell, d) \leq \prod_{j=1}^d \left(\frac{\ell \omega_1}{j \omega_j} + 1 \right). \tag{7}$$

We remark several points regarding this lemma. First, the assumption on the weights $\boldsymbol{\omega}$ to be ordered is obviously no restriction since we can always permute the directions. Second, in the isotropic case (4) the estimate is sharp, as we have with constant $\boldsymbol{\omega}$ using also (6),

$$\#\Lambda_{\boldsymbol{\omega}}(\ell, d) = \binom{\ell + d}{d} = \frac{(\ell + d)!}{d!\,\ell!} = \frac{1}{d!} \prod_{j=1}^{d} (\ell + j) = \prod_{j=1}^{d} \left(\frac{\ell}{j} + 1\right).$$

There are several other publications with different estimates on the cardinality of $\Lambda_{\boldsymbol{\omega}}(\ell, d)$, see for example [36, Lemma 2.8] or [34, Proposition 3.5]. However, they all exhibit the same asymptotic behaviour of ℓ^d for $\ell \to \infty$ and the bound given in Lemma 2.12 seems to be the sharpest upper bound yet obtained.

With all this at hand, we can now give a formal definition of Smolyak's construction. We will do this in the most general form using only arbitrary linear spaces and their algebraic tensor products.

Definition 2.13 Let $U^{(1)}, \ldots, U^{(d)}$ and $V^{(1)}, \ldots, V^{(d)}$ be linear spaces with algebraic tensor products $S = U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $\Lambda \subseteq \mathbb{N}^d$ be a monotone index set. For $1 \leq j \leq d$ let $L^{(j)} := \max\{\lambda_j : \boldsymbol{\lambda} \in \Lambda\}$. Assume that there are operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$ for $1 \leq j \leq d$ and $1 \leq i \leq L^{(j)}$. Set $A_0^{(j)} := 0$ and define the difference operators

$$\Delta_i^{(j)} := A_i^{(j)} - A_{i-1}^{(j)} : U^{(j)} \to V^{(j)}, \qquad 1 \leq i \leq L^{(j)}, \quad 1 \leq j \leq d. \qquad (8)$$

Then, the *Smolyak operator* $\mathcal{A}_\Lambda : S \to T$ is defined as the tensor product operator

$$\mathcal{A}_\Lambda := \sum_{\mathbf{i} \in \Lambda} \Delta_{i_1}^{(1)} \otimes \cdots \otimes \Delta_{i_d}^{(d)}. \qquad (9)$$

Sometimes one also speaks of the *Smolyak algorithm* or *Smolyak procedure*.

Remark 2.14 If the spaces $U^{(j)}$, $V^{(j)}$, $1 \leq j \leq d$, are normed spaces and if all the operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$ are bounded, then the Smolyak operator can uniquely be extended from the algebraic tensor products to the tensor products, i.e. to an operator $\mathcal{A}_\Lambda : \bigotimes U^{(j)} \to \bigotimes V^{(j)}$, as long as the norms on the algebraic tensor product spaces are compatible, see Remark 2.6. In this situation, we will not distinguish between the Smolyak operator and its extension.

Remark 2.15 In the univariate setting $d = 1$ any monotone index Λ has the form $\Lambda = \{1, \ldots, L\}$. Hence, in this situation, the Smolyak operator simply becomes

$$\mathcal{A}_\Lambda = \sum_{i=1}^{L} \Delta_i^{(1)} = A_L^{(1)},$$

i.e. we simply have the given univariate operator of the highest index. For the isotropic index set we have $\Lambda(q, 1) = \{i \in \mathbb{N} : i \leq q\}$, i.e. $L = q$ and for the anisotropic index set we have $\Lambda_{\boldsymbol{\omega}}(\ell, 1) = \{i \in \mathbb{N} : (i - 1)\omega_1 \leq \omega_1 \ell\}$, i.e. $L = \ell + 1$.

We now want to discuss a different representation of $\mathcal{A}_\Lambda$, expressing it in the $A_i^{(j)}$ instead of the $\Delta_i^{(j)}$. This is basically based on the identity, see [33],

$$\Delta_{i_1}^{(1)} \otimes \cdots \otimes \Delta_{i_d}^{(d)} = \sum_{\substack{\boldsymbol{\beta} \in \{0,1\}^d \\ \mathbf{i} - \boldsymbol{\beta} \geq \mathbf{1}}} (-1)^{|\boldsymbol{\beta}|} A_{i_1 - \beta_1}^{(1)} \otimes \cdots \otimes A_{i_d - \beta_d}^{(d)},$$

where we also used the fact that $A_{i_1 - \beta_1}^{(1)} \otimes \cdots \otimes A_{i_d - \beta_d}^{(d)} = 0$ whenever one of the indices $i_j - \beta_j$ is zero. By setting $\mathbf{j} = \mathbf{i} - \boldsymbol{\beta}$, this leads to

$$
\begin{aligned}
\mathcal{A}_\Lambda &= \sum_{\mathbf{i} \in \Lambda} \sum_{\substack{\boldsymbol{\beta} \in \{0,1\}^d \\ \mathbf{i} - \boldsymbol{\beta} \geq \mathbf{1}}} (-1)^{|\boldsymbol{\beta}|} A_{i_1 - \beta_1}^{(1)} \otimes \cdots \otimes A_{i_d - \beta_d}^{(d)} \\
&= \sum_{\substack{\mathbf{j} \in \Lambda, \boldsymbol{\beta} \in \{0,1\}^d \\ \mathbf{j} + \boldsymbol{\beta} \in \Lambda}} (-1)^{|\boldsymbol{\beta}|} A_{j_1}^{(1)} \otimes \cdots \otimes A_{j_d}^{(d)}
\end{aligned}
\tag{10}
$$

This can be further simplified if the index sets mentioned above are used. For the isotropic index set $\Lambda(q, d)$, the following result can be found in [3, 7, 33].

Lemma 2.16 *The operator $\mathcal{A}_\Lambda$ of Definition 2.13 has for $\Lambda = \Lambda(q, d) = \{\mathbf{i} \in \mathbb{N}^d : |\mathbf{i}| \leq q\}$ the representation*

$$\mathcal{A}_{\Lambda(q,d)} = \sum_{\mathbf{i} \in P(q,d)} (-1)^{q - |\mathbf{i}|} \binom{d-1}{q - |\mathbf{i}|} A_{i_1}^{(1)} \otimes \cdots \otimes A_{i_d}^{(d)}, \tag{11}$$

where $P(q, d) = \{\mathbf{i} \in \mathbb{N}^d : q - d + 1 \leq |\mathbf{i}| \leq q\}$.

For the anisotropic index set of (5), i.e. $\Lambda = \Lambda_{\boldsymbol{\omega}}(\ell, d)$ a generalisation of the previous result can be found in, e.g., [8].

Lemma 2.17 *For $\Lambda = \Lambda_{\boldsymbol{\omega}}(\ell, d) = \left\{\mathbf{i} \in \mathbb{N}^d : \sum_{j=1}^d (i_j - 1)\omega_j \leq \ell \min_{1 \leq j \leq d} \omega_j\right\}$, the operator $\mathcal{A}_\Lambda$ of Definition 2.13 has the representation*

$$\mathcal{A}_{\Lambda_{\boldsymbol{\omega}}(\ell,d)} = \sum_{\mathbf{i} \in Y_{\boldsymbol{\omega}}(\ell,d)} \sum_{\substack{\boldsymbol{\beta} \in \{0,1\}^d \\ \mathbf{i} + \boldsymbol{\beta} \in \Lambda_{\boldsymbol{\omega}}(\ell,d)}} (-1)^{|\boldsymbol{\beta}|} A_{i_1}^{(1)} \otimes \cdots \otimes A_{i_d}^{(d)} \tag{12}$$

with $Y_{\boldsymbol{\omega}}(\ell, d) = \Lambda_{\boldsymbol{\omega}}(\ell, d) \backslash \Lambda_{\boldsymbol{\omega}}(\ell - \|\boldsymbol{\omega}\|_1 / \min \omega_j, d)$.

These representations are also called *combination techniques*, see [12].

We finally need yet another representation, which will be helpful for the error analysis later on. We will do this only for the two types of index sets we have in mind. The first one is for the isotropic set. We will formulate the result in a more general form than the one in [3, 7, 33]. However, it is obvious that the proof given there holds in this situation, as well.

Lemma 2.18 *Let $U^{(1)}, \ldots, U^{(d)}$ and $V^{(1)}, \ldots, V^{(d)}$ be linear spaces with algebraic tensor products $S = U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $E^{(j)} : U^{(j)} \to V^{(j)}$, $1 \le j \le d$, be linear operators with tensor product $\mathcal{E}^{(d)} = E^{(1)} \otimes \cdots \otimes E^{(d)} : S \to T$. Finally, let $\mathcal{A}_{\Lambda(q,d)}$ be the Smolyak operator defined by operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$. Then, using the notation $\tilde{q} := q - d$,*

$$
\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)} = \sum_{k=1}^{d-1} \sum_{\mathbf{i} \in \Lambda(\tilde{q}+k,k)} \bigotimes_{j=1}^{k} \Delta_{i_j}^{(j)} \otimes \left(E^{(k+1)} - A_{\tilde{q}+k+1-|\mathbf{i}|}^{(k+1)} \right) \otimes \bigotimes_{j=k+2}^{d} E^{(j)}
$$
$$
+ \left(E^{(1)} - A_{\tilde{q}+1}^{(1)} \right) \otimes \bigotimes_{j=2}^{d} E^{(j)}.
$$

The corresponding version for the anisotropic index sets $\Lambda_{\boldsymbol{\omega}}(\ell, d)$ comes again from [8]. Again, we formulate it slightly more general and use also Remark 2.15. To simplify the notation, we will write for $\boldsymbol{\omega} \in \mathbb{R}_+^d$, $\ell \in \mathbb{N}$ and $1 \le k \le d$ also $\Lambda_{\boldsymbol{\omega}}(\ell, k) = \{\mathbf{i} \in \mathbb{N}^k : \sum_{j=1}^{k} (i_j - 1)\omega_j \le \min_{1 \le j \le d} \omega_j\}$.

Lemma 2.19 *Let $U^{(1)}, \ldots, U^{(d)}$ and $V^{(1)}, \ldots, V^{(d)}$ be linear spaces with algebraic tensor products $S = U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $E^{(j)} : U^{(j)} \to V^{(j)}$, $1 \le j \le d$, be linear operators with tensor product $\mathcal{E}^{(d)} = E^{(1)} \otimes \cdots \otimes E^{(d)} : S \to T$. Finally, let $\mathcal{A}_{\Lambda_{\boldsymbol{\omega}}(\ell,d)}$ be the Smolyak operator defined by operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$. Then,*

$$
\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_{\boldsymbol{\omega}}(\ell,d)} = \sum_{k=1}^{d-1} \sum_{\mathbf{i} \in \Lambda_{\boldsymbol{\omega}}(\ell,k)} \bigotimes_{j=1}^{k} \Delta_{i_j}^{(j)} \otimes \left(E^{(k+1)} - A_{\widehat{i}_{k+1}-1}^{(k+1)} \right) \otimes \bigotimes_{j=k+2}^{d} E^{(j)}
$$
$$
+ \left(E^{(1)} - A_{\ell+1}^{(1)} \right) \otimes \bigotimes_{j=2}^{d} E^{(j)},
$$

where for $\mathbf{i} \in \Lambda_{\boldsymbol{\omega}}(\ell, k)$ the index $\widehat{i}_{k+1}$ is defined as

$$
\widehat{i}_{k+1} = \left\lfloor 2 + \ell \frac{\min_{1 \le j \le d} \omega_j}{\omega_{k+1}} - \sum_{j=1}^{k} (i_j - 1) \frac{\omega_j}{\omega_{k+1}} \right\rfloor.
$$

3 The kernel-based multilevel method

In this section, we will introduce the operators that we will employ as the operators $A_{i_j}^{(j)}$ in the Smolyak formula. To simplify the notation, for this section, we omit the explicit dependency on the direction j.

3.1 Kernel-based approximation

Let $\Omega \subseteq \mathbb{R}^n$ be the domain of interest. We will use positive definite functions for building approximations to functions $f \in C(\Omega)$ which are only known at discrete points $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subseteq \Omega$. Details on this topic can be found in [23]. A function $\Phi : \mathbb{R}^n \to \mathbb{R}$ is positive definite, if all matrices of the form $M_X = (\Phi(\mathbf{x}_i - \mathbf{x}_j))$ are symmetric and positive definite for all possible sets $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ of pairwise distinct points. There is a close connection between positive definite functions and *reproducing kernel Hilbert spaces*. These are Hilbert spaces H of functions $f : \Omega \to \mathbb{R}$, which possess a unique kernel $K : \Omega \times \Omega \to \mathbb{R}$ with $K(\cdot, \mathbf{x}) \in H$ for all $\mathbf{x} \in \Omega$ and $f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x})\rangle_H$ for all $\mathbf{x} \in \Omega$ and all $f \in H$. Every positive definite function Φ defines a unique Hilbert space $\mathcal{N}_K(\mathbb{R}^n)$ of functions, called the *native space* of Φ, in which $K(\mathbf{x}, \mathbf{y}) := \Phi(\mathbf{x} - \mathbf{y})$ is the translation invariant reproducing kernel.

For us the following result will be particularly important. For a proof we again refer to [23].

Lemma 3.1 *Let $\sigma > n/2$ and let $\Phi : \mathbb{R}^n \to \mathbb{R}$ be an integrable and continuous function having a Fourier transform satisfying*

$$c_1 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^{-\sigma} \leq \widehat{\Phi}(\boldsymbol{\omega}) \leq c_2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^{-\sigma}, \qquad \boldsymbol{\omega} \in \mathbb{R}^n, \tag{13}$$

with two fixed constants $0 < c_1 \leq c_2$. Then, $K : \mathbb{R}^n \times \mathbb{R}^n$ defined by $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y})$ is the reproducing kernel of the Sobolev space $H^\sigma(\mathbb{R}^n)$ if the latter space is equipped with the inner product

$$\langle f, g\rangle_{\mathcal{N}_K(\mathbb{R}^n)} := \int_{\mathbb{R}^n} \frac{\widehat{f}(\boldsymbol{\omega})\overline{\widehat{g}(\boldsymbol{\omega})}}{\widehat{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega}.$$

The induced norm is equivalent to the standard norm on $H^\sigma(\mathbb{R}^n)$.

It is even possible to choose Φ to have compact support in the closed unit ball $B_1[\mathbf{0}] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1\}$. In this situation, we can also scale Φ as $\Phi_\delta = \delta^{-n}\Phi(\cdot/\delta)$ with $\delta > 0$. The resulting function has support in $B_\delta[\mathbf{0}]$. Moreover, since scaling does not change the decay of the Fourier transform, K_δ defined by $K_\delta(\mathbf{x}, \mathbf{y}) = \Phi_\delta(\mathbf{x} - \mathbf{y})$ is also a reproducing kernel of $H^\sigma(\mathbb{R}^n)$.

We are particularly interested in two reconstruction processes based on positive definite functions: interpolation and penalised least-squares. The following well-known result makes kernel-based approximations particularly appealing. A proof of the first part can be found in [23] and for the second part in [37].

Lemma 3.2 *Let* $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subseteq \mathbb{R}^n$ *be given. Let* $\Phi : \mathbb{R}^n \to \mathbb{R}$ *be a positive definite function. Then, given* $f_1, \ldots, f_N \in \mathbb{R}$, *there is a unique function*

$$s = \sum_{j=1}^{N} \alpha_j \Phi(\cdot - \mathbf{x}_j) \tag{14}$$

satisfying $s(\mathbf{x}_i) = f_i$, $1 \le i \le N$. *The coefficients* $\boldsymbol{\alpha} \in \mathbb{R}^N$ *are determined by a linear system* $M\boldsymbol{\alpha} = \mathbf{f}$ *with* $M = (\Phi(\mathbf{x}_i - \mathbf{x}_j))$.

Moreover, for every $\lambda > 0$ *there is a unique function* s *of the form* (14) *solving*

$$\min_{u \in \mathcal{N}_K(\mathbb{R}^n)} \sum_{j=1}^{N} |u(\mathbf{x}_j) - f_j|^2 + \lambda \|u\|_{\mathcal{N}_K(\mathbb{R}^n)}^2,$$

where $\mathcal{N}_K(\mathbb{R}^n)$ *is the reproducing kernel Hilbert space of the kernel* K *induced by* Φ. *In this case, the coefficients can be computed via the linear system* $(M + \lambda I)\boldsymbol{\alpha} = \mathbf{f}$.

When working with a compactly supported, positive definite and radial kernel, there is an unfortunate trade-off principle, see for example [23, 27]. If the support of the kernel is kept fixed and the number of data sites is increased then the sequence of interpolants converges to the target function. However, the interpolation matrices become denser and denser, i.e. the ratio between non-zero entries and all entries remains constant, and the evaluation of the interpolant becomes more and more expensive. Eventually, the advantage of the compact support is lost. If, however, the ratio between the support radius and the "data density" is kept fixed then the interpolation matrices remain sparse and well-conditioned, evaluation can be done in constant time but the sequence of interpolants does not converge.

The remedy to this problem is described in the next subsection and will form the basis of our definition of the approximation operators in the Smolyak algorithm.

3.2 Multilevel approximation

The idea of the kernel-based multilevel approximation is to combine different kernel-based approximation spaces with a simple residual correction scheme. The latter is rather general and we start with its definition.

Definition 3.3 Let V be a linear space. Let $W_1, W_2, \ldots \subseteq V$ be a sequence of linear subspaces and let $V_i = W_1 + \cdots + W_i$ for $i \in \mathbb{N}$. The *residual correction scheme* computes for a given $f \in V$ a sequence of *local approximations* $s_i \in W_i$, *global approximations* $f_i \in V_i$ and errors $e_i \in V$ as follows. For initialisation, $f_0 = 0$ and $e_0 = f$. Then, recursively, for $i = 1, 2, \ldots$

- Determine a local approximant $s_i \in W_i$ to e_{i-1}.
- Set $f_i = f_{i-1} + s_i$.
- Set $e_i = e_{i-1} - s_i$.

Throughout this paper, we will assume that the local approximants are given by an operator $\mathcal{I}_i : V \to W_i$. In our applications these operators will always be linear.

Lemma 3.4 *If the local approximations $s_i \in W_i$ are computed with a specific operator $\mathcal{I}_i : V \to W_i$, i.e. if we have $s_i = \mathcal{I}(e_{i-1})$, then, for each $L \in \mathbb{N}$, there is a multilevel approximation operator $A_L : V \to V_L$ defined by*

$$A_L(f) := \sum_{i=1}^{L} s_i = \sum_{i=1}^{L} \mathcal{I}_i(e_{i-1}) \tag{15}$$

such that $f_L = A_L(f)$.

Proof This follows immediately from the definition, as the sequences of functions generated by the residual correction scheme satisfy the additional relations

$$f_i = s_1 + \cdots + s_i,$$
$$e_i = f - (s_1 + \cdots + s_i).$$

$\square$

We will now describe how we will use kernel-based approximation to build the approximation spaces W_i for the residual correction scheme. As can be expected there are two ingredients to this: point sets and kernels. We follow mainly [26].

Let $\Omega \subseteq \mathbb{R}^n$ again be given. We assume that we have a sequence of discrete, not necessarily nested data sets $X_1, X_2, \ldots$ given by $X_i := \{\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,N_i}\}$. For each point set X_i we define two geometric quantities

$$h_i := h_{X_i, \Omega} := \sup_{\mathbf{x} \in \Omega} \min_{\mathbf{x}_k \in X_i} \|\mathbf{x} - \mathbf{x}_k\|_2,$$

$$q_i := q_{X_i} := \frac{1}{2} \min_{m \neq k} \|\mathbf{x}_m - \mathbf{x}_k\|_2.$$

The first quantity is called *fill distance* or *mesh norm* and is a measure of how well the points in X_i cover the domain Ω, as it is the radius of the largest ball with center in Ω without a point from X_i. The second quantity is called *separation radius* and denotes the half of the smallest distance between two data sites. We are especially interested in so-called *quasi-uniform* data sets, i.e. there exists a constant $c_{qu} > 0$ such that

$$q_i \leq h_i \leq c_{qu} q_i, \quad i \in \mathbb{N}.$$

Concerning the fill distances of the data sites $X_1, X_2, \ldots$, we assume them to be monotonically decreasing, i.e. we assume them to satisfy for some fixed $\mu \in (0, 1)$ the relation $h_{i+1} \approx \mu h_i$.

Next, we define a kernel $K_i : \Omega \times \Omega \to \mathbb{R}$ for each level $i \in \mathbb{N}$. In theory, we could pick different kernels for every level i, choosing a smooth kernel for the first few levels and then, for higher levels, switching to a rougher one. But in this paper, we will define

the kernels on all of $\mathbb{R}^n$ and build them using the same function $\Phi : \mathbb{R}^n \to \mathbb{R}$. We will assume that this function has the properties stated in Lemma 3.1, i.e. the associated reproducing kernel Hilbert space $\mathcal{N}_K(\mathbb{R}^n)$ of the kernel $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y})$ is the Sobolev space $H^\sigma(\mathbb{R}^n)$.

We will, in addition, assume that Φ is compactly supported with support given by the closed unit ball $B_1[\mathbf{0}] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1\}$. Then, we can pick a new support radius $\delta_i > 0$ for each level and define the level-dependent kernels as

$$K_i(\mathbf{x}, \mathbf{y}) := \delta_i^{-n} \Phi\left(\frac{\mathbf{x} - \mathbf{y}}{\delta_i}\right), \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{16}$$

Obviously, $K_i(\cdot, \mathbf{y})$ has support in $B_{\delta_i}[\mathbf{y}]$ for fixed $\mathbf{y}$ and all these kernels are reproducing kernels of $H^\sigma(\mathbb{R}^n)$ but with different inner products.

Since we assume that the data sites $X_1, X_2, \ldots$ become denser and denser, the support radii of Φ_i are chosen to mimic this behaviour, i.e. we choose them smaller and smaller. Usually, we couple them to the fill distance via a constant $\nu > 0$ such that $\delta_i = \nu h_i$. This way, the number of sites in the support of K_i can be bounded independently of the level.

Definition 3.5 Let $V = C(\Omega)$. The *kernel-based multilevel approximation method* is given by the residual correction scheme with local approximation spaces

$$W_i := \text{span}\{K_i(\cdot, \mathbf{x}) \ \mathbf{x} \in X_i\},$$

where $K_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined in (16) and the sequences $\{X_i\}, \{h_i\}, \{\delta_i\}$ satisfy the conditions outlined in the paragraphs above.

Though there is some resemblance to other multiscale methods like wavelets, there are significant differences. For example, the local approximation spaces W_i are in general finite-dimensional for every i and built using shifts and scales of a fixed function over scattered data rather than a regular grid. Moreover, the function Φ does not satisfy any kind of refinement equation. However, there are also recent multiscale constructions, for example based on geometric harmonics, see [38], which also employ finite dimensional spaces and scattered points, but differ significantly from our approach as they are based on eigenfunction expansions.

Nonetheless, the interpretation of the method is similar to that of other multiscale methods. The idea is that we start by computing an approximant from $V_1 = W_1$ to f on X_1 using a large support radius δ_1. In the next step, we want to add more details and hence compute an approximant from W_2 on X_2 using a smaller support radius δ_2 to the error of the first step. The sum of these two approximants belongs to $V_2 = W_1 + W_2$ and should be a better approximation than the first one. Proceeding in this way, we have the above defined kernel-based approximation method. The method has initially be suggested in [39] and has then further been investigated in [25, 26, 40–43].

If the local approximations $s_i \in W_i$ are computed via $s_i = \mathcal{I}_i(e_{i-1})$ using a linear operator $\mathcal{I}_i : C(\Omega) \to W_i$ then the kernel-based multilevel approximation scheme allows us once again to write f_L as $f_L = A_L(f)$ with global approximation operators $A_L : C(\Omega) \to V_L$ given by (15).

It remains to discuss how we choose these approximation operators $\mathcal{I}_i$. In this paper, we will deal with two types, interpolation and penalised least-squares approximation, see Lemma 3.2.

Any operator $\mathcal{I}_i : C(\Omega) \to W_i$ has the form

$$\mathcal{I}_i f = \sum_{k=1}^{N_i} \alpha_{i,k} K_i(\cdot, \mathbf{x}_{i,k}), \qquad f \in C(\Omega), \tag{17}$$

with certain coefficients $\boldsymbol{\alpha}_i = (\alpha_{i,1}, \ldots, \alpha_{i,N_i})^T \in \mathbb{R}^{N_i}$. In the case of interpolation, these coefficients are determined by a linear system $M_i \boldsymbol{\alpha}_i = \mathbf{f}$ with the symmetric, positive definite and sparse matrix $M_i = (K_i(\mathbf{x}_{i,j}, \mathbf{x}_{i,k})) \in \mathbb{R}^{N_i \times N_i}$. These matrices have, under the conditions described above, a condition number which is independent of the level. Moreover, evaluating $\mathcal{I}_i f$ can be done in $\mathcal{O}(1)$. Thus, the associated multilevel operator A_L can be computed in $\mathcal{O}(L N_L)$ time and evaluated in $\mathcal{O}(L)$ time, which makes this method numerically extremely appealing.

Similarly, in the case of penalised least-squares approximation, we have the local approximation as

$$s_i := \underset{s \in \mathcal{N}_{K_i}(\mathbb{R}^n)}{\arg\min} \sum_{k=1}^{N_i} |f(\mathbf{x}_{i,k}) - s(\mathbf{x}_{i,k})|^2 + \lambda_i \|s\|^2_{\mathcal{N}_{K_i}(\mathbb{R}^n)}, \tag{18}$$

which is also given by a linear operator $s_i = \mathcal{I}_i f$ of the form (17), where the coefficients are determined by solving the linear system $(M_i + \lambda_i I)\boldsymbol{\alpha}_i = \mathbf{f}$, where M_i is again the kernel matrix $M_i = (K_i(\mathbf{x}_{i,j}, \mathbf{x}_{i,k}))$, see Lemma 3.2.

The numerical complexity is comparable to the one for interpolation.

We end this section with collecting convergence results for the multilevel method. Again, these results are proven in [26].

Proposition 3.6 *Let $\Omega \subseteq \mathbb{R}^n$ be a bounded domain with Lipschitz boundary. Let $X_1, X_2, \ldots$ be point sets in Ω with mesh norms $h_1, h_2, \ldots$ satisfying $c\mu h_i \leq h_{i+1} \leq \mu h_i$ for $i = 1, 2, \ldots$ with fixed constants $\mu \in (0, 1)$, $c \in (0, 1]$ and h_1 sufficiently small. Let Φ be a kernel generating $H^\sigma(\mathbb{R}^n)$, i.e. its Fourier transform satisfies (13), and let $K_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be defined by (16). with scale factor $\delta_i = \nu h_i$. Assume $1/h_1 \geq \nu \geq \gamma/\mu$ with a fixed $\gamma > 0$.*

1. *If the target function f belongs to $H^\sigma(\Omega)$ and if the multilevel method is build using interpolation operators then there are constants $C_1, C > 0$ such that*

$$\|f - A_L(f)\|_{L_2(\Omega)} \leq C(C_1 \mu^\sigma)^L \|f\|_{H^\sigma(\Omega)}, \qquad L = 1, 2, \ldots. \tag{19}$$

2. *If the target function f belongs to $H^\beta(\Omega)$ with $n/2 < \beta < \sigma$ and if there is a constant $c_q > 0$ such that $q_i \leq h_i \leq c_q q_i$ for $i = 1, 2, \ldots$ then there are constants $C_1, C > 0$ such that*

$$\|f - A_L(f)\|_{L_2(\Omega)} \leq C(C_1 \mu^\beta)^L \|f\|_{H^\beta(\Omega)}, \qquad L = 1, 2, \ldots. \tag{20}$$

3. *If the target function f belongs to $H^\sigma(\Omega)$ and if the smoothing parameters λ_i in (18) satisfy $\lambda_i \leq \kappa(h_j/\delta_j)^{2\sigma}$ with a fixed constant $\kappa > 0$ then then there are constants $C_1, C > 0$ such that*

$$\|f - A_L(f)\|_{L_2(\Omega)} \leq C(C_1\mu^\sigma)^L \|f\|_{H^\sigma(\Omega)}, \qquad L = 1, 2, \ldots. \qquad (21)$$

In all three cases convergences appears if the parameter μ is chosen sufficiently small.

The first case in the theorem above corresponds to approximating functions from $H^\sigma(\Omega)$ where the smoothness aligns to the smoothness of the kernel. In the second case rougher target functions $f \in H^\beta(\Omega)$ are approximated with a smoother kernel. Here, an additional stability assumption on the point distributions is required. In the third case penalised least-squares approximation for functions from $H^\sigma(\Omega)$ is discussed. Again, the smoothness is aligned to the smoothness of the kernel. Convergence is guaranteed if the smoothing parameters are chosen correctly. It is also possible to show convergence for less smooth target functions but we omit the details here.

Finally, the L_2-norm on the left-hand side can be replaced by other L_p or Sobolev norms. Again, we omit the details here.

3.3 Further representations

For their application within the Smolyak construction, we need a better understanding on how the residual correction operator in general and the more specific multilevel approximation operator depend on the target function f. This is the goal of this section and we will start with the general residual correction procedure.

Definition 3.7 Let V be a linear space and $L \in \mathbb{N}$. Let $W_1, \ldots, W_L \subseteq V$ be subspaces. Let $\mathcal{I}_i : V \to W_i$, $1 \leq i \leq L$ be given operators. For a subset $\mathfrak{u} = \{u_1, \ldots, u_j\} \subseteq \{1, \ldots, L\}$ with $j = \#\mathfrak{u}$ elements, we will always assume that the elements are sorted as $u_1 < u_2 < \cdots < u_j$. Then, we define the *combined operator*

$$\mathcal{I}_\mathfrak{u} := \mathcal{I}_{u_j}\mathcal{I}_{u_{j-1}} \cdots \mathcal{I}_{u_1} : V \to W_{u_j}.$$

It is important to note that the ordering of $\mathfrak{u}$ is crucial in the definition of $\mathcal{I}_\mathfrak{u}$ as the operators $\mathcal{I}_i$ usually do not commute. Moreover, we stress that in the definition of $\mathcal{I}_\mathfrak{u}$ we first apply the operator with the smallest index then the one with the next bigger index and so on.

Theorem 3.8 *Assume that the residual correction scheme computes local approximations via $\mathcal{I}_i : V \to W_i$. Then, the residual correction approximation $A_L : V \to V_L$ from (15) has the representation*

$$A_L = \sum_{\substack{\mathfrak{u} \subseteq \{1,\ldots,L\} \\ 1 \leq \#\mathfrak{u} \leq L}} (-1)^{\#\mathfrak{u}+1} \mathcal{I}_\mathfrak{u} = \sum_{i=1}^{L}(-1)^{i+1} \sum_{\substack{\mathfrak{u} \subseteq \{1,\ldots,L\} \\ \#\mathfrak{u}=i}} \mathcal{I}_\mathfrak{u}. \qquad (22)$$

Proof Let $I : V \rightarrow V$ denote the identity. The definition of the error yields the recursion $e_i = (I - \mathcal{I}_i)e_{i-1}$, showing

$$e_L = (I - \mathcal{I}_L)(I - \mathcal{I}_{L-1}) \cdots (I - \mathcal{I}_1)f.$$

For the resulting product of operators it is easy to see by induction, that we have the identity

$$(I - \mathcal{I}_L)(I - \mathcal{I}_{L-1}) \cdots (I - \mathcal{I}_1) = I + \sum_{\substack{\mathfrak{u} \subseteq \{1,\ldots,L\} \\ \mathfrak{u} \neq \emptyset}} (-1)^{\#\mathfrak{u}} \mathcal{I}_{\mathfrak{u}}.$$

Finally, this, together with the identity $f_L = f - e_L$ yields the stated representation.
□

After this general representation, we will now derive a specific representation for the multilevel approximation method for both cases, interpolation and penalised least-squares approximation. Hence, we are in the situation described in Definition 3.5, i.e. we consider operators $\mathcal{I}_i : C(\Omega) \rightarrow W_i$ defining the local approximation with

$$W_i := \text{span}\{K_i(\cdot, \mathbf{x}) \ \mathbf{x} \in X_i\}.$$

We first look at interpolation. A change of basis allows us to write the operators $\mathcal{I}_i$ also in explicit form. For each i and $1 \leq k \leq N_i$ let $\chi_{i,k} \in W_i$ be the *cardinal* or *Lagrange* function satisfying $\chi_{i,k}(\mathbf{x}_{i,\ell}) = \delta_{k,\ell}$, $1 \leq k, \ell \leq N_i$, where $\delta_{k,\ell}$ is the Kronecker symbol. Then, obviously, we have

$$\mathcal{I}_i f = \sum_{k=1}^{N_i} f(\mathbf{x}_{i,k}) \chi_{i,k}. \tag{23}$$

Inserting this formula into the representation (22) yields the following result.

Theorem 3.9 *Let $\mathcal{I}_i : C(\Omega) \rightarrow W_i$ be interpolation operators of the form (23). Let $\mathfrak{u} = \{u_1, \ldots, u_{\#\mathfrak{u}}\}$ be an ordered set. Then, using the notation $\mathbf{k} = (k_1, \ldots, k_{\#\mathfrak{u}})^{\mathrm{T}} \in \mathbb{N}^{\#\mathfrak{u}}$ and $\mathbf{N}_{\mathfrak{u}} = (N_{u_1}, \ldots, N_{u_{\#\mathfrak{u}}})^{\mathrm{T}} \in \mathbb{N}^{\#\mathfrak{u}}$, the combined operator $\mathcal{I}_{\mathfrak{u}} : C(\Omega) \rightarrow W_{\#\mathfrak{u}}$ has the representation*

$$\mathcal{I}_{\mathfrak{u}} f = \sum_{\mathbf{k} \leq \mathbf{N}_{\mathfrak{u}}} a(\mathfrak{u}, \mathbf{k}) f(\mathbf{x}_{u_1,k_1}) \chi_{u_{\#\mathfrak{u}},k_{\#\mathfrak{u}}},$$

where the coefficients are given by $a(\mathfrak{u}, \mathbf{k}) = 1$ if $\#\mathfrak{u} = 1$ and

$$a(\mathfrak{u}, \mathbf{k}) = \prod_{\ell=1}^{\#\mathfrak{u}-1} \chi_{u_\ell,k_\ell}(\mathbf{x}_{u_{\ell+1},k_{\ell+1}})$$

for $\#\mathfrak{u} > 1$.

Moreover, the multilevel interpolation operator $A_L : C(\Omega) \to V_L$ has the representation

$$A_L(f) = \sum_{\substack{\mathfrak{u} \subseteq \{1,\ldots,L\} \\ 1 \leq \#\mathfrak{u} \leq L}} (-1)^{\#\mathfrak{u}+1} \sum_{\mathbf{k} \leq \mathbf{N_u}} a(\mathfrak{u}, \mathbf{k}) f(\mathbf{x}_{u_1,k_1}) \chi_{u_{\#\mathfrak{u}},k_{\#\mathfrak{u}}}. \tag{24}$$

Proof For $\mathfrak{u} = \{u_1, \ldots, u_i\}$ we have

$$
\begin{aligned}
\mathcal{I}_{\mathfrak{u}} f &= \mathcal{I}_{u_i} \cdots \mathcal{I}_{u_1} f \\
&= \mathcal{I}_{u_i} \cdots \mathcal{I}_{u_2} \sum_{k_1=1}^{N_{u_1}} f(\mathbf{x}_{u_1,k_1}) \chi_{u_1,k_1} \\
&= \mathcal{I}_{u_i} \cdots \mathcal{I}_{u_3} \sum_{k_2=1}^{N_{u_2}} \sum_{k_1=1}^{N_{u_1}} \chi_{u_1,k_1}(\mathbf{x}_{u_2,k_2}) f(\mathbf{x}_{u_1,k_1}) \chi_{u_2,k_2} \\
&= \cdots \\
&= \sum_{k_i=1}^{N_{u_i}} \cdots \sum_{k_1=1}^{N_{u_1}} \chi_{u_{i-1},k_{i-1}}(\mathbf{x}_{u_i,k_i}) \cdots \chi_{u_1,k_1}(\mathbf{x}_{u_2,k_2}) f(\mathbf{x}_{u_1,k_1}) \chi_{u_i,k_i} \\
&= \sum_{\mathbf{k} \leq \mathbf{N_u}} \left[\prod_{\ell=1}^{i-1} \chi_{u_\ell,k_\ell}(\mathbf{x}_{u_{\ell+1},k_{\ell+1}}) \right] f(\mathbf{x}_{u_1,k_1}) \chi_{u_i,k_i}.
\end{aligned}
$$

Inserting this into (22) gives the representation for the multilevel approximation. $\square$

Next, we are dealing with penalised least-squares approximation, see also Lemma 3.2. To find a representation for the multilevel approximation operator like the one in Theorem 3.9, we have to replace the cardinal or Lagrange functions $\chi_{i,k}$ appropriately. To this end, we note that, in the case of interpolation, the coefficient vector of the interpolant (17) is given by $\boldsymbol{\alpha}_i = M_i^{-1}\mathbf{f}$ with the kernel matrix $M_i = (K_i(\mathbf{x}_{i,j}, \mathbf{x}_{i,k}))$. Hence, the interpolant can also be written as

$$s_i(\mathbf{x}) = \mathbf{r}_i(\mathbf{x})^{\mathrm{T}} \boldsymbol{\alpha}_i = \mathbf{r}_i(\mathbf{x})^{\mathrm{T}} M_i^{-1}\mathbf{f},$$

where $\mathbf{r}_i(\mathbf{x}) = (K_i(\mathbf{x}, \mathbf{x}_{i,1}), \ldots, K_i(\mathbf{x}, \mathbf{x}_{i,N_i}))^{\mathrm{T}} \in \mathbb{R}^{N_i}$. Thus, we can conclude that the cardinal or Lagrange function can alternatively be written as

$$\chi_{i,k}(\mathbf{x}) = \mathbf{r}_i(\mathbf{x})^{\mathrm{T}} M_i^{-1} \mathbf{e}_k, \qquad 1 \leq k \leq N_i,$$

where $\mathbf{e}_k \in \mathbb{R}^{N_i}$ is the k-the unit vector. This idea now carries over to the penalised least-squares setting. As in this case the coefficient vector in (17) is the solution of $(M_i + \lambda_i I)\boldsymbol{\alpha}_i = \mathbf{f}$, we can now write the approximant as

$$s_i(\mathbf{x}) = \mathbf{r}_i(\mathbf{x})^{\mathrm{T}} (M_i + \lambda_i I)^{-1}\mathbf{f} = \sum_{k=1}^{N_i} f(\mathbf{x}_{i,k}) \widetilde{\chi}_{i,k}(\mathbf{x}),$$

with the modified cardinal functions

$$\widetilde{\chi}_{i,k}(\mathbf{x}) = \mathbf{r}_i(\mathbf{x})^{\mathrm{T}}(M_i + \lambda_i I)^{-1}\mathbf{e}_k, \qquad 1 \leq k \leq N_i. \tag{25}$$

This gives together with the representation (22) the following result.

Corollary 3.10 *In the case of penalised least-squares operators $\mathcal{I}_i : C(\Omega) \to W_i$, the multilevel interpolation operator $A_L : C(\Omega) \to V_L$ has the representation (24) if the cardinal functions $\chi_{i,k}$ are replaced by the functions $\widetilde{\chi}_{i,k}$ from (25).*

4 The tensor product multilevel method

In this section, we will combine the kernel multilevel method of the previous section with Smolyak's formula. First, we will describe the resulting tensor-product multilevel method in more detail. In the second subsection we will then study the convergence of this method.

4.1 The method and its representations

The main idea of the method is to use the multilevel method of Sect. 3 in Smolyak's construction (9). We will discuss various representations of the resulting operator. We will start this by first looking at the residual correction scheme with general monotone index sets Λ and then become more and more specific by looking at the isotropic and anisotropic index sets $\Lambda(q, d)$ and $\Lambda_{\boldsymbol{\omega}}(\ell, d)$ from (4) and (5), respectively. After that, we prove representations if the kernel-based multilevel method is employed with interpolation and penalised least-squares approximation.

In the most general case, we assume that the operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$ for $1 \leq j \leq d$ and $1 \leq i \leq L^{(j)}$ between linear spaces $U^{(j)}$ and $V^{(j)}$ are now given by our multilevel approximation operators from Lemma 3.4, i.e. by

$$A_i^{(j)}(f^{(j)}) := \sum_{k=1}^{i} \mathcal{I}_k^{(j)}\left(e_{k-1}^{(j)}\right), \qquad f^{(j)} \in U^{(j)}. \tag{26}$$

In its original form, the Smolyak construction

$$\mathcal{A}_\Lambda := \sum_{\mathbf{i} \in \Lambda} \Delta_{i_1}^{(1)} \otimes \cdots \otimes \Delta_{i_d}^{(d)} \tag{27}$$

heavily relies on the difference operators (8) rather than the approximation operators. In our setting, the difference operators simplify to the local approximation operators, i.e.

$$\Delta_i^{(j)}\left(f^{(j)}\right) = A_i^{(j)}\left(f^{(j)}\right) - A_{i-1}^{(j)}\left(f^{(j)}\right) = \mathcal{I}_i^{(j)}\left(e_{i-1}^{(j)}\right), \qquad 1 \leq i \leq L^{(j)}. \tag{28}$$

This is particularly helpful when approximating elements from an algebraic tensor product space, as in this case all computations can be done independently for each direction.

Proposition 4.1 *Let $U^{(1)}, \ldots, U^{(d)}$ and $V^{(1)}, \ldots, V^{(d)}$ be linear spaces with algebraic tensor products $S = U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $\Lambda \subseteq \mathbb{N}^d$ be a monotone index set. For $1 \leq j \leq d$ let $L^{(j)} := \max\{\lambda_j : \lambda \in \Lambda\}$. Assume that the operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$, $1 \leq i \leq L^{(j)}$, are given by the residual correction operators in (26). Assume finally, $f \in S$ has the form*

$$f = \sum_{k=1}^{n} f_k^{(1)} \otimes \cdots \otimes f_k^{(d)}, \qquad f_i^{(j)} \in U^{(j)}.$$

For $1 \leq j \leq d$ let $e_{i,k}^{(j)}$, $0 \leq i \leq L^{(j)}$, $1 \leq k \leq n$, be the errors defined in the standard residual correction scheme applied to $f_k^{(j)}$. Then, the tensor product multilevel operator applied to f is given by

$$\mathcal{A}_\Lambda(f) = \sum_{k=1}^{n} \sum_{\mathbf{i} \in \Lambda} \mathcal{I}_{i_1}^{(i)}(e_{i_1-1,k}^{(j)}) \otimes \cdots \otimes \mathcal{I}_{i_d}^{(d)}(e_{i_d-1,k}^{(d)}).$$

However, if the $U^{(j)}$ are normed spaces and we want to extend the Smolyak operator to the tensor product space $\bigotimes U^{(j)}$, we need the combination technique, to find a representation of the tensor product multilevel method. To this end, we need to express the dependence on the multilevel operator on f explicitly. This now follows immediately if we employ the representation from Theorem 3.8 in each direction $1 \leq j \leq d$.

Proposition 4.2 *Let $U^{(1)}, \ldots, U^{(d)}$ and $V^{(1)}, \ldots, V^{(d)}$ be linear spaces with algebraic tensor products $S = U^{(1)} \otimes \cdots \otimes U^{(d)}$ and $T = V^{(1)} \otimes \cdots \otimes V^{(d)}$. Let $\Lambda \subseteq \mathbb{N}^d$ be a monotone index set. For $1 \leq j \leq d$ let $L^{(j)} := \max\{\lambda_j : \lambda \in \Lambda\}$. Assume that the operators $A_i^{(j)} : U^{(j)} \to V^{(j)}$, $1 \leq i \leq L^{(j)}$, are given by the residual correction operators in (26). Then, the Smolyak operator (27) has for a general index set Λ, the isotropic index set $\Lambda(q, d)$ and the anisotropic index set $\Lambda_\omega(\ell, d)$ the representations*

$$\mathcal{A}_\Lambda = \sum_{\substack{\mathbf{i} \in \Lambda, \boldsymbol{\beta} \in \{0,1\}^d \\ \mathbf{i}+\boldsymbol{\beta} \in \Lambda}} \sum_{\substack{u^{(1)} \subseteq \{1,\ldots,i_1\} \\ 1 \leq \#u^{(1)} \leq i_1}} \cdots \sum_{\substack{u^{(d)} \subseteq \{1,\ldots,i_d\} \\ 1 \leq \#u^{(d)} \leq i_d}} c_{\boldsymbol{\beta}}(u^{(1)}, \ldots, u^{(d)}) \mathcal{I}_{u^{(1)}}^{(1)} \otimes \cdots \otimes \mathcal{I}_{u^{(d)}}^{(d)},$$

$$\mathcal{A}_{\Lambda(q,d)} = \sum_{\mathbf{i} \in P(q,d)} \sum_{\substack{u^{(1)} \subseteq \{1,\ldots,i_1\} \\ 1 \leq \#u^{(1)} \leq i_1}} \cdots \sum_{\substack{u^{(d)} \subseteq \{1,\ldots,i_d\} \\ 1 \leq \#u^{(d)} \leq i_d}} b_{\mathbf{i}}(u^{(1)}, \ldots, u^{(d)}) \mathcal{I}_{u^{(1)}}^{(1)} \otimes \cdots \otimes \mathcal{I}_{u^{(d)}}^{(d)},$$

$$\mathcal{A}_{\Lambda_\omega(\ell,d)} = \sum_{\substack{\mathbf{i} \in Y_\omega(\ell,d), \boldsymbol{\beta} \in \{0,1\}^d \\ \mathbf{i}+\boldsymbol{\beta} \in \Lambda_\omega(\ell,d)}} \sum_{\substack{u^{(1)} \subseteq \{1,\ldots,i_1\} \\ 1 \leq \#u^{(1)} \leq i_1}} \cdots \sum_{\substack{u^{(d)} \subseteq \{1,\ldots,i_d\} \\ 1 \leq \#u^{(d)} \leq i_d}} c_{\boldsymbol{\beta}}(u^{(1)}, \ldots, u^{(d)}) \mathcal{I}_{u^{(1)}}^{(1)} \otimes \cdots \otimes \mathcal{I}_{u^{(d)}}^{(d)},$$

respectively, where $P(q,d) = \{\mathbf{i} \in \mathbb{N}^d : q - d + 1 \le |\mathbf{i}| \le q\}$, $Y_{\boldsymbol{\omega}}(\ell, d) = \Lambda_{\boldsymbol{\omega}}(\ell, d)\backslash\Lambda_{\boldsymbol{\omega}}(\ell - \|\boldsymbol{\omega}\|_1/\min\omega_j, d)$ *and*

$$c_{\boldsymbol{\beta}}(\mathfrak{u}^{(1)}, \ldots, \mathfrak{u}^{(d)}) = (-1)^{|\boldsymbol{\beta}|+d+\#\mathfrak{u}^{(1)}+\cdots+\#\mathfrak{u}^{(d)}},$$

$$b_{\mathbf{i}}(\mathfrak{u}^{(1)}, \ldots, \mathfrak{u}^{(d)}) = (-1)^{q+d-|\mathbf{i}|+\#\mathfrak{u}^{(1)}+\ldots+\#\mathfrak{u}^{(d)}}\binom{d-1}{q-|\mathbf{i}|}.$$

Proof Using (22) for each direction shows

$$A_{i_1}^{(1)} \otimes \cdots \otimes A_{i_d}^{(d)} = \sum_{\substack{\mathfrak{u}^{(1)}\subseteq\{1,\ldots,i_1\}\\ 1\le\#\mathfrak{u}^{(1)}\le i_1}} \cdots \sum_{\substack{\mathfrak{u}^{(d)}\subseteq\{1,\ldots,i_d\}\\ 1\le\#\mathfrak{u}^{(d)}\le i_d}} (-1)^{\#\mathfrak{u}_1+\cdots+\#\mathfrak{u}_d+d}\mathcal{I}_{\mathfrak{u}^{(1)}}^{(1)} \otimes \cdots \otimes \mathcal{I}_{\mathfrak{u}^{(d)}}^{(d)}.$$

Inserting this into (10) yields the first stated representation for a general index set Λ. If we insert this formula into (11), this gives the second representation for the index set $\Lambda(q,d)$ and inserting it into (12) yields the final representation for the anisotropic set $\Lambda_{\boldsymbol{\omega}}(\ell, d)$. $\qquad\square$

After these general results for the residual correction scheme, we now turn to the kernel-based multilevel approximation method. Here, we can further specify the approximation operators.

Consequently, we fix for each direction $1 \le j \le d$ a sequence of data sites

$$X_i^{(j)} = \{\mathbf{x}_{i,1}^{(j)}, \ldots, \mathbf{x}_{i,N_i^{(j)}}^{(j)}\} \subseteq \Omega^{(j)} \subseteq \mathbb{R}^{n_j}, \qquad 1 \le i \le L^{(j)},$$

where $L^{(j)} = \max\{\lambda_j : \boldsymbol{\lambda} \in \Lambda\}$. The corresponding mesh norms $h_i^{(j)}$ and the employed support radii $\delta_i^{(j)}$ are coupled, as explained above by

$$h_{i+1}^{(j)} = \mu^{(j)}h_i^{(j)}, \qquad 1 \le i \le L^{(j)} - 1, \quad 1 \le j \le d,$$

$$\delta_i^{(j)} = \nu^{(j)}h_i^{(j)}, \qquad 1 \le i \le L^{(j)}, \quad 1 \le j \le d,$$

with fixed parameters $\mu^{(j)} \in (0,1)$ and $\nu^{(j)} > 1$.

Next, for each direction, we choose a smoothness parameter $\sigma_j > n_j/2$ and a function $\Phi^{(j)} : \mathbb{R}^{n_j} \to \mathbb{R}$ satisfying (13) with $\sigma = \sigma_j$. The kernels $K_i^{(j)} : \mathbb{R}^{n_j} \times \mathbb{R}^{n_j} \to \mathbb{R}$ defined by $K_i^{(j)}(\mathbf{x}^{(j)}, \mathbf{y}^{(j)}) = \delta_i^{-n_j}\Phi^{(j)}((\mathbf{x}^{(j)}-\mathbf{y}^{(j)})/\delta_i)$ yield the direction-dependent approximation spaces

$$W_i^{(j)} = \text{span}\left\{K_i^{(j)}(\cdot, \mathbf{x}^{(j)}) : \mathbf{x}^{(j)} \in X_i^{(j)}\right\}, \qquad V_i^{(j)} = W_1^{(j)} + \cdots + W_i^{(j)}.$$

Finally, for each $1 \le j \le d$ we have the local approximation operators

$$\mathcal{I}_i^{(j)}f^{(j)} = \sum_{k=1}^{N_i^{(j)}} \alpha_{i,k}^{(j)}K_i^{(j)}(\cdot, \mathbf{x}_{i,k}^{(j)}), \qquad f^{(j)} \in C(\Omega^{(j)}),$$

where the coefficients $\boldsymbol{\alpha}_i^{(j)} \in \mathbb{R}^{N_i^{(j)}}$ are either determined by interpolation or penalised least-squares, though other methods are obviously possible as well. This finally yields the also direction-dependent multilevel approximation operators $A_i^{(j)} : C(\Omega^{(j)}) \to V_i^{(j)}$ defined by

$$A_i^{(j)}(f^{(j)}) = \sum_{k=1}^{i} \mathcal{I}_k^{(j)}(e_{k-1}^{(j)}), \qquad f^{(j)} \in C(\Omega^{(j)}),$$

where $e_k^{(j)}$ are the error functions defined in the multilevel process, depending now also on the direction $1 \leq j \leq d$.

For interpolation, we can rewrite the interpolation operator as

$$\mathcal{I}_i^{(j)} f^{(j)} = \sum_{k=1}^{N_i^{(j)}} f^{(j)}(\mathbf{x}_{i,k}^{(j)}) \chi_{i,k}^{(j)},$$

using now also direction-dependent Lagrange functions $\chi_{i,k}^{(j)} \in W_i^{(j)}$. For penalised least-squares approximation, we have the same representation after replacing $\chi_{i,k}^{(j)}$ with the modified functions $\widetilde{\chi}_{i,k}^{(j)} \in W_i^{(j)}$. Hence, we will mainly concentrate on the interpolation case. To derive the representation of the Smolyak operator in this situation we can employ Proposition 4.2. Hence, we only need to derive representations for the combined operators $\mathcal{I}_{\mathfrak{u}^{(1)}}^{(1)} \otimes \cdots \otimes \mathcal{I}_{\mathfrak{u}^{(d)}}^{(d)}$ appearing in that proposition as functions of the target function. This representation, however, follows with Theorem 3.9. We have

$$\mathcal{I}_{\mathfrak{u}^{(j)}} f^{(j)} = \sum_{\mathbf{k}^{(j)} \leq \mathbf{N}_{\mathfrak{u}^{(j)}}} a^{(j)}\left(\mathfrak{u}^{(j)}, \mathbf{k}^{(j)}\right) f^{(j)}\left(\mathbf{x}_{u_1^{(j)}, k_1^{(j)}}^{(j)}\right) \chi_{u_{\#\mathfrak{u}^{(j)}}^{(j)}, k_{\#\mathfrak{u}^{(j)}}^{(j)}}^{(j)}.$$

This then gives the following result, which, together with Proposition 4.2 yields the combination technique for the multilevel approximation operator.

Proposition 4.3 *Let $\Omega^{(j)} \subseteq \mathbb{R}^{n_j}$, $1 \leq j \leq d$, and $\Omega = \Omega^{(1)} \times \cdots \times \Omega^{(d)}$. Let $\mathcal{I}_i^{(j)} : C(\Omega^{(j)}) \to V_i^{(j)}$ be the interpolation operators previously defined. Then, for any $f \in C(\Omega)$, and sets $\mathfrak{u}^{(1)}, \ldots, \mathfrak{u}^{(d)}$, the combined operator has the form*

$$\mathcal{I}_{\mathfrak{u}^{(1)}}^{(1)} \otimes \cdots \otimes \mathcal{I}_{\mathfrak{u}^{(d)}}^{(d)}(f)$$

$$= \sum_{\mathbf{k}^{(1)} \leq \mathbf{N}_{\mathfrak{u}^{(1)}}} \cdots \sum_{\mathbf{k}^{(d)} \leq \mathbf{N}_{\mathfrak{u}^{(d)}}} \prod_{j=1}^{d} a^{(j)}(\mathfrak{u}^{(j)}, \mathbf{k}^{(j)}) f\left(\mathbf{x}_{u_1^{(1)}, k_1^{(1)}}^{(1)}, \ldots, \mathbf{x}_{u_1^{(d)}, k_1^{(d)}}^{(d)}\right) \cdot$$

$$\cdot \chi_{u_{\#\mathfrak{u}^{(1)}}^{(1)}, k_{\#\mathfrak{u}^{(1)}}^{(1)}} \otimes \cdots \otimes \chi_{u_{\#\mathfrak{u}^{(d)}}^{(d)}, k_{\#\mathfrak{u}^{(d)}}^{(d)}}.$$

4.2 Convergence results

We will now analyse the error of our tensor product multilevel approximation method. We will do this first for the isotropic case, i.e. for index sets of the form $\Lambda = \Lambda(q, d)$ and then for the anisotropic case, i.e. for index sets $\Lambda = \Lambda_{\boldsymbol{\omega}}(\ell, d)$.

For both cases it is helpful to summarise the general set-up, though we will simplify it significantly for the isotropic case.

As outlined above, we consider function spaces over domains $\Omega^{(j)} \subseteq \mathbb{R}^{n_j}$. To be more precise, we will use $U^{(j)} = H^{\beta_j}\left(\Omega^{(j)}\right)$ with $\beta_j > n_j/2$ and $V^{(j)} = L_2(\Omega^{(j)})$. Then, with $\Omega = \Omega^{(1)} \times \cdots \times \Omega^{(j)}$, the associated tensor product spaces are,

$$\bigotimes_{j=1}^{d} H^{\beta_j}(\Omega^{(j)}) = H^{\boldsymbol{\beta}}_{\text{mix}}(\Omega), \quad \bigotimes_{j=1}^{d} L_2(\Omega^{(j)}) = L_2(\Omega).$$

This follows for the L_2 case and for the case that $\boldsymbol{\beta} \in \mathbb{N}_0^d$ from Lemma 2.10. For general $\boldsymbol{\beta} \in \mathbb{R}_+^d$ this can be derived in the same way if the standard Sobolev spaces $H^{\beta_j}(\Omega^{(j)})$ are defined, for example, by interpolation.

Note that the condition $\beta_j > n_j/2$ ensures $U^{(j)} \subseteq C(\Omega^{(j)})$ by the Sobolev embedding theorem such that all our operators are well defined.

Next, we need bounds on all involved operators.

Lemma 4.4 *Assume that the local approximations $A_i^{(j)} : U^{(j)} \to V^{(j)}$ satisfy the conditions of Proposition 3.6. For a uniform notation we will simply set $\beta_j = \sigma_j$ for the first and third case of Proposition 3.6.*

Let $E^{(j)} : U^{(j)} \to V^{(j)}$ be the embedding operator. Then, for $1 \le j \le d$, the following bounds on the norm hold

$$\|E^{(j)}\|_{H^{\beta_j}(\Omega^{(j)}) \to L_2(\Omega^{(j)})} = 1,$$

$$\|E^{(j)} - A_i^{(j)}\|_{H^{\beta_j}(\Omega^{(j)}) \to L_2(\Omega^{(j)})} \le C^{(j)} \left[C_1^{(j)}(\mu^{(j)})^{\beta_j} \right]^i =: C^{(j)}\epsilon_j^i, \quad 0 \le i \le L^{(j)},$$

$$\|\Delta_i^{(j)}\|_{H^{\beta_j}(\Omega^{(j)}) \to L_2(\Omega^{(j)})} \le C^{(j)} \left[\epsilon_j^i + \epsilon_j^{i-1} \right] = C^{(j)} \left(1 + \frac{1}{\epsilon_j} \right) \epsilon_j^i$$

$$\le 2C^{(j)}\epsilon_j^{i-1}, \quad 1 \le i \le L^{(j)}.$$

Proof The bound on the embedding $E^{(j)} : H^{\beta_j}\left(\Omega^{(j)}\right) \to L_2\left(\Omega^{(j)}\right)$ is obviously true. The second bound follows directly from Proposition 3.6. The final bound follows because of

$$\Delta_i^{(j)} = A_i^{(j)} - A_{i-1}^{(j)} = A_i^{(j)} - E^{(j)} + E^{(j)} - A_{i-1}^{(j)}$$

immediately from the second one.   □

4.2.1 The isotropic case

in the isotropic case we use the index set $\Lambda = \Lambda(q, d) = \{\mathbf{i} \in \mathbb{N}^d : |\mathbf{i}| \le q\}$. As the index set does not distinguish between the directions, it is reasonable and

custom to assume this also for everything else. Hence, in this situation, we have $\Omega^{(j)} = \Omega^{(1)} \subseteq \mathbb{R}^n$ and $\beta_j = \beta$ for $1 \leq j \leq d$. Thus, we have $U^{(j)} = H^{\beta}(\Omega^{(1)})$ and $V^{(j)} = L_2(\Omega^{(1)})$.

In the multilevel method we also choose the point sets also direction-independent, i.e. $X_i^{(j)} = X_i = \{\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,N_i}\} \subseteq \Omega^{(1)}$. This means that the fill distances $h_i^{(j)} = h_i$ are independent of j and hence we choose the support radii $\delta_i^{(j)} = \delta_i$ and the parameters $\mu^{(j)} = \mu$ and $\nu^{(j)} = \nu$ also direction-independent.

In this simplified situation Lemma 4.4 simplifies as follows.

Lemma 4.5 *In the isotropic case, for $1 \leq j \leq d$, the following bounds on the norm hold*

$$\|E^{(1)}\|_{H^{\beta}(\Omega^{(1)}) \to L_2(\Omega^{(1)})} = 1,$$

$$\|E^{(1)} - A_i\|_{H^{\beta}(\Omega^{(1)}) \to L_2(\Omega^{(1)})} \leq C \left[C_1 \mu_j^{\beta} \right]^i =: C\epsilon^i, \qquad 0 \leq i \leq L$$

$$\|\Delta_i\|_{H^{\beta}(\Omega^{(1)}) \to L_2(\Omega^{(1)})} \leq C \left(1 + \frac{1}{\epsilon} \right) \epsilon^i, \qquad 1 \leq i \leq L.$$

Using the fact that all involved norms are reasonable and compatible crossnorms, it is now possible to derive a bound on the error of the tensor product multilevel approximation, i.e. on $\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)}$ using Lemma 2.18. However, this has already been in done in a general context in [33, Lemma 2] which is as follows.

Lemma 4.6 *In the situation of Lemma 2.18, assume that there are constants B, c, $\widetilde{c}$ and D such that the bounds*

$$\|E^{(1)}\|_{U_i \to V_i} \leq B$$
$$\|E^{(1)} - A_i\|_{U_i \to V_i} \leq cD^i, \qquad i \geq 0,$$
$$\|\Delta_i\|_{U_1 \to V_i} \leq \widetilde{c}D^i, \qquad i \geq 1,$$

hold. Then, the error for the Smolyak operator can be bounded by

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)}\|_{S \to T} \leq cB^{d-1}D^{q-d+1} \sum_{j=0}^{d-1} \left(\frac{\widetilde{c}D}{B} \right)^j \left(\begin{array}{c} 1 - d + j \\ j \end{array} \right)$$

$$\leq cH^{d-1} \left(\begin{array}{c} q \\ d-1 \end{array} \right) D^q,$$

with $H = \max\{B/D, \widetilde{c}\}$.

Combining Lemmas 4.5 and 4.6 gives our first main error estimate.

Theorem 4.7 *Let $\Omega^{(1)} \subseteq \mathbb{R}^n$ be a bounded domain with Lipschitz boundary and let $\sigma \geq \beta > n/2$. Let $\Omega = \Omega^{(1)} \times \cdots \times \Omega^{(1)}$. Assume that the low-dimensional reconstructions $A_i : C(\Omega^{(1)}) \to V_i$ are given by the multilevel RBF interpolants (15). Assume further that these multilevel RBF interpolants are built using compactly*

supported reproducing kernels of $H^\sigma(\mathbb{R}^n)$ and support radii and data sets as outlined in Proposition 3.6. *Finally, assume that the parameter $\mu \in (0, 1)$ is chosen in such a way that $\epsilon := C_1 \mu^\beta < 1$. Then, the error $\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)} : H^\beta_{mix}(\Omega) \to L_2(\Omega)$ can be bounded by*

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)}\|_{H^\beta_{mix}(\Omega) \to L_2(\Omega)} \le 2^{d-1} C^d \binom{q}{d-1} \epsilon^{q-d+1}, \tag{29}$$

where C is the constant from Lemma 4.5.

Proof From Lemma 4.5 we know that the constants in Lemma 4.6 are given by $B = 1$, $c = C$, $\widetilde{c} = C(1 + 1/\epsilon)$ and $D = \epsilon$. This leads for $C \ge 1$ to

$$H = \max\{B/D, \widetilde{C}\} = \max\left\{1/\epsilon, C(1 + \frac{1}{\epsilon})\right\} = C\left(1 + \frac{1}{\epsilon}\right) \le 2C/\epsilon.$$

Hence, Lemma 4.6 yields

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)}\|_{H^\beta_{mix}(\Omega) \to L_2(\Omega)} \le C H^{d-1} D^q \binom{q}{d-1} = 2^{d-1} C^d \epsilon^{q-d+1} \binom{q}{d-1},$$

which is the stated estimate. □

As we assume $q \ge d$, estimate (29) shows also convergence for $\epsilon \to 0$. However, this is not desirable as we have $\epsilon = C_1 \mu^\beta$, where μ is the refinement factor of the mesh norms of the involved grids, i.e. it is a fixed number in $(0, 1)$ and hence so is ϵ. A smaller and smaller μ would lead to a too expensive refinement of the data sets.

The involved constant, unfortunately but also expectedly, depends exponentially on the space dimension d. Using the bound $\binom{q}{d-1} \le \frac{q^{d-1}}{(d-1)!}$, the error bound becomes

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(q,d)}\|_{H^\beta_{mix}(\Omega) \to L_2(\Omega)} \le c_d q^{d-1} \epsilon^{q-d+1}, \tag{30}$$

with $c_d = \frac{2^{d-1} C^d}{(d-1)!}$, which converges for $q \to \infty$.

4.2.2 The anisotropic case

We will now discuss the more general case of the anisotropic index $\Lambda_{\boldsymbol\omega}(\ell, d)$ from (5). The purpose here is to analyse the approximate reconstructions of functions $f \in H^\beta_{mix}(\Omega)$ with a possibly different smoothness β_j in each direction $1 \le j \le d$. This, of course, is reflected by choosing direction-dependent functions $\Phi^{(j)}$ leading to reproducing kernels of $H^{\sigma_j}(\Omega^{(j)})$ as outlined at the beginning of this section.

In the rest of this subsection, we will, without restriction, assume that the elements of the weight vector $\boldsymbol\omega$ are ordered as $\omega_1 \le \omega_2 \le \cdots \le \omega_d$ such that we in particular have $\min_j \omega_j = \omega_1$.

Though the representation in Lemma 2.19 was used in [8] to derive an error analysis for univariate operators $A_i^{(j)}$ given by polynomial interpolation in Chebyshev points, a general bound like the one in Lemma 4.6 has not been provided. Hence, we will use Lemma 2.19 directly. Accordingly, we have the error representation

$$\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)} = \sum_{k=1}^{d-1} \left(R(\ell,k) \otimes \bigotimes_{j=k+2}^{d} E^{(j)} \right) + \left(E^{(1)} - A_{\ell+1}^{(1)} \right) \otimes \bigotimes_{j=2}^{d} E^{(j)}$$

(31)

with

$$R(\ell,k) := \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \bigotimes_{j=1}^{k} \Delta_{i_j}^{(j)} \otimes \left(E^{(k+1)} - A_{\widehat{i}_{k+1}-1}^{(k+1)} \right)$$

and, for $\mathbf{i} \in \Lambda_\omega(\ell,k)$,

$$\widehat{i}_{k+1} = \left\lfloor 2 + \ell\frac{\omega_1}{\omega_{k+1}} - \sum_{j=1}^{k}(i_j-1)\frac{\omega_j}{\omega_{k+1}} \right\rfloor.$$

We will now use this representation to bound the error in the situation that the low-dimensional interpolation operators are given by the multilevel RBF interpolant. Using the fact that all involved norms are reasonable and compatible crossnorms we can use the estimates from Lemma 4.4, to bound the single terms in (31). To simplify the notation, we will from now on suppress the index at the occuring norms. Essentially, we are only dealing with operator norms $\|\cdot\|_{H^{\beta_j}(\Omega^{(j)})\to L_2(\Omega^{(j)})}$ or $\|\cdot\|_{H^\beta_{\mathrm{mix}}(\Omega)\to L_2(\Omega)}$ and it should be clear from the context, which norm is meant. Defining $C := \max_{1\le j\le d} C^{(j)}$ we have

$$\left\| \left(E^{(1)} - A_{\ell+1}^{(1)} \right) \otimes \bigotimes_{j=2}^{d} E^{(j)} \right\| = \| E^{(1)} - A_{\ell+1}^{(1)} \| \le C\epsilon_1^{\ell+1}$$

and for $1 \le k \le d-1$ also

$$\left\| R(\ell,k) \otimes \bigotimes_{j=k+2}^{d} E^{(j)} \right\| = \| R(\ell,k) \| \le \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \prod_{j=1}^{k} \|\Delta_{i_j}^{(j)}\| \| E^{(k+1)} - A_{\widehat{i}_{k+1}-1}^{(k+1)} \|$$

$$\le \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \prod_{j=1}^{k} \left[2C\epsilon_j^{i_j-1} \right] C\epsilon_{k+1}^{\widehat{i}_{k+1}-1} = 2^k C^{k+1} \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \epsilon_{k+1}^{\widehat{i}_{k+1}-1} \prod_{j=1}^{k} \epsilon_j^{i_j-1}.$$

This gives the first bound

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)}\| \leq \sum_{k=1}^{d-1} \left\| R(\ell,k) \otimes \bigotimes_{j=k+2}^{d} E^{(j)} \right\| + \left\| \left(E^{(1)} - A_{\ell+1}^{(1)} \right) \otimes \bigotimes_{j=2}^{d} E^{(j)} \right\|$$

$$\leq C\epsilon_1^{\ell+1} + \sum_{k=1}^{d-1} \left[2^k C^{k+1} \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \epsilon_{k+1}^{\widehat{i}_{k+1}-1} \prod_{j=1}^{k} \epsilon_j^{i_j-1} \right]. \qquad (32)$$

To understand this estimate, we need to look at the coefficient $\widehat{i}_{k+1} - 1$ and, as $\epsilon_{k+1} \in (0,1)$, we need a lower bound. This can, for example, be reached by noticing that $\omega_1 \leq \omega_k$ for all $2 \leq k \leq d$ and hence

$$\widehat{i}_{k+1} - 1 = \left\lfloor 2 + \ell \frac{\omega_1}{\omega_{k+1}} - \sum_{j=1}^{k}(i_j-1)\frac{\omega_j}{\omega_{k+1}} \right\rfloor - 1$$

$$\geq \left\lfloor \ell \frac{\omega_1}{\omega_{k+1}} \right\rfloor + 1 - \sum_{j=1}^{k}(i_j-1).$$

With this, (32) becomes

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)}\| \leq C\epsilon_1^{\ell+1} + \sum_{k=1}^{d-1} \left[2^k C^{k+1} \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \epsilon_{k+1}^{\left\lfloor \ell\frac{\omega_1}{\omega_{k+1}} \right\rfloor + 1 - \sum_{j=1}^{k}(i_j-1)} \prod_{j=1}^{k} \epsilon_j^{i_j-1} \right]$$

$$= C\epsilon_1^{\ell+1} + \sum_{k=1}^{d-1} \left[2^k C^{k+1} \sum_{\mathbf{i}\in\Lambda_\omega(\ell,k)} \epsilon_{k+1}^{\left\lfloor \ell\frac{\omega_1}{\omega_{k+1}} \right\rfloor + 1} \prod_{j=1}^{k} \left(\frac{\epsilon_j}{\epsilon_{k+1}}\right)^{i_j-1} \right].$$

To simplify this further we will make the assumption that we choose the local refinement factors $\mu^{(j)}$ in such a way that $\epsilon_j = \epsilon$ for all $1 \leq j \leq d$. Such a choice seems natural as it reflects also the different smoothness β_j in each direction. The larger β_j is, the larger we can also choose $\mu^{(j)} \in (0,1)$. Under the assumption $\omega_1 \leq \omega_2 \leq \cdots \leq \omega_d$ we can use (7) to bound the number of indices in $\Lambda_\omega(\ell,k)$ as

$$\#\Lambda_\omega(\ell,k) \leq \prod_{j=1}^{k}\left(\frac{\ell\omega_1}{j\omega_j}+1\right) \leq \prod_{j=1}^{k}\left(\frac{\ell}{j}+1\right) = \binom{\ell+k}{\ell} \leq \frac{(\ell+k)^k}{k!}.$$

With this, our error bound becomes

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)}\| \leq C\epsilon^{\ell+1} + \sum_{k=1}^{d-1} 2^k C^{k+1} \epsilon^{\left\lfloor \ell\frac{\omega_1}{\omega_{k+1}} \right\rfloor + 1} \#\Lambda_\omega(\ell,k)$$

$$\leq C\epsilon^{\ell+1} + \epsilon^{\left\lfloor \ell\frac{\omega_1}{\omega_d} \right\rfloor + 1} \sum_{k=1}^{d-1} 2^k C^{k+1} \frac{(\ell+k)^k}{k!},$$

where we also have used $\omega_{k+1} \leq \omega_d$ and hence $\omega_1/\omega_{k+1} \geq \omega_1/\omega_d$. The final sum can further be bounded by

$$\sum_{k=1}^{d-1} 2^k C^{k+1} \frac{(\ell+k)^k}{k!} \leq (\ell+d-1)^{d-1} \sum_{k=1}^{d-1} \frac{2^k C^{k+1}}{k!} \leq C(\ell+d-1)^{d-1} e^{2C},$$

Taking this all together yields the following error bound.

Theorem 4.8 *For $1 \leq j \leq d$ let $\Omega^{(j)} \subseteq \mathbb{R}^{n_j}$ be bounded domains with Lipschitz boundary and let $\sigma_j \geq \beta_j > n_j/2$. Set $\Omega = \Omega^{(1)} \times \cdots \times \Omega^{(d)}$. Let $\omega_1 \leq \cdots \leq \omega_d$. Assume that the low-dimensional reconstructions $A_i^{(j)} : C(\Omega^{(j)}) \to V_i^{(j)}$ are given by the multilevel RBF interpolants* (15). *Assume further that these multilevel RBF interpolants are built using compactly supported reproducing kernels of $H^{\sigma_j}(\mathbb{R}^{n_j})$ and support radii $\delta_i^{(j)}$ and data sets $X_i^{(j)}$ as outlined in Proposition* 3.6. *Finally, assume that the factors $\mu^{(j)}$ are chosen so that all $\epsilon_j := C_1^{(j)}[\mu^{(j)}]^{\beta_j}$ satisfy $\epsilon_j = \epsilon \in (0,1)$. Then, the interpolation error $\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)} : H_{mix}^{\beta}(\Omega) \to L_2(\Omega)$ satisfies the bound*

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)}\|_{H_{mix}^{\beta}(\Omega) \to L_2(\Omega)} \leq c(\ell+d-1)^{d-1} \epsilon^{\left\lfloor \ell \frac{\omega_1}{\omega_d} \right\rfloor + 1} \tag{33}$$

with $c := C\left(1 + e^{2C}\right)$.

For $\ell \to \infty$, the asymptotic behaviour of the error bound is given by

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)}\|_{H_{mix}^{\beta}(\Omega) \to L_2(\Omega)} \leq c_d \ell^{d-1} \epsilon^{\lfloor \ell \frac{\omega_1}{\omega_d} \rfloor + 1}$$

with c_d from (30). Hence, we have again exponential convergence for $\ell \to \infty$. Moreover, in the case of a constant weight vector $\boldsymbol{\omega} = (\omega, \ldots, \omega)^{\mathrm{T}} \neq \mathbf{0}$ we have $\Lambda_\omega(\ell,d) = \Lambda(q,d)$ with $q = \ell + d$. In this case, the error bound (33) becomes

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_\omega(\ell,d)}\|_{H_{mix}^{\beta}(\Omega) \to L_2(\Omega)} \leq c(q-1)^{d-1} \epsilon^{q-d+1},$$

which recovers exactly the same behaviour that we have seen in (30).

A closer look at the proof above shows that we can rewrite one of the intermediate steps of the error as follows

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(\ell,d)}\| \leq \sum_{k=1}^{d} 2^{k-1} C^k \sum_{\mathbf{i} \in \Lambda_\omega(\ell,k)} \epsilon_k^{L^{(k)}} \prod_{j=1}^{k} \left(\frac{\epsilon_j}{\epsilon_k}\right)^{i_j-1},$$

using $L^{(k)} = \max\{\lambda_k : \boldsymbol{\lambda} \in \Lambda_\omega(\ell,d)\} = \lfloor \ell \frac{\omega_1}{\omega_k} \rfloor + 1$. Obviously, from our assumption $\omega_1 \leq \omega_2 \leq \cdots \leq \omega_d$ it follows that the maximum number of levels in each direction is decreasing, i.e. $L^{(1)} \geq L^{(2)} \geq \cdots \geq L^{(d)}$.

Unfortunately, in the above formula we have two competing terms. We want each $\epsilon_k^{L^{(k)}}$ to become small. If we want to balance them, it seems natural to assume that

$\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_k$, i.e. we can choose a larger ϵ_j for smaller indices as the corresponding exponents $L^{(j)}$ are also larger. However, this collides with the other term $\prod_{j=1}^k (\epsilon_j/\epsilon_k)^{i_j-1}$ as these products then become larger than one. We have resolved this issue by assuming that all ϵ_j are equal. This allowed us to neglect the damaging products but leads to the bound

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda(\ell,d)}\| \leq \sum_{k=1}^{d} 2^{k-1} C^k \binom{\ell+k-1}{\ell} \epsilon^{L^{(k)}},$$

which is dominated by the worst term $\epsilon^{L^{(d)}}$.

Remark 4.9 So far, we have not coupled the smoothness $\boldsymbol{\beta}$ of the target function with the weight vector $\boldsymbol{\omega}$ of the index set. However, our arguments above suggest that we should choose $\boldsymbol{\omega} = \boldsymbol{\beta}$. This leads to all the desired properties. Because of $L^{(1)} \geq L^{(2)} \geq \cdots \geq L^{(d)}$ we use more levels in the direction of the least smoothness, as expected. Because of $\epsilon_j = C^{(j)}[\mu^{(j)}]^{\beta_j} = \epsilon$ we also have $\mu^{(1)} \leq \mu^{(2)} \leq \cdots \leq \mu^{(d)}$. As the number of points $N_i^{(j)}$ is proportional to $[\mu^{(j)}]^{-i/n_j}$, this means that we also have more points in those directions of lower smoothness. The error can then be expressed using the smoothness of the target function via

$$\|\mathcal{E}^{(d)} - \mathcal{A}_{\Lambda_{\boldsymbol{\beta}}(\ell,d)}\|_{H^{\beta}_{\text{mix}}(\Omega) \to L_2(\Omega)} \leq c_d \ell^{d-1} \epsilon^{\lfloor \ell \frac{\beta_1}{\beta_d} \rfloor + 1}.$$

5 Numerical tests

We will now provide numerical examples to demonstrate that the new tensor product multilevel method works as theoretically predicted and to validate the error bounds we derived in the previous section.

The details of the algorithmic realisation of the method are mostly already described in Sect. 4. However, we recall the important parts here. Once we fixed the index set $\Lambda_{\boldsymbol{\omega}}(\ell, d)$, where the weight vector $\boldsymbol{\omega} \in \mathbb{R}^d_+$ and threshold $\ell \in \mathbb{N}$ serve as input parameters for the algorithm, we determine the low-dimensional point sets $X_i^{(j)}$ with refinement parameter μ or $\mu^{(j)}$ such that the conditions of Theorems 4.7 and 4.8, respectively, are satisfied. After choosing a reproducing kernel $K^{(j)}$ for every direction, we compute the Lagrange functions $\chi_{i,k}^{(j)}$ for every level $1 \leq i \leq L^{(j)}$ and every anchor $\mathbf{x}_k \in X_i^{(j)}$, $1 \leq k \leq N_i^{(j)}$. All this can be done in an offline phase as soon as the input parameters are determined. The interpolants then are computed following the formulas of Propositions 4.2 and 4.3.

We test two different settings: First, an isotropic, high-dimensional example and second a fully anisotropic one. In any case we use the compactly supported Wendland kernels
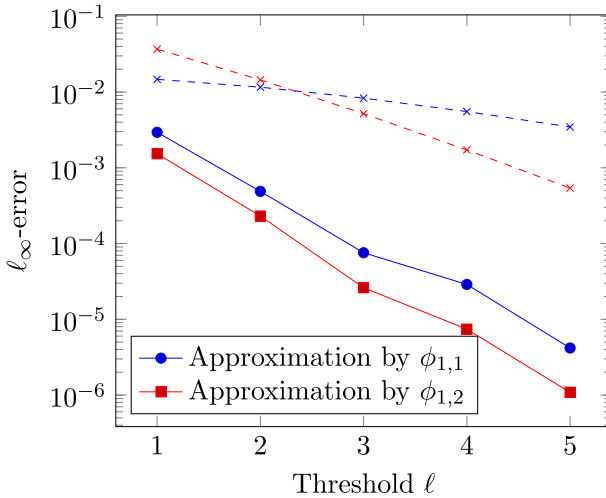
$$\phi_{1,1}(r) = (1-r)^3_+ (3r+1),$$

Fig. 1 Error plot for the isotropic case: $\ell_\infty([0, 1]^7)$-error evaluated on 50 randomly drawn points. The dashed lines are the theoretical error bounds $(\ell + 6)^6 \varepsilon^{\ell+1}$ with $\varepsilon = C_1 \mu^{\beta-0.5}$

$$\phi_{1,2}(r) = (1 - r)_+^5 (8r^2 + 5r + 1),$$
$$\phi_{1,3}(r) = (1 - r)_+^7 (21r^3 + 19r^2 + 7r + 1),$$

which are the reproducing kernels of $H^2(\mathbb{R})$, $H^3(\mathbb{R})$ and $H^4(\mathbb{R})$, respectively.

5.1 Seven-dimensional isotropic example

For the isotropic example we follow the ideas of [3] and use the *oscillatory* function $f : [0, 1]^7 \to \mathbb{R}$ with the choice $w_1 = 1$, i.e.,

$$f(\mathbf{x}) = \cos\left(\sum_{j=1}^{7} c_j x^{(j)}\right),$$

as the target function. The components of the *difficulty parameter* $\mathbf{c}$ are chosen randomly in $[0, 1]$ and then renormalised such that $\sum_{j=1}^{7} c_j = 9.0$. Obviously, $f \in C^\infty([0, 1]^7)$ and thus $f \in H_{\text{mix}}^{\beta}((0, 1)^7)$ for every $\boldsymbol{\beta} \in \mathbb{R}^7$ with $\beta_j > 1/2$, $1 \le j \le 7$.

As direction-wise point sets $X_i^{(j)} \subseteq [0, 1]$ we use uniformly distributed points with $\#X_i^{(j)} = 2^{i+2}$, i.e. we use a uniform refinement parameter $\mu = \mu^{(j)} = 0.5$. The overlap parameter, which dictates how the support radius δ_i is coupled to the fill distance h_i, is chosen as $\nu = \nu^{(j)} = 8.0$.

In Fig. 1 we show an ℓ_∞-error plot for the interpolation with the first two kernels given above. We follow again [3] and compute the error for both examples on the same 50 randomly drawn points. Though our theory is based on L_2-estimates, it can

easily be modified for L_∞. Additionally, we give the theoretical bound $(\ell + 6)^6 \varepsilon^{\ell+1}$ as the dashed lines. We use $\varepsilon = (C_1\mu)^{2-0.5}$, for the case $\phi_{1,1}$ and $\varepsilon = (C_1\mu)^{3-0.5}$ if we use $\phi_{1,2}$. In both cases we use $C_1 = 1$. As expected, approximating with a smoother kernel leads to smaller errors and a faster convergence. We see that in the case of the less smooth kernel $\phi_{1,1}$ the method converges faster than the theory predicts. Approximating with $\phi_{1,2}$ yields a convergence order similar to the one predicted by the theory.

5.2 Two-dimensional anisotropic example

Looking at (33) in Theorem 4.8, we see that the convergence order depends only on the weights of the first and last direction, ω_1 and ω_d, respectively. Hence, it suffices to test a two-dimensional problem.

This time, we follow the ideas of [44] and use the function $f_{\boldsymbol{\alpha}} : [-1, 1]^2 \to \mathbb{R}$ defined by

$$f_{\boldsymbol{\alpha}}(\mathbf{x}) := \prod_{j=1}^{2} f^{(j)}(x_j) = \prod_{j=1}^{2} |x_j|^{\alpha_j}$$

with $\boldsymbol{\alpha} = (1.6, 3.6)^{\mathrm{T}}$. We see that $f^{(1)} \in H^2((-1, 1))$ but not in $H^3((-1, 1))$ and $f^{(2)} \in H^4((-1, 1))$ but $f^{(2)} \notin H^5((-1, 1))$. This means that $f_{\boldsymbol{\alpha}}$ is an element of $H^{\boldsymbol{\beta}}_{\mathrm{mix}}(\Omega) = H^{(2,4)}_{\mathrm{mix}}(\Omega)$, but not in any Sobolev space of mixed, higher-order regularity. Furthermore, we now use in the first direction the kernel $\phi_{1,1}$ and in the second direction $\phi_{1,3}$. This way, the $f^{(j)}$ are elements of the native spaces of the respective kernels.

As point sets $X_i^{(j)} \subseteq [-1, 1]$ we use again uniformly distributed points such that, for numerical reasons, $\#X_i^{(j)} = \left\lceil \frac{1}{\mu^{(j)}} \right\rceil^{2+i}$. We fix the refinement parameter in the first direction to be $\mu^{(1)} = 0.4$ and compute $\mu^{(2)}$ according to the assumption in Theorem 4.8, i.e. such that

$$\epsilon^{(1)} = \left(C_1^{(1)}\mu^{(1)}\right)^{\beta_1} = \left(C_1^{(2)}\mu^{(2)}\right)^{\beta_2} = \epsilon^{(2)}.$$

Assuming that $C_1^{(1)} = C_1^{(2)}$, this means

$$\mu^{(2)} = \left(\mu^{(1)}\right)^{\frac{\beta_1}{\beta_2}} = 0.4^{\frac{1}{2}}.$$

This leads to cardinalities $\#X_i^{(1)} = 2^{2+i}$ and $\#X_i^{(2)} = 3^{2+i}$. Additionally, we set the overlap parameter $\nu^{(j)}$ to 8.0 for every direction j. This results in direction-independent condition numbers of the kernel matrices.

To determine the anisotropy, characterised by the quotient of the components of the weight vector $\boldsymbol{\omega} \in \mathbb{R}^d_+$, we couple ω_j to β_j, the smoothness of $f_{\boldsymbol{\alpha}}$ in direction j, and normalise it with respect to ω_1. This leads to the weight vector $\boldsymbol{\omega} = (1, 2)^{\mathrm{T}}$.
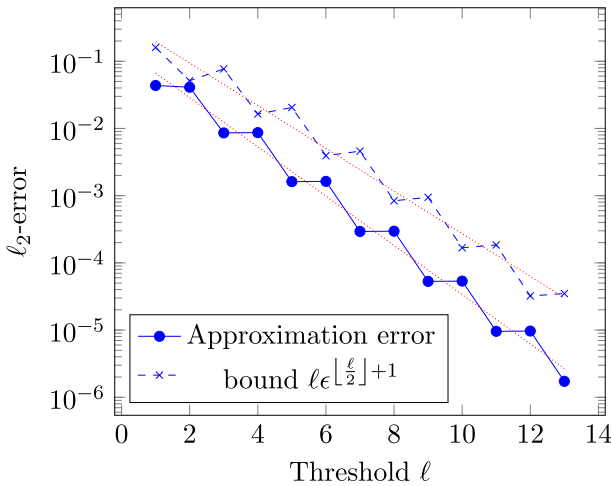
Fig. 2 Error plot for the anisotropic case: $\ell_2([-1, 1]^2)$-error evaluated on 15000×15000 grid points

We compute the ℓ_2-errors on a uniform 15000×15000 grid and show the error in Fig. 2. Additionally, we give again as the dashed line the values of the expected behaviour $\ell \varepsilon^{\lfloor \frac{\ell}{2} \rfloor + 1}$. To gain an impression of the convergence speed we provide a linear fit for each graph, shown as a dotted red line. We note that the obvious steps in the error graph are not surprising. We have the formula

$$L^{(j)} = \left\lfloor \frac{\ell \omega_1}{\omega_j} \right\rfloor + 1,$$

which indicates the maximum level $L^{(j)}$ per direction for a given ℓ. We see that this $L^{(j)}$ exhibits steps due to the Gaussian braces. The mild but unexpected increase in the error steps is most likely due to numerical cancellation errors caused by parallelisation.

Again, we can see that the method converges slightly faster than the theory suggests.

6 Summary

In this paper, we have, for the first time, combined the multilevel radial basis function method with Smolyak's algorithm. We have provided a rigorous error analysis for both the isotropic and anisotropic situation. Some advantages of this new tensor product multilevel method are as follows. The underlying domains Ω_j do not have to be intervals nor one-dimensional. Though the error analysis requires the low-dimensional data to be quasi-uniform, this is not necessary for the actual computation. The method can deal with arbitrary point locations in the low dimensional domains, leading to sparse grids based on scattered data. The numerical examples provided show that the method has the potential to perform well in moderately high-dimensional situations. However, more research is necessary for example to find computationally more

efficient representations of the high-dimensional approximation operator. Here, optimised combination techniques could and should be studied. Moreover, we plan to further study the anisotropic case and make better use of the different smoothnesses in different directions.

References

1. Bungartz, H.-J., Griebel, M.: Sparse grids. In: Iserles, A. (ed.) Acta Numerica, vol. 13, pp. 1–123. Cambridge University Press, Cambridge (2004)
2. Smolyak, S.A.: Quadrature and interpolation formulas for tensor products of certain classes of functions. Soviet Math. Dokl. **4**, 240–243 (1963)
3. Barthelmann, V., Novak, E., Ritter, K.: High dimensional polynomial interpolation on sparse grids. Adv. Comput. Math. **12**, 273–288 (2000)
4. Conrad, P.R., Marzouk, Y.M.: Adaptive Smolyak pseudospectral approximations. SIAM J. Sci. Comput. **35**(6), A2643–A2670 (2013)
5. Tempone, R., Wolfers, S.: Smolyak's algorithm: a powerful black box for the acceleration of scientific computations. In: Sparse Grids and Applications–Miami, Lect. Notes Comput. Sci. Eng., vol. 123, pp. 201–228. Springer, Cham (2016)
6. Ullrich, T.: Smolyak's algorithm, sampling on sparse grids and Sobolev spaces of dominating mixed smoothness. East J. Approx. **14**(1), 1–38 (2008)
7. Nobile, F., Tempone, R., Webster, C.G.: A sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal. **46**(5), 2309–2345 (2008)
8. Nobile, F., Tempone, R., Webster, C.G.: An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal. **46**(5), 2411–2442 (2008)
9. Bungartz, H.-J., Griebel, M., Röschke, D., Zenger, C.: Two proofs of convergence for the combination technique for the efficient solution of sparse grid problems. In: Domain Decomposition Methods in Scientific and Engineering Computing, University Park, PA: Contemp. Math., vol. 180, Amer. Math. Soc. Providence, RI, pp. 15–20 (1993)
10. Garcke, J.: A dimension adaptive sparse grid combination technique for machine learning. ANZIAM J. **48**, C725–C740 (2009)
11. Garcke, J., Hegland, M.: Fitting multidimensional data using gradient penalties and the sparse grid combination technique. Computing **84**, 1–25 (2009)
12. Griebel, M., Schneider, M., Zenger, C.: A combination technique for the solution of sparse grid problems. In: Iterative Methods in Linear Algebra (Brussels, 1991), pp. 263–281. North-Holland, Amsterdam (1992)
13. Hegland, M., Garcke, J., Challis, V.: The combination technique and some generalisations. Linear Algebra Appl. **420**(2), 249–275 (2007)
14. Hegland, M., Harding, B., Kowitz, C., Pflüger, C., Strazdins, P.: Recent developments in the theory and application of the sparse grid combination technique. In: Software for Exascale computing—SPPEXA 2013–2015, Lect. Notes Comput. Sci. Eng., vol. 113, pp. 143–163. Springer, Cham (2016)

15. Lastdrager, B., Koren, B., Verwer, J.: The sparse-grid combination technique applied to time-dependent advection problems. Appl. Numer. Math. **38**(4), 377–401 (2001)
16. Obersteiner, M., Bungartz, H.J.: spatially adaptive sparse grid combination technique for numerical quadrature. In: Sparse Grids and Applications–Munich, Lect. Notes Comput. Sci. Eng., vol. 144, pp. 161–185. Springer, Cham (2018)
17. Obersteiner, M., Bungartz, H.-J.: A generalized spatially adaptive sparse grid combination technique with dimension-wise refinement. SIAM J. Sci. Comput. **43**(4), A2381–A2403 (2021)
18. Reisinger, C.: Analysis of linear difference schemes in the sparse grid combination technique. IMA J. Numer. Anal. **33**(2), 544–581 (2013)
19. Rüttgers, A., Griebel, M.: Multiscale simulation of polymeric fluids using the sparse grid combination technique. Appl. Math. Comput. **319**, 425–443 (2018)
20. Hinojosa, A.P., Harding, B., Hegland, M., Bungartz, H.J.: Handling silent data corruption with the sparse grid combination technique. In: Software for Exascale Computing—SPPEXA 2013–2015, Lect. Notes Comput. Sci. Eng., vol. 113, pp. 187–208. Springer, Cham (2016)
21. Avila, G., Carrington, T.: Solving the Schroedinger equation using Smolyak interpolants. J. Chem. Phys. **139**, 134114 (2013)
22. Narayan, A., Jakeman, J.D.: Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation. SIAM J. Sci. Comput. **36**(6), A2952–A2983 (2014)
23. Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge (2005)
24. Fasshauer, G.: Meshfree Approximation Methods with MATLAB. World Scientific Publishers, Singapore (2007)
25. Le Gia, Q.T., Sloan, I., Wendland, H.: Multiscale analysis in Sobolev spaces on the sphere. SIAM J. Numer. Anal. **48**, 2065–2090 (2010)
26. Wendland, H.: Multiscale analysis in Sobolev spaces on bounded domains. Numer. Math. **116**, 493–517 (2010)
27. Wendland, H.: Multiscale radial basis functions. In: Pesenson, I., Le Gia, Q.T., Mayeli, A., Mhaskar, H., Zhou, D.-X. (eds.) Frames and Other Bases in Abstract and Function Spaces—Novel Methods in Harmonic Analysis, vol. 1, pp. 265–299. Birkhäuser, Cham (2017)
28. Georgoulis, E.H., Levesley, J., Subhan, F.: Multilevel sparse kernel-based interpolation. SIAM J. Sci. Comput. **35**, A815–A831 (2013)
29. Usta, F., Levesley, J.: Multilevel quasi-interpolation on a sparse grid with the Gaussian. Numer. Algorithms **77**, 793–808 (2018)
30. Cheney, E.W., Light, W.P.: Approximation Theory in Tensor Product Spaces. Springer, Berlin (1985)
31. Aubin, J.P.: Applied Functional Analysis, 2nd edn. Wiley, New York (2000)
32. Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus. Springer, Berlin (2012)
33. Wasilkowski, G.W., Wozniakowski, H.: Explicit cost bounds of algorithms for multivariate tensor product problems. J. Complex. **11**, 1–56 (1995)
34. Rieger, C., Wendland, H.: Sampling inequalities for anisotropic tensor product grids. IMA J. Numer. Anal. **40**, 285–321 (2020)
35. Haji-Ali, A.-L., Harbrecht, H., Peters, M.D., Siebenmorgen, M.: Novel results for the anisotropic sparse grid quadrature. J. Complex. **47**, 62–85 (2018)
36. Griebel, M., Oettershagen, J.: On tensor product approximation of analytic functions. J. Approx. Theory **207**, 348–379 (2016)
37. Wahba, G.: *Spline models for observational data*. CBMS-NSF, Regional Conference Series in Applied Mathematics, Siam, Philadelphia, (1990)
38. Coifman, R.R., Lafon, S.: Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions. Appl. Comput. Harmon. Anal. 31–52 (2006)
39. Floater, M.S., Iske, A.: Multistep scattered data interpolation using compactly supported radial basis functions. J. Comput. Appl. Math. **73**, 65–78 (1996)
40. Le Gia, Q.T., Sloan, I., Wendland, H.: Multiscale RBF collocation for solving PDEs on spheres. Numer. Math. **121**, 99–125 (2012)
41. Le Gia, Q.T., Wendland, H.: Data compression on the sphere using multiscale radial basis functions. Adv. Comput. Math. **40**, 923–943 (2014)
42. Farrell, P., Wendland, H.: RBF multiscale collocation for second order elliptic boundary value problems. SIAM J. Numer. Anal. **51**, 2403–2425 (2013)

43. Farrell, P., Gillow, K., Wendland, H.: Multilevel interpolation of divergence-free vector fields. IMA J. Numer. Anal. **37**, 332–353 (2017)
44. Rieger, C., Wendland, H.: Sampling inequalities for sparse grids. Numer. Math. **136**, 439–466 (2017)