ON THE APPROXIMABILITY AND CURSE OF DIMENSIONALITY OF CERTAIN CLASSES OF HIGH-DIMENSIONAL FUNCTIONS

CHRISTIAN RIEGER* AND HOLGER WENDLAND†

Abstract. In this paper, we study the approximability of high-dimensional functions that appear, for example, in the context of many body expansions and hdmr (high-dimensional model representation). Such functions, though high dimensional, can be represented as finite sums of lower dimensional functions. We will derive sampling inequalities for such functions, give explicit advice on the location of good sampling points and show that such functions do not suffer from the curse of dimensionality.

Key words. high-dimensional approximation, high-dimensional model representation, curse of dimensionality, mixed regularity Sobolev spaces, sampling inequalities.

1. Introduction. The approximation of high-dimensional functions $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ using only a finite number of samples $f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_N) \in \Omega$ is an extremely challenging problem because of the *curse of dimensionality*, see [4] as the original source of this term and [14] for an overview in approximation theory.
In this paper, we will discuss this topic for *Sobolev* and *mixed regularity Sobolev functions*.
In the first case, functions $f \in H^\sigma(\Omega)$ are approximated. To give the occuring point evaluations $f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_N)$ a proper meaning, we need the *Sobolev embedding theorem*, which states that for $\sigma > d/2$ the functions can be interpreted as continuous functions. This assumption itself is already critical for high dimensions d. Nonetheless, it is well-known that such functions allow good approximations, for example with radial basis functions (see [21]). Such approximations $I_X f \in H^\sigma(\Omega)$ use only the information of f on $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ and satisfy error estimates of the form

$$\|f - I_X f\|_{L_\infty(\Omega)} \le C h_{X,\Omega}^{\sigma - d/2} \|f\|_{H^\sigma(\Omega)}. \tag{1.1}$$

Here, $h_{X,\Omega}$ is the *fill distance* or *mesh size* of X in Ω, to be defined more precisely later on. If the data sites form a quasi-uniform grid on Ω then the number of points satisfies $N \approx h_{X,\Omega}^{-d}$ and the approximation result becomes

$$\|f - I_X f\|_{L_\infty(\Omega)} \le C N^{-\sigma/d + 1/2} \|f\|_{H^\sigma(\Omega)}. \tag{1.2}$$

The smoothness requirement $\sigma > d/2$ obviously is quite restrictive in high dimensions. Nonetheless, for illustration purposes, assume that $\sigma = d$. Then the error estimate becomes

$$\|f - I_X f\|_{L_\infty(\Omega)} \le C N^{-1/2} \|f\|_{H^\sigma(\Omega)}.$$

This seems not too bad as the convergence rate, though slow, is independent of the dimension. However, there are two caveats to this. First of all, the constant C depends on the space dimension and it can depend exponentially on the space dimension. Secondly, such estimates usually only hold, if the fill distance is sufficiently small. Most theoretical bounds require

$$h_{X,\Omega} \le h_0 := c/\sigma^2. \tag{1.3}$$

This means that in the case of $N \approx h_{X,\Omega}^{-d}$, the estimates hold if

$$N \ge \sigma^{2d}/c^d \ge d^{2d}/(2c)^d$$

* FB Mathematik und Informatik, Philipps-Universität Marburg, 35032 Marburg, Germany
† Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany

which is more than exponential in d.
The situation is slightly better if mixed-regularity Sobolev spaces are considered. First of all, the Sobolev embedding theorem for such spaces shows that $H^\sigma_{\text{mix}}(\Omega) \subseteq C(\overline{\Omega})$ is already satisfied if $\sigma > 1/2$. Hence, this is entirely decoupled from the space dimension.
Standard tensor product approximations on quasi-uniform point sets yield again error estimates similar to those in (1.1) and (1.2). This time, however, the smoothness σ decouples from the space dimension d. Showing that the rate σ/d becomes arbitrarily slow for a fixed $\sigma > 1/2$ with increasing space dimension d. Moreover, the requirement of the points to form a quasi-uniform set means that the curse of dimensionality can not be avoided, as the number of points is once again growing exponentially in d. For example, if the points form a regular grid with $m_{q-d+1} := 2^{q-d} + 1$, $q \geq d$ points (this rather peculiar notation is due to the upcoming application in sparse grids), the total number is given by $N = (2^{q-d} + 1)^d$. In the context of mixed regularity Sobolev spaces usually sparse grids are employed as the location of the data sites to achieve a better cost benefit ratio. Such sparse grids allow estimates of the form

$$\|f - I_X f\|_{L_\infty(\Omega)} \leq C(\log N)^{\rho(\sigma,d)} N^{-\sigma+1/2} \|f\|_{H^\sigma_{\text{mix}}(\Omega)},$$

where $\rho(\sigma, d)$ is a certain, given exponent. Obviously, except for the logarithmic term, these estimates are almost optimal and hardly suffer from the curse of dimensionality. However, as mentioned above, this kind of convergence is achieved with data sites on sparse grids. If, for example, these sparse grids are formed from univariate grids with $m_1 = 1$ and $m_j = 2^{j-1} + 1$, $2 \leq j \leq q - d + 1$, points then the number of points in the sparse grid still grows exponentially. It asymptotically behaves like

$$N \approx \frac{q^{d-1}}{(d-1)!} 2^{q+1-2d}, \qquad q \to \infty, \tag{1.4}$$

see [12, Lemma 1]. Hence, even if they reduce the number of points significantly when compared to full grid methods, the curse of dimensionality still applies.
However, in many applications the involved high-dimensional functions have a specific structure. It is a common observation that functions defined on a high-dimensional domain can be decomposed or fragmented into suitable combinations of several functions depending only on subsets of the input variables.
In statistics such decompositions are often named after Hoeffding and Sobol. The most prominent example of such decompositions is the *Analysis of Variance (ANOVA)*. In the physical literature one often finds such decompositions under the name *Many body expansion*, see for instance [8] for a brief overview from the standpoint of high dimensional approximation.
We will focus on a specific decomposition which is often called *High-dimensional model representation*, see [16] for an overview on this method. Nowadays, these decomposition methods find also applications in data analysis. See for instance [24] for specific software solutions to employ such methods in sensitivity analysis. Lately, these methods gained also attention in the machine learning context, see for instance [3] for specific examples employing the HDMR decomposition and [11] for the combination of HDMR and kernel based learning.
All these methods have in common that a function $f : \mathbb{R}^d \to \mathbb{R}$ can be represented or approximated as a sum of functions, which depend on fewer variables

$$\begin{aligned} f(x_1, \ldots, x_d) = f_0 &+ f_1(x_1) + f_2(x_2) + \cdots + f_d(x_d) \\ &+ f_{1,2}(x_1, x_2) + f_{1,3}(x_1, x_3) + \cdots + f_{d-1,d}(x_{d-1}, x_d) + \cdots \end{aligned}$$

The most general form, which we will discuss in this paper uses the following setup. Let $\mathfrak{D} = \{1, \ldots, d\}$ and let $\mathcal{P}(d) = \{\mathfrak{u} : \mathfrak{u} \subseteq \mathfrak{D}\}$ be the set of all subsets of $\mathfrak{D}$. If $\Lambda \subseteq \mathcal{P}(d)$ is a specific collection of subsets, we will discuss functions which can be represented as

$$f = \sum_{\mathfrak{u} \in \Lambda} f_{\mathfrak{u}},$$

where the components $f_{\mathfrak{u}}$ depend only on the variables with indices in $\mathfrak{u}$. We will derive certain *sampling inequalities* for such functions from Sobolev spaces $H^\sigma_\Lambda(\Omega)$ and mixed regularity Sobolev spaces $H^\sigma_{\mathrm{mix},\Lambda}(\Omega)$ given a fixed $\Lambda \subset \mathcal{P}(d)$. The precise definitions of these spaces follow later on.
We will then use these sampling inequalities to show that for such function classes, under mild assumption on the set Λ, the curse of dimensionality does not apply, i.e. The cost of approximating a normed function from such a space up to an accuracy $\epsilon > 0$ depends only polynomially on the space dimension d. Similar considerations were previously done in a probabilistic context using deep neural networks, see for example [18]. However, while dealing with a more general situation, the results there are non constructive, as they do not show how the approximations actually can be computed. Other considerations specifically for periodic functions using Fourier transformation techniques can, for example, be found in [15].
The paper itself is organised as follows. In the next section, we review the necessary result on function decompositions. In the third section, we collect relevant known and prove new results on Sobolev spaces and mixed-regularity Sobolev spaces. The fourth section deals with lower order sub-classes of Sobolev spaces. Here, we derive our first main result, namely a sampling inequality for such sub-classes. We also address the curse of dimensionality for these functions. In the fifth section, we prove the corresponsing results for sub-classes of mixed-regularity Sobolev spaces. In the final section, we give an illustrative example.

2. Decomposition of Functions. For a very high dimensional function, it is often possible or attempted to write such a function as a sum of functions that depend on less variables. Typical decompositions of this type are the so-called *ANOVA decomposition* and the so-called *anchored decomposition*. As pointed out in [10], both can more generally be describe using *eliminating projections*. We will follow their approach and cite the following results and definitions.

DEFINITION 2.1. *Let V be a linear space of real functions $f : \Omega \to \mathbb{R}$. Let $\mathfrak{D} := \{1, \dots, d\}$. For $\mathfrak{u} \subseteq \mathfrak{D}$ and $\boldsymbol{x} = (x_1, \dots, x_d) \in \Omega$ let $\boldsymbol{x}_{\mathfrak{u}} = (x_j : j \in \mathfrak{u})$ be the vector containing only those components with index in $\mathfrak{u}$.*
A function $f \in V$ does not depend on the variables with index in $\mathfrak{u} \subseteq \mathfrak{D}$ if for every $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ the equality $\boldsymbol{x}_{\mathfrak{D}\setminus\mathfrak{u}} = \boldsymbol{y}_{\mathfrak{D}\setminus\mathfrak{u}}$ always implies $f(\boldsymbol{x}) = f(\boldsymbol{y})$. The function depends (only) on the variables with index in $\mathfrak{u}$ if it does not depend on the variables with index in $\mathfrak{D} \setminus \mathfrak{u}$.

We will use projections to eliminate variables from a given function $f \in V$.

DEFINITION 2.2. *Let $P_j : V \to V$, $1 \le j \le d$, be linear and bounded operators. Then, P_j is said to* eliminate *the j-th variable if $P_j f$ does not depend on the j-th variable and if $P_j f = f$ for every $f \in V$, which does not depend on the j-th variable.*

In what follows, we will also use the notation $P_\emptyset = I$ for the identity operator and for $\mathfrak{u} \subseteq \mathfrak{D}$, $\mathfrak{u} \neq \emptyset$,

$$P_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} P_j.$$

In our applications, the operators P_j will always commute such that the ordering in the product above is irrelevant.
The following result from [10] will be crucial to our approach later on. From now on, we will denote the number of elements in a set $\mathfrak{u}$ by $\#\mathfrak{u}$.

THEOREM 2.3. *Let V be a linear space of real functions defined on $\Omega \subseteq \mathbb{R}^d$ and let $P_j : V \to V$, $1 \le j \le d$, be commuting operators, eliminating the j-th variable, respectively. For any function*

$f \in V$ and any subset $\mathfrak{u} \subseteq \mathfrak{D} = \{1, \dots, d\}$ let

$$f_{\mathfrak{u}} = \left(\prod_{j \in \mathfrak{u}} (I - P_j) \right) P_{\mathfrak{D} \setminus \mathfrak{u}} f. \tag{2.1}$$

Then, $f_{\mathfrak{u}}$ only depends on the variable with indices in $\mathfrak{u}$ and f can be decomposed into

$$f = \sum_{\mathfrak{u} \subseteq \mathfrak{D}} f_{\mathfrak{u}}. \tag{2.2}$$

Moreover, the following holds.

1. *The functions $f_{\mathfrak{u}}$, $\mathfrak{u} \subseteq \mathfrak{D}$, are recursively given by $f_{\emptyset} = P_{\mathfrak{D}} f$ and*
$$f_{\mathfrak{u}} = P_{\mathfrak{D} \setminus \mathfrak{u}} f - \sum_{\mathfrak{v} \subsetneq \mathfrak{u}} f_{\mathfrak{v}},$$
for $\mathfrak{u} \subseteq \mathfrak{D}$, $\mathfrak{u} \neq \emptyset$.
2. *The functions $f_{\mathfrak{u}}$ have the representation*
$$f_{\mathfrak{u}} = \sum_{\mathfrak{v} \subseteq \mathfrak{u}} (-1)^{\#\mathfrak{u} - \#\mathfrak{v}} P_{\mathfrak{D} \setminus \mathfrak{v}} f. \tag{2.3}$$
3. *The functions $f_{\mathfrak{u}}$ satisfy the* annihilation property
$$P_j f_{\mathfrak{u}} = 0, \qquad j \in \mathfrak{u}.$$
Moreover, if $f = \sum_{\mathfrak{u} \subseteq \mathfrak{D}} g_{\mathfrak{u}}$ is any decomposition in which $g_{\mathfrak{u}} \in V$ only depends on the variables with indices in $\mathfrak{u}$ and if the functions $g_{\mathfrak{u}}$ satisfy the annihilation property then $g_{\mathfrak{u}} = f_{\mathfrak{u}}$ for all $\mathfrak{u} \subseteq \mathfrak{D}$. In this way, the decomposition (2.2) is unique.
4. *If $f = \sum_{\mathfrak{u} \subseteq \mathfrak{D}} g_{\mathfrak{u}}$ is any representation with terms $g_{\mathfrak{u}} \in V$ depending only on the variables with indices in $\mathfrak{u}$ and if $\mathfrak{v} \subseteq \mathfrak{D}$ is a subset such that $g_{\mathfrak{u}} = 0$ for all $\mathfrak{u} \subseteq \mathfrak{D}$ with $\mathfrak{v} \subseteq \mathfrak{u}$ then $f_{\mathfrak{u}} = 0$ for all such $\mathfrak{u}$, as well.*

The last property means that decompositions based on such projections are minimal in the sense that they try to avoid terms with a large number of variables. This can be made more precise by introducing the following concept.

DEFINITION 2.4. *Let $d \in \mathbb{N}$ be given and let $\mathcal{P}(d) = \{\mathfrak{u} \subseteq \mathfrak{D}\}$ be the set of all subsets of $\mathfrak{D} = \{1, \dots, d\}$. Let $\Lambda \subseteq \mathcal{P}(d)$ be a fixed set of subsets of $\mathfrak{D}$.*
A function $f : \Omega \to \mathbb{R}$ has a Λ-representation*, if it can be written as*

$$f = \sum_{\mathfrak{u} \in \Lambda} \widetilde{f}_{\mathfrak{u}}, \tag{2.4}$$

where each $\widetilde{f}_{\mathfrak{u}}$ is a function depending only on the variables with indices in $\mathfrak{u}$.

With this concept, we are able to describe very high-dimensional functions, as long as the number $\#\Lambda$ of *building blocks* can be bounded reasonably. The last point of Theorem 2.3 means for functions with Λ-representation that their projection form does not employ terms with more variables.

COROLLARY 2.5. *If $\Lambda \subseteq \mathcal{P}(d)$ is a given set of subsets and if $\overline{\Lambda} = \{\mathfrak{v} \subseteq \mathfrak{u} : \mathfrak{u} \in \Lambda\}$ contains all sets from Λ and their subsets, then any function of the form (2.4) has a decomposition of the form*

$$f = \sum_{\mathfrak{u} \in \overline{\Lambda}} f_{\mathfrak{u}},$$

with $f_{\mathfrak{u}}$ from (2.3).

Hence, the projection form of Theorem 2.3 may contain additional terms $f_{\mathfrak{u}}$ of fewer variables. To have exactly the same variable sets, we need to assume that with each set $\mathfrak{u} \in \Lambda$ also the subsets $\mathfrak{v} \subseteq \mathfrak{u}$ are elements of Λ.

DEFINITION 2.6. *A set of sets Λ is called* downward closed *if $\mathfrak{u} \in \Lambda$ and $\mathfrak{v} \subseteq \mathfrak{u}$ always implies $\mathfrak{v} \in \Lambda$.*

If starting with a general set Λ then we would have to employ the set $\overline{\Lambda}$, which only increases the number $\#\Lambda$ of sets to

$$\#\overline{\Lambda} = \sum_{\mathfrak{u}\in\Lambda} 2^{\#\mathfrak{u}}.$$

In our applications this is still an acceptable increase of sets. For this reason, from now on, we will include all these sets beforehand and will look only at downward closed sets Λ.
The two most common classes of such downward closed sets are given for a number $n \le d$ by the index sets

$$\Lambda = \{\mathfrak{u} \subseteq \mathfrak{D} : \#\mathfrak{u} \le n\},$$
$$\Lambda = \{\mathfrak{u} \subseteq \mathfrak{D} : \mathfrak{u} \subseteq \{1, \dots, n\}\}.$$

The corresponding functions are functions which can be written as sums of functions depending only on n or only on the first n variables. Such functions were, for example, further studied in [10].

DEFINITION 2.7.

1. *A function $f \in V$, $f : \Omega \to \mathbb{R}$ has* cutoff dimension n *if it depends only on the first n variables, i.e. if it has a representation*

$$f = \sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} f_{\mathfrak{u}}.$$

and n is the smallest such number.

2. *A function $f \in V$ is of* order n *if it can be written as a sum of functions depending only on n variables, i.e. if it has a representation*

$$f = \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D}\\ \#\mathfrak{u}\le n}} f_{\mathfrak{u}}$$

and n is the smallest such number.

Hence, if f is, for example, an order-n-function, then its representation (2.2) contains only terms $\mathfrak{u} \subseteq \mathfrak{D}$ with $\#\mathfrak{u} \le n$. A corresponding result holds if f has cutoff dimension n. Another way of expressing Corollary 2.5 in this context is the following.

COROLLARY 2.8. *The decomposition in Theorem 2.3 has both the smallest cutoff dimension and order.*

From the point of view of approximating high-dimensional functions, order-n-functions are more interesting, as they allow arbitrary space dimensions d. Functions with cutoff dimension n are essentially functions defined on n-dimensional domains. Hence, in this paper, we will particularly pursue the case of order-n-functions, as they describe the worst case of all Λ with sets $\mathfrak{u} \in \Lambda$ having at most n elements. The number of elements in this set has the following, obvious bounds, which are, for a fixed $n < d$ only polynomial in d.

LEMMA 2.9. *The number of elements in $\Lambda := \{\mathfrak{u} \subseteq \mathfrak{D} : \#\mathfrak{u} \le n\}$ is bounded by*

$$\left(\frac{d}{n}\right)^n \le \#\Lambda = \sum_{j=0}^{n} \binom{d}{j} \le \left(\frac{ed}{n}\right)^n,$$

Proof. This immediately follows from

$$\sum_{j=0}^{n}\binom{d}{j} \geq \binom{d}{n} \geq \left(\frac{d}{n}\right)^{n}, \tag{2.5}$$

where the last inequality is easily proven by induction, and from

$$\left(\frac{n}{d}\right)^{n}\sum_{j=0}^{n}\binom{d}{j} \leq \sum_{j=0}^{n}\binom{d}{j}\left(\frac{n}{d}\right)^{j} \leq \sum_{j=0}^{d}\binom{d}{j}\left(\frac{n}{d}\right)^{j} = \left(1+\frac{n}{d}\right)^{d} \leq e^{n}.$$

∎

These general concepts apply to any projections as described above. For example, we could use an ANOVA decomposition, which is achieved by using integration for the projection operators. However, here, we will work with specific projections, which are sometimes called *freezing*.

DEFINITION 2.10. *Let $\boldsymbol{a} \in \Omega$ be fixed. Then, freezing at a_j is defined by the projection*

$$P_j f(x_1,\dots,x_d) = f(x_1,\dots,x_{j-1},a_j,x_{j+1},\dots,x_d), \qquad \boldsymbol{x} \in \Omega.$$

Obviously, for this to be well-defined, we need that the argument on the right-hand side also belongs to Ω. We will make sure that this is always the case.
Freezing corresponds to so-called *anchored spaces* and the point $\boldsymbol{a} = (a_1,\dots,a_d)$ is called the anchor. For these projections, we particularly have

$$f_{\mathfrak{u}}(\boldsymbol{x}) = \sum_{\mathfrak{v}\subseteq\mathfrak{u}}(-1)^{\#\mathfrak{u}-\#\mathfrak{v}} f((\boldsymbol{x};\boldsymbol{a})_{\mathfrak{v}}), \tag{2.6}$$

where we denoted the vector whose j-th component is x_j if $j \in \mathfrak{u}$ and a_j if $j \notin \mathfrak{u}$ by $(\boldsymbol{x};\boldsymbol{a})_{\mathfrak{u}}$. With this notation, it is easy to see that the the representation of Theorem 2.3 has the following form for order-n-functions, see also [10].

COROLLARY 2.11. *Let $\Omega \subseteq \mathbb{R}^d$ and $V \subseteq C(\Omega)$ such that an anchor $\boldsymbol{a} \in \Omega$ exists with $(\boldsymbol{x};\boldsymbol{a})_{\mathfrak{u}} \in \Omega$ for all $\boldsymbol{x} \in \Omega$ and all $\mathfrak{u} \subseteq \mathfrak{D}$. Then, any order-$n$-function $f \in V$ has a representation*

$$f = \sum_{j=0}^{n}(-1)^{n-j}\binom{d-1-j}{n-j}\sum_{\substack{\mathfrak{u}\in\mathfrak{D}\\ \#\mathfrak{u}=j}} f((\cdot;\boldsymbol{a})_{\mathfrak{u}}). \tag{2.7}$$

3. Sobolev and mixed regularity Sobolev spaces. We will consider functions that are defined on a bounded domain $\Omega \subseteq \mathbb{R}^d$ which is of Cartesian product structure, i.e. which has the form

$$\Omega = \Omega_1 \times \cdots \times \Omega_d,$$

with $\Omega_j \subseteq \mathbb{R}$. In theory, we could also consider the more general situation $\Omega_j \subseteq \mathbb{R}^{n_j}$. However, while this is more general, it also complicates the notations significantly. Hence, we will refrain from discussing this more general approach but the reader should keep in mind that such a generalisation is easily possible.
For each $\mathfrak{u} = \{u_1,\dots,u_n\} \subseteq \mathfrak{D} := \{1,\dots,d\}$ we define the Cartesian products

$$\Omega_{\mathfrak{u}} := \begin{cases} \Omega_{u_1} \times \cdots \times \Omega_{u_n} & \text{if } \mathfrak{u} \neq \emptyset, \\ \Omega_1 & \text{if } \mathfrak{u} = \emptyset \end{cases}$$

and will, throughout this text, make the following assumption.

ASSUMPTION 3.1. *Let $\Omega = \Omega_1 \times \cdots \times \Omega_d$ with $\Omega_j \subseteq \mathbb{R}$, $1 \le j \le d$ be given.*

- *For each $\mathfrak{u} \subseteq \mathfrak{D}$ the sets $\Omega_{\mathfrak{u}}$ are bounded and have a Lipschitz boundary. Thus, they satisfy an interior cone condition with angle $\theta_{\mathfrak{u}}$ and radius $r_{\mathfrak{u}}$.*
- *There is an element $\boldsymbol{a} \in \Omega$, called* anchor*, which satisfies*

$$(\boldsymbol{x}; \boldsymbol{a})_{\mathfrak{u}} \in \Omega$$

for all $\boldsymbol{x} \in \Omega$ and all $\mathfrak{u} \subseteq \mathfrak{D}$.

This is obviously satisfied if $\Omega_j = [0, 1]$, or, for that matter, any other interval, where $\Omega_{\mathfrak{u}}$ is essentially simply $[0, 1]^n$ if $n = \#\mathfrak{u}$ is the number of indices in $\mathfrak{u}$.

3.1. Sobolev spaces of anchored functions. We want to shows that the components (2.6) of an anchored decomposition of a Sobolev function are Sobolev functions themselves. Hence, we begin with a short recap of Sobolev spaces, mainly to settle the notation. The space $H^m(\Omega)$ consists of all functions $f \in L_2(\Omega)$ having weak derivatives $D^{\boldsymbol{\alpha}} f \in L_2(\Omega)$ for all $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ with $|\boldsymbol{\alpha}| \le m$. It is equipped with the norm

$$\|f\|_{H^m(\Omega)} = \left(\sum_{|\boldsymbol{\alpha}| \le m} \|D^{\boldsymbol{\alpha}} f\|_{L_2(\Omega)}^2 \right)^{1/2}.$$

There are different ways of extending the definition to *fractional order* Sobolev spaces $H^\sigma(\Omega)$ with $\sigma \in \mathbb{R}$, $\sigma > 0$. We will assume that the reader is familiar with this and refer only, for example, to [1] for details. Here, it is important that in the case of $\Omega = \mathbb{R}^d$, there is an equivalent norm on $H^\sigma(\mathbb{R}^d)$, which is given by

$$\|f\|_{\widetilde{H}^\sigma(\mathbb{R}^d)} = \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{\omega})|^2 (1 + \|\boldsymbol{\omega}\|_2^2)^\sigma d\boldsymbol{\omega},$$

using the Fourier transform

$$\widehat{f}(\boldsymbol{\omega}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(\boldsymbol{x}) e^{-i\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}} d\boldsymbol{x}.$$

We need the following results which are essentially generalisations of trace and extension results of Sobolev functions.

PROPOSITION 3.2. *Let $1 \le n < d$. Let $\Omega \subseteq \mathbb{R}^d$ be decomposed into $\Omega = \widetilde{\Omega}_1 \times \widetilde{\Omega}_2$ with $\widetilde{\Omega}_1 \subseteq \mathbb{R}^n$ and $\widetilde{\Omega}_2 \subseteq \mathbb{R}^{d-n}$. Assume that Ω and $\widetilde{\Omega}_1$ have Lipschitz boundaries. For $\boldsymbol{\omega} \in \Omega$ let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ be the decomposition with $\boldsymbol{\omega}_1 \in \widetilde{\Omega}_1$ and $\boldsymbol{\omega}_2 \in \widetilde{\Omega}_2$. Let $\sigma > (d-n)/2$ and $\tau = \sigma - (d-n)/2$. Fix $\boldsymbol{a} \in \Omega$.*

1. *There exists a continuous linear trace operator $T_{\Omega, \widetilde{\Omega}_1} : H^\sigma(\Omega) \to H^\tau(\widetilde{\Omega}_1)$ satisfying*

$$T_{\Omega, \widetilde{\Omega}_1} f(\boldsymbol{x}) = f(\boldsymbol{x}_1, \boldsymbol{a}_2), \qquad f \in H^\sigma(\Omega),$$

   for almost all $\boldsymbol{x} \in \Omega$.

2. *There exists a continuous linear extension operator $Z_{\widetilde{\Omega}_1, \Omega} : H^\tau(\widetilde{\Omega}_1) \to H^\sigma(\Omega)$ satisfying*

$$(T_{\Omega, \widetilde{\Omega}_1} \circ Z_{\widetilde{\Omega}_1, \Omega}) g = g, \qquad g \in H^\tau(\widetilde{\Omega}_1).$$

Proof. For both statements, the proof is divided into two steps. In the first step, the result is shown for $\Omega = \mathbb{R}^d = \mathbb{R}^n \times \mathbb{R}^{d-n}$. In the second step, an extension operator is used to finalise the proof. We only prove the result for the trace operator. The proof of the second part is similar.

As pointed out, we start with the case $\Omega = \mathbb{R}^d$ and mainly follow the ideas of [23, Theorem 8.1]. It suffices to consider $f \in C_0^\infty(\mathbb{R}^d)$, where $C_0^\infty(\mathbb{R}^d)$ denotes the space of all infinitely often differentiable functions with compact support, and to use density of these test functions to define the operator $T_0 : H^\sigma(\mathbb{R}^d) \to H^\tau(\mathbb{R}^n)$. For such an f, we can use the Fourier transform to express the Sobolev norm. Noting that for a fixed $\boldsymbol{x}_1 \in \mathbb{R}^n$, the function $\boldsymbol{x}_2 \to f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ belongs to $C_0^\infty(\mathbb{R}^{d-n})$, we can use the inverse Fourier transform to recover $f(\boldsymbol{x}_1, \cdot)$ from its $(n-d)$-variate Fourier transform, i.e., we have

$$\begin{aligned}
f(\boldsymbol{x}_1, \boldsymbol{a}_2) &= (2\pi)^{-(d-n)/2} \int_{\mathbb{R}^{d-n}} f(\boldsymbol{x}_1, \cdot)^\wedge(\boldsymbol{y}_2) e^{i\boldsymbol{y}_2^{\mathrm{T}} \boldsymbol{a}_2} d\boldsymbol{y}_2 \\
&= (2\pi)^{-(d-n)} \int_{\mathbb{R}^{d-n}} \int_{\mathbb{R}^{d-n}} f(\boldsymbol{x}_1, \boldsymbol{z}_2) e^{i\boldsymbol{y}_2^{\mathrm{T}}(\boldsymbol{a}_2 - \boldsymbol{z}_2)} d\boldsymbol{z}_2 d\boldsymbol{y}_2.
\end{aligned}$$

Hence, the Fourier transform with respect to the first n variables is given by

$$\begin{aligned}
& f(\cdot, \boldsymbol{a}_2)^\wedge(\boldsymbol{\omega}_1) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(\boldsymbol{x}_1, \boldsymbol{a}_2) e^{-i\boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{\omega}_1} d\boldsymbol{x}_1 \\
&= (2\pi)^{-d+\frac{n}{2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^{d-n}} \int_{\mathbb{R}^{d-n}} f(\boldsymbol{x}_1, \boldsymbol{z}_2) e^{i\boldsymbol{y}_2^{\mathrm{T}}(\boldsymbol{a}_2 - \boldsymbol{z}_2)} d\boldsymbol{z}_2 d\boldsymbol{y}_2 e^{-i\boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{\omega}_1} d\boldsymbol{x}_1 \\
&= (2\pi)^{-d+\frac{n}{2}} \int_{\mathbb{R}^{d-n}} \left[\int_{\mathbb{R}^n} \int_{\mathbb{R}^{d-n}} f(\boldsymbol{x}_1, \boldsymbol{z}_2) e^{-i(\boldsymbol{y}_2^{\mathrm{T}} \boldsymbol{z}_2 + \boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{\omega}_1)} d\boldsymbol{z}_2 d\boldsymbol{x}_1 \right] e^{i\boldsymbol{y}_2^{\mathrm{T}} \boldsymbol{a}_2} d\boldsymbol{y}_2 \\
&= (2\pi)^{-\frac{d-n}{2}} \int_{\mathbb{R}^{d-n}} \widehat{f}(\boldsymbol{\omega}_1, \boldsymbol{y}_2) e^{i\boldsymbol{y}_2^{\mathrm{T}} \boldsymbol{a}_2} d\boldsymbol{y}_2. \qquad (3.1)
\end{aligned}$$

With this, we can now compute the Sobolev norm of the restricted function. With $\tau = \sigma - (d-n)/2$ we have

$$\begin{aligned}
& \|f(\cdot, \boldsymbol{a}_2)\|_{H^\tau(\mathbb{R}^n)}^2 \le \int_{\mathbb{R}^n} |f(\cdot, \boldsymbol{a}_2)^\wedge(\boldsymbol{\omega}_1)|^2 (1 + \|\boldsymbol{\omega}_1\|_2^2)^\tau d\boldsymbol{\omega}_1 \\
&= (2\pi)^{-(d-n)} \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^{d-n}} \widehat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}_2^{\mathrm{T}} \boldsymbol{a}_2} d\boldsymbol{\omega}_2 \right|^2 (1 + \|\boldsymbol{\omega}_1\|_2^2)^\tau d\boldsymbol{\omega}_1 \\
&= (2\pi)^{-(d-n)} \int_{\mathbb{R}^n} \left[(1 + \|\boldsymbol{\omega}_1\|_2^2)^\tau \int_{\mathbb{R}^{d-n}} |\widehat{f}(\boldsymbol{\omega})|^2 (1 + \|\boldsymbol{\omega}\|_2^2)^\sigma d\boldsymbol{\omega}_2 \times \right. \\
&\qquad \left. \times \int_{\mathbb{R}^{d-n}} (1 + \|\boldsymbol{\omega}\|_2^2)^{-\sigma} d\boldsymbol{\omega}_2 \right] d\boldsymbol{\omega}_1.
\end{aligned}$$

A simple change of variables shows that there is a constant $C_{n,d,\sigma} > 0$ such that the latter integral becomes

$$\int_{\mathbb{R}^{d-n}} (1 + \|\boldsymbol{\omega}\|_2^2)^{-\sigma} d\boldsymbol{\omega}_2 = C_{n,d,\sigma} (1 + \|\boldsymbol{\omega}_1\|_2^2)^{-\tau}$$

As a matter of fact, we can compute this constant in case of $\sigma \in \mathbb{N}$ using [7, 3.241.4], which states for $p, q \neq 0$ and $0 < \mu/\nu < m+1$ that

$$\int_0^\infty \frac{x^{\mu-1}}{(p + qx^\nu)^{m+1}} dx = \frac{1}{\nu} \left(\frac{p}{q}\right)^{\mu/\nu} \frac{\Gamma\left(\frac{\mu}{\nu}\right) \Gamma\left(1 + m - \frac{\mu}{\nu}\right)}{\Gamma(1+m)}. \qquad (3.2)$$

If we set $\mu = d - n$, $p = 1 + \|\boldsymbol{\omega}_1\|_2^2$, $q = 1$, $\nu = 2$ and $m = \sigma - 1$, we have $0 < \mu/\nu = (d-n)/2 < \sigma = m + 1$ and hence derive

$$
\begin{aligned}
\int_{\mathbb{R}^{d-n}} (1+\|\boldsymbol{\omega}\|_2^2)^{-\sigma} d\boldsymbol{\omega}_2 &= \int_{\mathbb{R}^{d-n}} (1+\|\boldsymbol{\omega}_1\|_2^2+\|\boldsymbol{\omega}_2\|_2^2)^{-\sigma} d\boldsymbol{\omega}_2 \\
&= \frac{(d-n)\pi^{\frac{d-n}{2}}}{\Gamma\left(\frac{d-n}{2}\right)} \int_0^\infty (1+\|\boldsymbol{\omega}_1\|_2^2+r^2)^{-\sigma} r^{d-n-1} dr \\
&= \frac{(d-n)\pi^{\frac{d-n}{2}}}{\Gamma\left(\frac{d-n}{2}\right)} \frac{1}{2} \frac{\Gamma\left(\frac{d-n}{2}\right)\Gamma\left(\sigma-\frac{d-n}{2}\right)}{\Gamma(\sigma)} (1+\|\boldsymbol{\omega}_1\|_2^2)^{-\sigma+\frac{d-n}{2}} \\
&=: C_{n,d,\sigma}(1+\|\boldsymbol{\omega}_1\|_2^2)^{-\tau}.
\end{aligned}
$$

Inserting this in the above calculation yields

$$
\begin{aligned}
\|f(\cdot,\boldsymbol{a}_2)\|_{H^\tau(\mathbb{R}^n)}^2 &\le (2\pi)^{-(d-n)} c_1 C_{n,d,\sigma} \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{\omega})|^2 (1+\|\boldsymbol{\omega}\|_2^2)^\sigma d\boldsymbol{\omega} \\
&\le C\|f\|_{H^\sigma(\mathbb{R}^d)}^2
\end{aligned}
$$

for any $f \in C_0^\infty(\mathbb{R}^d)$. For $f \in H^\sigma(\mathbb{R}^d)$ we can find a sequence $(f_k) \in C_0^\infty(\mathbb{R}^d)$ with $\|f - f_k\|_{H^\sigma(\mathbb{R}^d)} \to 0$ for $k \to \infty$. As $(f_k(\cdot,\boldsymbol{a}_2))$ is a Cauchy sequence in $H^\tau(\mathbb{R}^n)$, we can define $T_0 f$ as the $H^\tau(\mathbb{R}^n)$ limit of this sequence. It is easy to see that this is independent of the original sequence and indeed defines the desired operator $T_0 : H^\sigma(\mathbb{R}^d) \to H^\tau(\mathbb{R}^n)$ in the case $\Omega = \mathbb{R}^d$. First of all, the operator is obviously bounded and linear. Moreover, if $f \in H^\sigma(\Omega)$ and $(f_k) \subseteq C_0^\infty(\Omega)$ be a defining sequence, then we have from $\|f - f_k\|_{H^\sigma(\Omega)} \to 0$ and $\|T_0 f - f_k(\cdot,\boldsymbol{a}_2)\|_{H^\tau(\Omega)} \to 0$ that there is a sub-sequence, which we will denote by (f_k) again, satisfying $f_k(\boldsymbol{x}) \to f(\boldsymbol{x})$ for almost all $\boldsymbol{x} \in \Omega$ and $f_k(\boldsymbol{x}_1,\boldsymbol{a}_2) \to T_0 f(\boldsymbol{x}_1)$ for almost all $\boldsymbol{x}_1 \in \widetilde{\Omega}_1$. This shows $T_0 f(\boldsymbol{x}) = f(\boldsymbol{x}_1,\boldsymbol{a}_2)$ for almost all $\boldsymbol{x} \in \Omega$.

For bounded $\Omega \subseteq \mathbb{R}^d$ with Lipschitz boundary we use the existence of a continuous and linear extension operator $E_\Omega : H^\sigma(\Omega) \to H^\sigma(\mathbb{R}^d)$, see [19, Chapter VI] for integer σ and [5] for general σ. The fact that the standard restriction operator $R_{\widetilde{\Omega}_1} : H^\tau(\mathbb{R}^n) \to H^\tau(\widetilde{\Omega}_1)$, $f \mapsto R_{\widetilde{\Omega}_1} f = f|\widetilde{\Omega}_1$ is continuous with bound one, allows us to define $T_{\Omega,\widetilde{\Omega}_1} : H^\sigma(\Omega) \to H^\tau(\widetilde{\Omega}_1)$ by $T_{\Omega,\widetilde{\Omega}_1} := R_{\widetilde{\Omega}_1} \circ T_0 \circ E_\Omega$, which gives the desired operator.

As pointed out at the beginning of the proof. The second part about the extension operator $Z_{\widetilde{\Omega}_1,\Omega}$ is proven in the same fashion. First an extension operator $Z_0 : H^\tau(\mathbb{R}^n) \to H^\sigma(\mathbb{R}^d)$ is defined similarly to [23, Theorem 8.3]. Then, using again a universal extension operator $E_{\widetilde{\Omega}_1} : H^\tau(\widetilde{\Omega}_1) \to H^\tau(\mathbb{R}^n)$ yields together with the restriction operator $R_\Omega : H^\sigma(\mathbb{R}^d) \to H^\sigma(\Omega)$ the desired operator $Z_{\widetilde{\Omega}_1,\Omega} := R_\Omega \circ Z_0 \circ E_{\widetilde{\Omega}_1}$. ∎

With this, we can now state the regularity of the functions $f_\mathfrak{u}$ of the decomposition if we consider them as functions on $\Omega_\mathfrak{u}$.

THEOREM 3.3. *Let $\Omega \subseteq \mathbb{R}^d$ satisfy Assumption 3.1. Let $\sigma > d/2$ be fixed.*

1. *For $f \in H^\sigma(\Omega)$ and $\mathfrak{u} \subseteq \mathfrak{D}$, the functions $f((\cdot;\boldsymbol{a})_\mathfrak{u})$ belong, as functions on $\Omega_\mathfrak{u}$, to $H^{\sigma-(d-\#\mathfrak{u})/2}(\Omega_\mathfrak{u})$.*
2. *For $f \in H^\sigma(\Omega)$ let $f_\mathfrak{u}$, $\mathfrak{u} \subseteq \mathfrak{D}$, be the anchored components (2.6) of the decomposition (2.2). Then, for each $\mathfrak{u} \subseteq \mathfrak{D}$ with $n = \#\mathfrak{u}$ it holds that $f_\mathfrak{u} \in H^{\sigma-(d-n)/2}(\Omega_\mathfrak{u})$ if considered as a function on $\Omega_\mathfrak{u}$.*

Proof. The first statement follows for $\mathfrak{u} \neq \emptyset$ immediately from Proposition 3.2. For $\mathfrak{u} = \emptyset$ we have $f((\cdot,\boldsymbol{a})_\emptyset) = f(\boldsymbol{a})$, i.e. the constant function, which obviously belongs to any $H^\tau(\Omega_1)$, in particular for $\tau = \sigma - d/2$.

By the Sobolev embedding theorem we have $f \in C(\Omega)$. Moreover, any such component $f_\mathfrak{u}$ has a representation (2.6). By the first statement we know that the functions $\boldsymbol{x}_\mathfrak{v} \mapsto f((\boldsymbol{x};\boldsymbol{a})_\mathfrak{v})$ belong to

$H^{\sigma-\frac{d-\#\mathfrak{v}}{2}}(\Omega_{\mathfrak{v}})$. The second statement of Proposition 3.2, interpreting these functions as functions on $\Omega_{\mathfrak{u}}$, yields functions belonging to $H^{\sigma-\frac{d-n}{2}}(\Omega_{\mathfrak{u}})$. ∎

3.2. Mixed regularity Sobolev spaces of anchored functions. We will now derive the corresponding results for mixed regularity Sobolev spaces. To this end, we recall the definition of a mixed regularity Sobolev space.

DEFINITION 3.4. *Let $\Omega \subseteq \mathbb{R}^d$ and $\boldsymbol{m} \in \mathbb{N}_0^d$ be given. The space $H^{\boldsymbol{m}}_{\text{mix}}(\Omega)$ consists of all function $f \in L_2(\Omega)$ having weak derivatives $D^{\boldsymbol{\alpha}} f \in L_2(\Omega)$ for all $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ with $\boldsymbol{\alpha} \leq \boldsymbol{m}$, i.e. with $\alpha_j \leq m_j$, $1 \leq j \leq d$. The space is equipped with the norm*

$$\|f\|_{H^{\boldsymbol{m}}_{\text{mix}}(\Omega)} := \left(\sum_{\boldsymbol{\alpha} \leq \boldsymbol{m}} \|D^{\boldsymbol{\alpha}} f\|^2_{L_2(\Omega)} \right)^{1/2}.$$

The following results on mixed regularity Sobolev spaces are well-known.

LEMMA 3.5.
1. *The space $H^{\boldsymbol{m}}_{\text{mix}}(\Omega)$ is a Hilbert space.*
2. *If $\Omega = \Omega_1 \times \cdots \times \Omega_d$ with open, nonempty intervals $\Omega_j \subseteq \mathbb{R}$, then $H^{\boldsymbol{m}}_{\text{mix}}(\Omega)$ is the tensor product of the univariate Sobolev spaces $H^{m_j}(\Omega_j)$, i.e.*

$$H^{\boldsymbol{m}}_{\text{mix}}(\Omega) = \bigotimes_{j=1}^d H^{m_j}(\Omega_j).$$

The second point can be used to introduce also mixed regularity Sobolev space of non-integer order.

DEFINITION 3.6. *Let $\boldsymbol{\sigma} \in \mathbb{R}^d$ with $\sigma_j \geq 0$, $1 \leq j \leq d$, be given. If $\Omega = \Omega_1 \times \cdots \times \Omega_d$, then the mixed regularity Sobolev space of order σ is defined as*

$$H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) = \bigotimes_{j=1}^d H^{\sigma_j}(\Omega_j).$$

If Ω is $\mathbb{R}^d$, then it is again possible to use the Fourier transform to introduce an equivalent norm.

LEMMA 3.7. *For $f \in H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)$ define*

$$\|f\|_{\widetilde{H}^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)} := \left(\int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{\omega})|^2 \prod_{j=1}^d (1+\omega_j^2)^{\sigma_j} d\boldsymbol{\omega} \right)^{1/2}.$$

Then, this defines an equivalent norm on $H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)$. To be more precise, there is a constant $C_{\boldsymbol{\sigma}} > 0$ such that

$$\|f\|_{H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)} \leq \|f\|_{\widetilde{H}^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)} \leq C_{\boldsymbol{\sigma}} \|f\|_{H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)}.$$

We need the Sobolev embedding theorem and an extension operator for mixed regularity spaces.

PROPOSITION 3.8. *Let $\Omega = \Omega_1 \times \cdots \times \Omega_d$ with open, non-empty intervals Ω_j.*
1. *If $\sigma_j > 1/2$, $1 \leq j \leq d$, then $H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) \subseteq C(\overline{\Omega})$ and the embedding is continuous.*
2. *There is a continuous and linear extension operator $E_\Omega : H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) \to H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)$. The operator itself is* universal *in the sense that it does not depend on $\boldsymbol{\sigma} \in \mathbb{R}^d$ with $\sigma_j \geq 0$.*

Proof. The first part is well-known and essentially follows from the fact that the embedding operators $\iota_j : H^{\sigma_j}(\Omega_j) \to C(\overline{\Omega_j}) \cap L_\infty(\Omega_j)$ are continuous if $\sigma_j > 1/2$ by the standard Sobolev embedding theorem. Standard tensor product theory shows

$$H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) = \bigotimes_{j=1}^d H^{\sigma_j}(\Omega_j), \qquad C(\overline{\Omega}) = \left(\bigotimes C(\overline{\Omega_j}), \|\cdot\|_\vee \right),$$

where the tensor-product norm for $H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega)$ is induced by the inherited inner products and $\|\cdot\|_\vee$ is the so-called *injective norm*, see, for example, [9] for more details. As the former norm is uniformly compatible with the latter norm, this implies that the induced embedding

$$\iota := \iota_1 \otimes \cdots \otimes \iota_d : H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) \to C(\overline{\Omega})$$

is continuous.

The proof of the second part follows along the same lines. As we have for each $1 \le j \le d$ a universal linear and continuous extension operator $E_j : H^{\sigma}(\Omega_j) \to H^{\sigma}(\mathbb{R})$, the arguments outlined above for the embedding also show that the operator

$$E_\Omega := E_1 \otimes \cdots \otimes E_d : H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) \to H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^d)$$

is well-defined and continuous. It is indeed an extension operator. To see this, we first note that we have for any elementary tensor $g_1 \otimes \cdots \otimes g_d \in H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega)$ by definition

$$E_\Omega(g_1 \otimes \cdots \otimes g_d) = (E_1 g_1) \otimes \cdots \otimes (E_d g_d) = g_1 \otimes \cdots \otimes g_d$$

almost everywhere on Ω. This extends by continuity. For every $f \in H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega)$ there is a sequence $(f_n) \subseteq \operatorname{span}\{g_1 \otimes \cdots \otimes g_d : g_j \in H^{\sigma_j}(\Omega_j)\}$ with $\|f - f_n\|_{H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega)} \to 0$. Continuity and linearity of the extension operator thus shows $\|E_\Omega f - E_\Omega f_n\|_{H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^n)} \to 0$. The Lebesgue theory implies that we can find sub-sequences of (f_n) and $E_\Omega f_n$ which converge almost everywhere to f and $E_\Omega f$, respectively. Thus, we have $E_\Omega f = f$ almost everywhere on Ω. ∎

With this result, we are able to derive a version of Theorem 3.3 for mixed regularity Sobolev spaces. We start again by discussing trace and extension theorems.

PROPOSITION 3.9. *Let $\Omega = \Omega_1 \times \cdots \times \Omega_d$ with open, non-empty intervals $\Omega_j \subseteq \mathbb{R}$. For a fixed $1 \le n < d$ let $\widetilde{\Omega}_1 = \Omega_1 \times \cdots \times \Omega_n$ and $\widetilde{\Omega}_2 = \Omega_{n+1} \times \cdots \times \Omega_d$. Let $\boldsymbol{\sigma} \in \mathbb{R}^d$ with $\sigma_j \ge 0$, $1 \le j \le d$. For $\boldsymbol{\omega} \in \Omega$ let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ be the decomposition with $\boldsymbol{\omega}_1 \in \widetilde{\Omega}_1$ and $\boldsymbol{\omega}_2 \in \widetilde{\Omega}_2$. Fix $\boldsymbol{a} \in \Omega$.*

1. *There exists a bounded linear trace operator $T_{\Omega,\widetilde{\Omega}_1} : H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega) \to H^{\boldsymbol{\sigma}}_{\text{mix}}(\widetilde{\Omega}_1)$ satisfying*

$$T_{\Omega,\widetilde{\Omega}_1} f(\boldsymbol{x}) = f(\boldsymbol{x}_1, \boldsymbol{a}_2), \qquad f \in H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega),$$

for almost all $\boldsymbol{x} \in \Omega$.

2. *There exists a bounded linear extension operator $Z_{\widetilde{\Omega}_1,\Omega} : H^{\boldsymbol{\sigma}}_{\text{mix}}(\widetilde{\Omega}_1) \to H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega)$ satisfying*

$$(T_{\Omega,\widetilde{\Omega}_1} \circ Z_{\widetilde{\Omega}_1,\Omega}) g = g, \qquad g \in H^{\boldsymbol{\sigma}}_{\text{mix}}(\widetilde{\Omega}_1).$$

Proof. Again, we restrict ourselves to proving the result for the trace operator and start with the case $\Omega = \mathbb{R}^d$ and $\widetilde{\Omega}_1 = \mathbb{R}^n$. As in the case of standard Sobolev spaces we consider $f \in C_0^\infty(\mathbb{R}^d)$ and employ the Fourier transform representation (3.1) to express the norm. Using also Lemma 3.7, we find

$$\begin{aligned}
&\|f(\cdot, \boldsymbol{a}_2)\|^2_{H^{\boldsymbol{\sigma}}_{\text{mix}}(\mathbb{R}^n)} \\
&\le \int_{\mathbb{R}^n} \prod_{j=1}^n (1+\omega_j^2)^{\sigma_j} \left| (2\pi)^{-(d-n)/2} \int_{\mathbb{R}^{d-n}} \widehat{f}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) e^{i\boldsymbol{\omega}_2^{\mathrm{T}} \boldsymbol{a}_2} d\boldsymbol{\omega}_2 \right|^2 d\boldsymbol{\omega}_1 \\
&= (2\pi)^{n-d} \int_{\mathbb{R}^n} \left[\prod_{j=1}^n (1+\omega_j^2)^{\sigma_j} \int_{\mathbb{R}^{d-n}} |\widehat{f}(\boldsymbol{\omega})|^2 \prod_{j=1}^d (1+\omega_j^2)^{\sigma_j} d\boldsymbol{\omega}_2 \times \right. \\
&\qquad \left. \times \int_{\mathbb{R}^{d-n}} \prod_{j=1}^d (1+\omega_j^2)^{-\sigma_j} d\boldsymbol{\omega}_2 \right] d\boldsymbol{\omega}_1.
\end{aligned}$$

Next, we obviously have

$$\int_{\mathbb{R}^{d-n}} \prod_{j=1}^{d} (1+\omega_j^2)^{-\sigma_j}\, d\boldsymbol{\omega}_2 = \prod_{j=1}^{n} (1+\omega_j^2)^{-\sigma_j} \prod_{j=n+1}^{d} \int_{\mathbb{R}} (1+\omega_j^2)^{-\sigma_j}\, d\omega_j$$
$$= C_{n,d,\boldsymbol{\sigma}} \prod_{j=1}^{n} (1+\omega_j^2)^{-\sigma_j},$$

where the constant $C_{n,d,\boldsymbol{\sigma}}$ can again be computed using (3.2). Inserting this into the above bound yields

$$\|f(\cdot, \boldsymbol{a}_2)\|^2_{H^{\boldsymbol{\sigma}}_{\mathrm{mix}}(\mathbb{R}^n)} \le (2\pi)^{n-d} C_{n,d,\boldsymbol{\sigma}} \int_{\mathbb{R}^n} |\widehat{f}(\boldsymbol{\omega})|^2 \prod_{j=1}^{n} (1+\omega_j^2)^{\sigma_j}\, d\boldsymbol{\omega}$$
$$\le \widetilde{C}_{n,d,\boldsymbol{\sigma}} \|f\|^2_{H^{\boldsymbol{\sigma}}_{\mathrm{mix}}(\mathbb{R}^d)}.$$

Density of the test functions then again shows the result for $\Omega = \mathbb{R}^d$. The case of a general Ω is then again dealt with as in the proof of Proposition 3.2. Of course, this time, we employ the extension operator from Proposition 3.8.
The proof of the existence of a continuous extension operator $Z_{\widetilde{\Omega}_1,\Omega}$ is again left for the reader. □

Having a trace and extension operator at hand, we can proceed exactly as in the case of standard Sobolev spaces.

THEOREM 3.10. *Let $\Omega \subseteq \mathbb{R}^d$ satisfy Assumption 3.1. Let $\boldsymbol{\sigma} \in \mathbb{R}^d$ with $\sigma_j \ge 1/2$, $1 \le j \le d$ and let $\mathfrak{u} \subseteq \mathfrak{D}$.*

1. *For $f \in H^{\boldsymbol{\sigma}}_{\mathrm{mix}}(\Omega)$, the functions $f((\cdot; \boldsymbol{a})_{\mathfrak{u}})$ belong, as functions on $\Omega_{\mathfrak{u}}$, to $H^{\boldsymbol{\sigma}}_{\mathrm{mix}}(\Omega_{\mathfrak{u}})$.*
2. *For $f \in H^{\boldsymbol{\sigma}}_{\mathrm{mix}}(\Omega)$ let $f_{\mathfrak{u}}$, $\mathfrak{u} \subseteq \mathfrak{D}$, be the anchored components (2.6) of the decomposition (2.2). Then, $f_{\mathfrak{u}} \in H^{\boldsymbol{\sigma}}_{\mathrm{mix}}(\Omega_{\mathfrak{u}})$ if considered as a function on $\Omega_{\mathfrak{u}}$.*

4. Sampling inequalities for Sobolev sub-classes. We are now in the position to describe sampling inequalities for function classes which are specific sub-classes of Sobolev or mixed regularity functions and specific point sets adequately tailored for these function classes. In this section, we deal with sub-classes of standard Sobolev functions. The next section is devoted to sub-classes of mixed regularity Sobolev spaces.

4.1. Preliminaries. Sampling inequalities allow us to to determine the general behaviour of a smooth function which is only known on discrete data sites. To describe this in more details. Let $\Omega \subseteq \mathbb{R}^d$ be a given domain of interest and let $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subseteq \Omega$ be a given set of data sites. The *fill distance* and *separation radius* are two geometric quantities, which describe how well X "covers" Ω. They are defined by

$$h_{X,\Omega} := \sup_{\boldsymbol{x}\in\Omega} \min_{1\le j\le N} \|\boldsymbol{x} - \boldsymbol{x}_j\|_2,$$
$$q_X = \frac{1}{2} \min_{j\ne k} \|\boldsymbol{x}_j - \boldsymbol{x}_k\|_2.$$

Sampling inequalities infer information on a function f known only at X. They have the following form and were first introduced in [13, 22].

THEOREM 4.1 (Sampling Inequality). *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary. Let $\sigma > d/2$. Then, there are constants $C > 0$, $h_0 > 0$ such that for all $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subseteq \Omega$ with $h_{X,\Omega} \le h_0$ and all $f \in H^\sigma(\Omega)$ the estimate*

$$\|f\|_{L_\infty(\Omega)} \le C \left[h^{\sigma-d/2}_{X,\Omega} \|f\|_{H^\sigma(\Omega)} + \|f\|_{\ell_\infty(X)} \right]$$

holds.

Sampling inequalities can be used to derive error estimates for stable approximation processes, which employ only function value. For example, if $I_X f \in H^\sigma(\Omega)$ is an interpolant to f on X satisfying also $\|I_X f\|_{H^\sigma(\Omega)} \le C\|f\|_{H^\sigma(\Omega)}$ then, the above sampling inequality yields for $f \in H^\sigma(\Omega)$,

$$\|f - I_X f\|_{L_\infty(\Omega)} \le C h_{X,\Omega}^{\sigma-d/2} \|f - I_X f\|_{H^\sigma(\Omega)} \le C h_{X,\Omega}^{\sigma-d/2} \|f\|_{H^\sigma(\Omega)}.$$

However, estimates based on such standard sampling inequalities inevitably suffer from the curse of dimensionality, as explained in the introduction.

4.2. Sobolev sub-classes.

DEFINITION 4.2. *Let $\Lambda \subseteq \mathcal{P}(d)$ be a downward closed set of subsets of $\mathcal{D} = \{1, \dots, d\}$. Let $H_\Lambda^\sigma(\Omega)$ be the set of all functions $f \in H^\sigma(\Omega)$ such that f has a Λ-representation (2.4).*
For $1 \le n \le d$ let $H_n^\sigma(\Omega)$ be the set of all order-n-functions $f \in H^\sigma(\Omega)$.

It is important to note that if $f \in H_\Lambda^\sigma(\Omega)$ with a downward closed set Λ, its anchored representation also employs only terms with index sets $\mathfrak{u} \in \Lambda$. Hence, we need point sets, which reflect this.

DEFINITION 4.3. *Let $\mathfrak{D} = \{1, \dots, d\}$ and let $\Lambda \subseteq \mathcal{P}(d)$ be downward closed. For each $\mathfrak{u} \in \Lambda$ choose a point set*

$$\widetilde{X}_\mathfrak{u} = \left\{ \widetilde{\boldsymbol{x}}_1^{(\mathfrak{u})}, \dots, \widetilde{\boldsymbol{x}}_{N_\mathfrak{u}}^{(\mathfrak{u})} \right\} \subseteq \Omega_\mathfrak{u}$$

and, using an anchor $\boldsymbol{a} \in \Omega$, extend this point set to the anchored set

$$X_\mathfrak{u} = \left\{ \boldsymbol{x}_1^{(\mathfrak{u})}, \dots, \boldsymbol{x}_{N_\mathfrak{u}}^{(\mathfrak{u})} \right\} \subseteq \Omega,$$

where the components of the points $\boldsymbol{x}_j^{(\mathfrak{u})}$ are defined, as expected, as

$$\boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{x}_j^{(\mathfrak{u})} = \begin{cases} \boldsymbol{e}_k^{\mathrm{T}} \widetilde{\boldsymbol{x}}_j^{(\mathfrak{u})} & \text{if } k \in \mathfrak{u}, \\ \boldsymbol{e}_k^{\mathrm{T}} \boldsymbol{a} & \text{if } k \notin \mathfrak{u}. \end{cases}$$

Here, $\boldsymbol{e}_k$ denotes the k-th unit vector in $\mathbb{R}^d$.
Then, a sampling point set for Sobolev Λ-functions in $\Omega \subseteq \mathbb{R}^d$ *is given by*

$$X_\Lambda^{(d)} = \bigcup_{\mathfrak{u} \in \Lambda} X_\mathfrak{u}.$$

For any $\boldsymbol{x} \in \Omega$, we obviously have

$$\|(\boldsymbol{x}; \boldsymbol{a})_\mathfrak{u} - \boldsymbol{x}_j^{(\mathfrak{u})}\|_2 = \|\boldsymbol{x}_\mathfrak{u} - \widetilde{\boldsymbol{x}}_j^{(\mathfrak{u})}\|_2,$$

which shows in particular that the data set $\widetilde{X}_\mathfrak{u} \subseteq \Omega_\mathfrak{u}$ and the extended set $X_\mathfrak{u}$ have the same separation radius, i.e.

$$q_{\widetilde{X}_\mathfrak{u}} = q_{X_\mathfrak{u}}.$$

Of course, the associated fill distances are quite different. As a matter of fact, we can only expect $\widetilde{X}_\mathfrak{u}$ to fill out $\Omega_\mathfrak{u}$, as $X_\mathfrak{u}$ is restricted to hyper-subsets of Ω and hence can never fill out Ω. However, this will be our advantage as it will suffice for us to fill out the low-dimensional domains.

DEFINITION 4.4. *The separation radius and the fill distance of a sampling point set $X_\Lambda^{(d)} \subseteq \Omega \subseteq \mathbb{R}^d$ of Λ-functions are defined as*

$$q_{X_\Lambda^{(d)}} := \min_{\mathfrak{u} \in \Lambda} q_{X_\mathfrak{u}},$$
$$h_{X_\Lambda^{(d)}} := \max_{\mathfrak{u} \in \Lambda} h_{\widetilde{X}_\mathfrak{u}, \Omega_\mathfrak{u}}.$$

We will call the sampling point set $X_\Lambda^{(d)}$ quasi-uniform *if $q_{X_\Lambda^{(d)}}$ and $h_{X_\Lambda^{(d)}}$ are of comparable size, i.e. if there are constants $c_1, c_2 > 0$ such that*

$$c_1 q_{X_\Lambda^{(d)}} \le h_{X_\Lambda^{(d)}} \le c_2 q_{X_\Lambda^{(d)}}. \tag{4.1}$$

With this, we are able to derive a general sampling inequality for Λ-depending functions.

THEOREM 4.5. *Let $\Omega \subseteq \mathbb{R}^d$ satisfy Assumption 3.1. Let $\Lambda \subseteq \mathcal{P}(d)$ be downward closed and let $X_\Lambda^{(d)} \subseteq \Omega$ be a sampling point set as in Definition 4.3. Then, there is a constant $C_\Lambda > 0$, such that*

$$\|f\|_{L_\infty(\Omega)} \le C_\Lambda \left[h_{X_\Lambda^{(d)}}^{\sigma - d/2} \|f\|_{H^\sigma(\Omega)} + \|f\|_{\ell_\infty(X_\Lambda^{(d)})} \right]$$

for all $f \in H_\Lambda^\sigma(\Omega)$. The constant C_Λ has the form

$$C_\Lambda = \max_{\mathfrak{u}\in\Lambda} C_{\mathfrak{u}} \sum_{\mathfrak{u}\in\Lambda} 2^{\#\mathfrak{u}},$$

where $C_{\mathfrak{u}}$ is the constant from the sampling inequality, Theorem 4.1, when applied to $\Omega_{\mathfrak{u}}$.

Proof. As $f \in H_\Lambda^\sigma(\Omega)$, it has a representation

$$f = \sum_{\mathfrak{u}\in\Lambda} f_{\mathfrak{u}}.$$

By Corollary 2.5, we may assume that the components $f_{\mathfrak{u}}$ are given in anchored form (2.6), i.e. by

$$f_{\mathfrak{u}}(\boldsymbol{x}) = \sum_{\mathfrak{v}\subseteq\mathfrak{u}} (-1)^{\#\mathfrak{u}-\#\mathfrak{v}} f((\boldsymbol{x};\boldsymbol{a})_{\mathfrak{v}})$$

By Theorem 3.3, these components satisfy $f((\cdot;\boldsymbol{a})_{\mathfrak{v}}) \in H^{\sigma-(d-n)/2}(\Omega_{\mathfrak{v}})$ with $n = \#\mathfrak{v}$ when considered as functions on $\Omega_{\mathfrak{v}}$. Moreover, by definitions of the anchor and the point sets we have

$$\|f((\cdot;\boldsymbol{a})_{\mathfrak{v}}\|_{L_\infty(\Omega_{\mathfrak{v}})} \le \|f\|_{L_\infty(\Omega)}, \qquad \|f((\cdot;\boldsymbol{a})_{\mathfrak{v}})\|_{\ell_\infty(\widetilde{X}_{\mathfrak{v}})} \le \|f\|_{\ell_\infty(X_\Lambda^{(d)})}.$$

Thus, we can employ the sampling theorem, Theorem 4.1, for each of these components, yielding

$$\begin{aligned}
\|f\|_{L_\infty(\Omega)} &\le \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} \|f((\cdot;\boldsymbol{a})_{\mathfrak{v}})\|_{L_\infty(\Omega_{\mathfrak{v}})} \\
&\le \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} C_{\mathfrak{v}} \left[h_{\widetilde{X}_{\mathfrak{v}},\Omega_{\mathfrak{v}}}^{\sigma-\frac{d-\#\mathfrak{v}}{2}-\frac{\#\mathfrak{v}}{2}} \|f((\cdot;\boldsymbol{a})_{\mathfrak{v}})\|_{H^{\sigma-\frac{d-\#\mathfrak{v}}{2}}(\Omega_{\mathfrak{v}})} + \|f((\cdot;\boldsymbol{a})_{\mathfrak{v}})\|_{\ell_\infty(\widetilde{X}_{\mathfrak{v}})} \right] \\
&\le \left(\max_{\mathfrak{u}\in\Lambda} C_{\mathfrak{u}}\right) \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} \left[h_{X_\Lambda^{(d)}}^{\sigma-\frac{d}{2}} \|f\|_{H^\sigma(\Omega)} + \|f\|_{\ell_\infty(X_\Lambda^{(d)})} \right].
\end{aligned}$$

As the number of subsets $\mathfrak{v} \subseteq \mathfrak{u}$ is given by $2^{\#\mathfrak{u}}$, this gives the stated constant. ∎

The constant C_Λ is not particular sharp, as it ignores the fact that several terms $f(\cdot;\boldsymbol{a})_{\mathfrak{v}}$ from different $\mathfrak{u} \supseteq \mathfrak{v}$ appear in the representation of f with possible different signs. Making use of these combinations might give a sharper bound. Nonetheless, this constant even works for very high-dimensional functions f as long as the size of each set $\mathfrak{u} \in \Lambda$ can be bounded by a constant n independent of d. In this situation, we have the obvious bound

$$C_\Lambda \le 2^n \#\Lambda \max_{\mathfrak{u}\in\Lambda} C_{\mathfrak{u}}. \tag{4.2}$$

The constant might also depend on the space dimension d via the set Λ. However, the "worst" case in this setting is given by order n-functions, i.e. by the set $\Lambda = \{\mathfrak{u} \subseteq \mathfrak{D} : \#\mathfrak{u} \leq n\}$. Here, using (2.5),

$$C_\Lambda \leq 2^n \sum_{j=0}^{n} \binom{d}{j} \max_{\#\mathfrak{u}\leq n} C_\mathfrak{u} \leq \left(\frac{2ed}{n}\right)^n \max_{\#\mathfrak{u}\leq n} C_\mathfrak{u}.$$

This bound already has the right asymptotic behaviour $(2d)^n$. However, it can slightly be improved using the representation (2.7). In this situation, we will write $X_n^{(d)}$ instead of $X_\Lambda^{(d)}$ and can derive the following more accurate bound.

THEOREM 4.6. *Let $\Omega \subseteq \mathbb{R}^d$ satisfy Assumption 3.1. Let $1 \leq n < d$ and let $X_n^{(d)} \subseteq \Omega$ be a sampling point set as in Definition 4.3. Then, there is a constant $C_n > 0$, not depending on d, such that*

$$\|f\|_{L_\infty(\Omega)} \leq C_n \frac{(2d)^n}{n!} \left[h_{X_n^{(d)}}^{\sigma-d/2} \|f\|_{H^\sigma(\Omega)} + \|f\|_{\ell_\infty(X_n^{(d)})} \right]$$

for all $f \in H_n^\sigma(\Omega)$.

Proof. As in the proof of Theorem 4.5 we may assume, by Corollary 2.8, that $f \in H_n^\sigma(\Omega)$ has a representation with components $f_\mathfrak{u}$ in anchored form (2.6). However, for order-n-functions, the components may be written as in (2.7), i.e. we have

$$f = \sum_{j=0}^{n} (-1)^{n-j} \binom{d-1-j}{n-j} \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D} \\ \#\mathfrak{u}=j}} f((\cdot;\boldsymbol{a})_\mathfrak{u}). \tag{4.3}$$

With this representation, we can now proceed as in the proof of Theorem 4.5 and derive the bound

$$\begin{aligned} \|f\|_{L_\infty(\Omega)} &\leq \sum_{j=0}^{n} \binom{d-1-j}{n-j} \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D} \\ \#\mathfrak{u}=j}} \|f((\cdot;\boldsymbol{a})_\mathfrak{u}\|_{L_\infty(\Omega_\mathfrak{u})} \\ &\leq C_n \sum_{j=0}^{n} \binom{d-1-j}{n-j} \binom{d}{j} \left[h_{X_n^{(d)}}^{\sigma-\frac{d}{2}} \|f\|_{H^\sigma(\Omega)} + \|f\|_{\ell_\infty(X_n^{(d)})} \right]. \end{aligned}$$

From this, the stated inequality follows if we also use

$$\sum_{j=0}^{n} \binom{d-1-j}{n-j} \binom{d}{j} \leq \sum_{j=0}^{n} \binom{d-1}{n-j} \binom{d}{j} = \binom{2d-1}{n} \leq \frac{(2d)^n}{n!}, \tag{4.4}$$

where the identity comes from [7, 0.156 formula 1]. ∎

The bound $(2d)^n/n!$ is indeed sharper than the general bound $(2ed/n)^n$. This follows immediately from *Stirling's formula*

$$1 \leq \frac{n!}{\sqrt{2\pi n} n^n e^{-n}}.$$

4.3. Curse of Dimensionality. In this section, we will use the above sampling inequality to show that there is no curse of dimensionality for functions from $H_n^\sigma(\Omega)$ with $\sigma > d/2$.

We will achieve this by constructing an interpolant using a positive definite kernel. To this end, recall that a Hilbert space $\mathcal{H}$ of functions $f : \Omega \to \mathbb{R}$ is called a *reproducing kernel Hilbert space* if there is a unique function $K : \Omega \times \Omega \to \mathbb{R}$ with $K(\cdot, \boldsymbol{x}) \in \mathcal{H}$ for every $\boldsymbol{x} \in \Omega$ and

$$f(\boldsymbol{x}) = \langle f, K(\cdot, \boldsymbol{x}) \rangle, \qquad \boldsymbol{x} \in \Omega, \quad f \in \mathcal{H}.$$

It is well known that such a reproducing kernel is always positive semi-definite in the sense that for any point set $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subseteq \Omega$ of distinct points, the matrices $A = (K(\boldsymbol{x}_i, \boldsymbol{x}_j)) \in \mathbb{R}^{N\times N}$ are positive semi-definite. The kernel is even positive definite, i.e. all the matrices are positive definite, if point-evaluations are linearly independent over $\mathcal{H}$, see for example [21].
Next, recall that $H^\sigma(\Omega)$ with $\sigma > d/2$ is itself a reproducing kernel Hilbert space due to the Sobolev embedding theorem. However, we cannot use the reproducing kernel of $H^\sigma(\Omega)$ to form the corresponding interpolant, as the latter would not be contained in $H^\sigma_n(\Omega)$. We can, however, use the fact that $H^\sigma_n(\Omega)$ is a closed subspace of $H^\sigma(\Omega)$ and hence a reproducing kernel Hilbert space itself. We will discuss this in the following, more general setting.

THEOREM 4.7. *Let $\Omega \subseteq \mathbb{R}^d$ satisfy Assumption 3.1. Let $\Lambda \subseteq \mathcal{P}(d)$ be a downward closed set and let $\sigma > d/2$. Then, $H^\sigma_\Lambda(\Omega)$ is a closed subspace of $H^\sigma(\Omega)$ and hence a reproducing kernel Hilbert space. If $K_\sigma : \Omega \times \Omega \to \mathbb{R}$ is the reproducing kernel of $H^\sigma(\Omega)$ and $P : H^\sigma(\Omega) \to H^\sigma_\Lambda(\Omega)$ is the orthogonal projection, then the reproducing kernel $K_{\sigma,\Lambda} : \Omega \times \Omega \to \mathbb{R}$ of $H^\sigma_\Lambda(\Omega)$ is given by*

$$K_{\sigma,\Lambda}(\cdot, \boldsymbol{x}) = PK_\sigma(\cdot, \boldsymbol{x}) \qquad \boldsymbol{x} \in \Omega.$$

The kernel is positive definite.

Proof. Any $f \in H^\sigma_\Lambda(\Omega)$ has a representation

$$f = \sum_{\mathfrak{u}\in\Lambda} \sum_{\mathfrak{v}\subseteq\mathfrak{u}} (-1)^{\#\mathfrak{u}-\#\mathfrak{v}} f((\cdot; \boldsymbol{a})_\mathfrak{v}). \tag{4.5}$$

By Proposition 3.2 and Theorem 3.3, there are constants $C_\mathfrak{v} > 0$ for $\mathfrak{v} \subseteq \mathfrak{D}$ such that

$$\|f(\cdot; \boldsymbol{a})_\mathfrak{v}\|_{H^{\sigma-(d-\#\mathfrak{v})/2}(\Omega_\mathfrak{v})} \le C_\mathfrak{v}\|f\|_{H^\sigma(\Omega)}. \tag{4.6}$$

If $(f_k) \subseteq H^\sigma_\Lambda(\Omega)$ is a convergent sequence with limit $f \in H^\sigma(\Omega)$ we need to show that $f \in H^\sigma_\Lambda(\Omega)$. From (4.6), we see that the components $(f_n(\cdot; \boldsymbol{a})_\mathfrak{v})$ converge in $H^{\sigma-(d-\#\mathfrak{v})/2}(\Omega)$ to $f((\cdot; \boldsymbol{a})_\mathfrak{v})$. Now, assume that the limit function has a non-vanishing component $f((\cdot; \boldsymbol{a})_\mathfrak{v})$ with $\mathfrak{v} \notin \Lambda$, i.e., if we set $\tau := \sigma - (d - \#\mathfrak{v})/2$, with $\|f((\cdot; \boldsymbol{a})_\mathfrak{v})\|_{H^\tau(\Omega_\mathfrak{v})} > 0$. Using (4.6) again, this gives the contradiction

$$\begin{aligned}
0 &= \|f_n((\cdot; \boldsymbol{a})_\mathfrak{v})\|_{H^\tau(\Omega_\mathfrak{v})}\\
&\ge \|f((\cdot; \boldsymbol{a})_\mathfrak{v})\|_{H^\tau(\Omega_\mathfrak{v})} - \|f((\cdot; \boldsymbol{a})_\mathfrak{v}) - f_n((\cdot; \boldsymbol{a})_\mathfrak{v})\|_{H^\tau(\Omega_\mathfrak{v})}\\
&\ge \|f((\cdot; \boldsymbol{a})_\mathfrak{v})\|_{H^\tau(\Omega_\mathfrak{v})} - C_\mathfrak{v}\|f - f_n\|_{H^\sigma(\Omega)}\\
&\ge \frac{1}{2}\|f((\cdot; \boldsymbol{a})_\mathfrak{v})\|_{H^\tau(\Omega_\mathfrak{v})}\\
&> 0,
\end{aligned}$$

which holds for all sufficiently large n. Thus, the limit function f has indeed a representation of the form (4.5). The other statements follow directly from standard reproducing kernel Hilbert space theory. First of all, as a closed subspace, $H^\sigma_\Lambda(\Omega)$ is indeed a reproducing kernel Hilbert space with the stated reproducing kernel, see for example [2] and particularly [20, Theorem 1 (d)]. Finally, to see that the kernel is positive definite and not only positive semi-definite, we can use standard arguments employing bump functions to show that point evaluation functionals are indeed linearly independent. Details can, for example, be found in [6] for standard Sobolev spaces. The techniques carry immediately over to this case by employing projections. ∎

The next thing we need is a bound on the number of points in a set $X^{(d)}_n$ using the separation radius and the fill distance. To this end, recall that for any $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subseteq \Omega \subseteq \mathbb{R}^d$, the open balls

$B_d(\boldsymbol{x}_j, q_X) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{x}_j\|_2 < q_X\}$ about $\boldsymbol{x}_j \in X$ with radius q_X are disjoint. Thus, for a bounded $B_d(\mathbf{0}, r) \subseteq \Omega \subseteq B_d(\mathbf{0}, R)$, a volume argument shows

$$h_{X,\Omega}^{-d} r^d \leq N \leq q_X^{-d}(R+1)^d,$$

where the last estimate holds for $q_X \leq 1$.

PROPOSITION 4.8. *Let $1 \leq n \leq d$ be fixed. Assume there are $R \geq r > 0$ such that $B_d(\mathbf{0}, r) \subseteq \Omega \subseteq B_d(\mathbf{0}, R)$. Let $X_n^{(d)} \subseteq \Omega$ be a point set as in Definition 4.3. The number $N = \#X_n^{(d)}$ of points in this set is bounded by*

$$\left(\frac{d}{n}\right)^n r^n h_{X_n^{(d)}}^{-n} \leq N \leq \left(\frac{ed}{n}\right)^n (R+1)^n q_{X_n^{(d)}}^{-n}.$$

Proof. First of all $\Omega \subseteq B_d(\mathbf{0}, R)$ implies $\Omega_\emptyset \subseteq B_1(\mathbf{0}, R)$ and $\Omega_\mathfrak{u} \subseteq B_{\#\mathfrak{u}}(\mathbf{0}, R)$ for $\emptyset \neq \mathfrak{u} \subseteq \mathfrak{D}$. Hence, the above considerations show with $q := q_{X_n^{(d)}}$,

$$\begin{aligned}
N &= \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D}\\ \#\mathfrak{u}\leq n}} \#\widetilde{X}_\mathfrak{u} = \sum_{k=0}^{n} \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D}\\ \#\mathfrak{u}=k}} \#\widetilde{X}_\mathfrak{u} \\
&\leq q_{\widetilde{X}_\emptyset}^{-1}(R+1) + \sum_{k=1}^{n} \binom{d}{k} q_{\widetilde{X}_\mathfrak{u}}^{-k}(R+1)^k \\
&\leq q^{-n}(R+1)^n \sum_{k=0}^{n} \binom{d}{k} \\
&\leq q^{-n}(R+1)^n \left(\frac{ed}{n}\right)^n,
\end{aligned}$$

using (2.5). In the same way, we can establish the lower bound. This time, we have with $h = h_{X_n^{(d)}}$,

$$\begin{aligned}
N &= \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D}\\ \#\mathfrak{u}\leq n}} \#\widetilde{X}_\mathfrak{u} = \sum_{k=0}^{n} \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D}\\ \#\mathfrak{u}=k}} \#\widetilde{X}_\mathfrak{u} \geq h^{-1} r + \sum_{k=1}^{n} \binom{d}{k} h^{-k} r^k \\
&\geq \binom{d}{n} h^{-n} r^n \geq \left(\frac{d}{n}\right)^n h^{-n} r^n.
\end{aligned}$$

Having these two ingredients, the reproducing kernel Hilbert space set-up and the bounds on the number of points, we can now show that the curse of dimensionality does not apply to $H_n^\sigma(\Omega)$.

If $K_{\sigma,n} : \Omega \times \Omega \to \mathbb{R}$ is the reproducing kernel and $X_n^{(d)} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ is our sampling point set then the matrices

$$A_{X_n^{(d)}} = (K_{\sigma,n}(\boldsymbol{x}_i, \boldsymbol{x}_j)) \in \mathbb{R}^{N \times N}$$

are symmetric and positive definite. Hence, we can compute an interpolant

$$s_f = \sum_{j=1}^{N} \alpha_j K_{\sigma,n}(\cdot, \boldsymbol{x}_j) \tag{4.7}$$

in $\mathcal{O}(N^3)$ time. As a matter of fact, there are numerical techniques, which allow us to reduce this cost significantly but as this is right now not important, we will leave this for future research. This interpolant is known to satisfy the stability condition

$$\|f - s_f\|_{H^\sigma(\Omega)} \leq \|f\|_{H^\sigma(\Omega)}, \qquad f \in H_n^\sigma(\Omega),$$

so that the sampling inequality leads to the error estimate

$$\|f - s_f\|_{L_\infty(\Omega)} \le C_n \frac{(2d)^n}{n!} h_{X_n^{(d)}}^{\sigma-d/2} \|f\|_{H^\sigma(\Omega)}. \tag{4.8}$$

THEOREM 4.9. *Let $\Omega \subseteq \mathbb{R}^d$ satisfy Assumption 3.1. Assume there are $R \ge r > 0$ such that $B_d(\mathbf{0}, r) \subseteq \Omega \subseteq B_d(\mathbf{0}, R)$. Let $1 \le n < d$ and let $X_n^{(d)} \subseteq \Omega$ be a quasi-uniform sampling point set as in Definition 4.3 with $h_{\widetilde{X}_\mathfrak{u}, \Omega_\mathfrak{u}} \le \min\{1, r\}$ for all $\mathfrak{u} \subseteq \mathfrak{D}$. Let $\sigma > d/2$ and $\tau := \sigma - d/2$. Let s_f be the interpolant to $f \in H_n^\sigma(\Omega)$ using the reproducing kernel of $H_n^\sigma(\Omega)$ and the set $X_n^{(d)}$. Then,*

$$\|f - s_f\|_{L_\infty(\Omega)} \le C_{n,\tau} d^{2n+\tau} N^{-\tau/n} \|f\|_{H^\sigma(\Omega)}.$$

with $C_{n,\tau} = \frac{C_n 2^n}{n!} c_2^\tau (e/n)^\tau (R+1)^\tau$, where C_n is the constant from Theorem 4.6 and c_2 is the constant from (4.1).
Hence, if $\|f\|_{H^\sigma(\Omega)} = 1$ and $\epsilon > 0$ are given,

$$N \ge C_{n,\tau}^{n/\tau} d^{(n+\tau)n/\tau} \epsilon^{-n/\tau} \tag{4.9}$$

points are required to achieve accuracy ϵ.

Proof. Writing again $q = q_{X_n^{(d)}}$ and $h = h_{X_n^{(d)}}$. From Proposition 4.8 we have

$$q \le N^{-1/n} \frac{ed}{n} (R+1).$$

Using this and (4.1) in (4.8) yields

$$\begin{aligned} \|f - s_f\|_{L_\infty(\Omega)} &\le C_n \frac{(2d)^n}{n!} c_2^\tau q^\tau \|f\|_{H^\sigma(\Omega)} \\ &\le C_n \frac{(2d)^n}{n!} c_2^\tau N^{-\tau/n} \left(\frac{ed}{n}\right)^\tau (R+1)^\tau \|f\|_{H^\sigma(\Omega)} \\ &= C_{n,\tau} d^{n+\tau} N^{-\tau/n} \|f\|_{H^\sigma(\Omega)}. \end{aligned}$$

The estimate on the required number of points N to achieve an accuracy ϵ follows immediately from this. ∎

We now want to discuss the constant in some more detail with regard to the curse of dimensionality. To study the latter, we note that we need to keep $\tau := \sigma - d/2$ fixed when looking at $d \to \infty$. Then, the constant

$$C_{n,\tau}^{n/\tau} = C_n^{n/\tau} 2^{n^2/\tau} c_2^n \left(\frac{e}{n}\right)^n (R+1)^n (n!)^{-n/\tau}$$

is indeed independent of the space dimension d. This means, the bound in (4.9) is only polynomial in the space dimension d. However, to conclude that the curse of dimensionality does not apply for functions in $H_n^\sigma(\Omega)$, we also need to address the topic of "sufficiently small" h introduced at the beginning of this paper.
For each point set $\widetilde{X}_\mathfrak{u} \subseteq \Omega_\mathfrak{u}$, we have the assumption

$$h_{\widetilde{X}_\mathfrak{u}, \Omega_\mathfrak{u}} \le h_{0,\mathfrak{u}} = \frac{c_\mathfrak{u}}{(\sigma - (d - \#\mathfrak{u})/2)^2},$$

or the coarser assumption

$$h_{\widetilde{X}_\mathfrak{u}, \Omega_\mathfrak{u}} \le \widetilde{h}_{0,\mathfrak{u}} := \frac{c_\mathfrak{u}}{\sigma^2}.$$

In the latter case, assuming also quasi-uniformity for each point set $X_{\mathfrak{u}}$, we have

$$\#\widetilde{X}_{\mathfrak{u}} \approx h_{\widetilde{X}_{\mathfrak{u}},\Omega_{\mathfrak{u}}}^{-\#\mathfrak{u}} \geq \sigma^{2\#\mathfrak{u}}/c_{\mathfrak{u}}^{\#\mathfrak{u}}$$

and thus we need, up to constants independent of d,

$$N \geq \sum_{k=0}^{n} \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D} \\ \#\mathfrak{u}=k}} \#\widetilde{X}_{\mathfrak{u}} = \sum_{k=0}^{n} \sigma^{2k} \sum_{\substack{\mathfrak{u}\subseteq\mathfrak{D} \\ \#\mathfrak{u}=k}} c_{\mathfrak{u}}^{-\#\mathfrak{u}},$$

which is satisfied if, with $C = \max c_{\mathfrak{u}}^{-\#\mathfrak{u}}$,

$$N \geq C \sum_{k=0}^{n} \sigma^{2k} \binom{d}{k}.$$

As seen several times before, this lower bound is also only polynomial in d.

COROLLARY 4.10. *For $H_n^\sigma(\Omega)$ with $\sigma > d/2$ and a fixed $n < d$, the curse of dimensionality does not apply.*

5. Sampling inequalities for mixed regularity Sobolev sub-classes. We will now turn to subsets of mixed regularity Sobolev spaces. While the theory will work for subsets of $H^{\boldsymbol{\sigma}}_{\text{mix}}(\Omega)$, we will make the simplification that all components σ_j of $\boldsymbol{\sigma}$ are the same. Hence, throughout this section, we will consider subsets of

$$H^\sigma_{\text{mix}}(\Omega) = \bigotimes_{j=1}^{d} H^\sigma(\Omega_j).$$

These subsets are the corresponding subsets of standard Sobolev spaces.

DEFINITION 5.1. *Let $\Lambda \subseteq \mathcal{P}(d)$ be a downward closed set of subsets of $\mathcal{D}$. Let $H^\sigma_{\text{mix},\Lambda}(\Omega)$ be the set of all functions $f \in H^\sigma_{\text{mix}}(\Omega)$ having a Λ-representation (2.4).*
For $1 \leq n \leq d$ let $H^\sigma_{\text{mix},n}(\Omega)$ be the set of all order-n-functions $f \in H^\sigma_{\text{mix}}(\Omega)$.

The procedure in this section is very similar to the one in the last section. We first define sampling sets which are well-suited for these function classes.

5.1. Sampling inequalities. Mixed regularity Sobolev spaces are naturally connected to *sparse grids*. Hence, we will use sparse grids to build the higher-dimensional sets in $\Omega_{\mathfrak{u}}$, which are then extended to the high-dimensional point sets in Ω. As usual in this context, we consider *Clenshaw-Curtis* points. This also means that we will restrict ourselves to

$$\Omega_j = [-1, 1], \qquad 1 \leq j \leq d,$$

though other intervals can obviously simply be treated by linearly transforming them to $[-1, 1]$. In this context isotropic sparse grids are constructed as follows. They involve an integer $q \in \mathbb{N}$ with $q \geq d$. This integer has nothing to do with the separation radius of the last section. Instead, it indicates the level of the finest univariate grid used in the construction of the sparse grid.

DEFINITION 5.2. *Let $\Omega_j = [-1, 1]$, $1 \leq j \leq d$. Let $n_1 = 1$ and $n_j = 2^{j-1} + 1$ for $j \geq 2$. Let $\Lambda \subseteq \mathcal{P}(d)$ be downward closed.*

1. *The* Clenshaw-Curtis points *or* Chebyshev points of the second kind *are the extremal points of the Chebyshev polynomials given by $Y_1 = \{0\}$ and*

$$Y_j = \left\{ x_i^{(j)} = -\cos\left(\pi \frac{i-1}{n_j - 1}\right) : 1 \leq i \leq m_j \right\}, \qquad j > 1.$$

2. *For $\emptyset \neq \mathfrak{u} \subseteq \mathfrak{D}$ and $q \in \mathbb{N}$ with $q \geq n := \#\mathfrak{u}$, the* sparse grid $\widetilde{X}_{\mathfrak{u},q} \subseteq \Omega_{\mathfrak{u}}$ *based on the points from $\{Y_j\}$ is defined as*

$$\widetilde{X}_{\mathfrak{u},q} = \bigcup_{\substack{\boldsymbol{i}\in\mathbb{N}^n \\ |\boldsymbol{i}|=q}} Y_{\mathfrak{u}_{i_1}} \times \cdots \times Y_{\mathfrak{u}_{i_n}}.$$

   In the case of $\mathfrak{u} = \emptyset$ we set $\widetilde{X}_{\emptyset,q} = Y_q$. These point sets are extended using the anchor $\boldsymbol{a} \in \Omega$ to a point set $X_{\mathfrak{u},q}$ as in Definition 4.3.
3. *For $\boldsymbol{q} = \{q_{\mathfrak{u}} : \mathfrak{u} \in \Lambda\}$ with $q_{\mathfrak{u}} \geq \#\mathfrak{u}$, a* sampling point set for mixed regularity Sobolev Λ-functions *is given by*

$$X^{(d)}_{\Lambda,\boldsymbol{q}} = \bigcup_{\mathfrak{u}\in\Lambda} X_{\mathfrak{u},q_{\mathfrak{u}}}.$$

As in the last section, we need a sampling inequality for lower-dimensional domains and, this time, sparse grids. A close inspection of the proof of the *non-oversampling* case in [17] shows that [17, Theorem 9] can more accurately and improvedly be formulated as.

THEOREM 5.3. *Let $\Omega = [-1,1]^d$. Let $\sigma > 1/2$ and $q \geq d$. Then, for $f \in H^{\sigma}_{\mathrm{mix}}(\Omega)$ and sampling point sets $\widetilde{X}_{\mathcal{D},q}$ with $N = \#\widetilde{X}_{\mathcal{D},q}$ points, the bounds*

$$\begin{aligned}\|f\|_{L_\infty(\Omega)} &\leq C\left[q^{3d-2}2^{-(\sigma-1/2)(q-d)}\|f\|_{H^{\sigma}_{\mathrm{mix}}(\Omega)} + q^{2d-1}\|f\|_{\ell_\infty(\widetilde{X}_{\mathcal{D},q})}\right] \\ &\leq C\left[(\log N)^{\rho_1(\sigma,d)}N^{-\sigma+1/2}\|f\|_{H^{\sigma}_{\mathrm{mix}}(\Omega)} + (\log N)^{\rho_2(d)}\|f\|_{\ell_\infty(\widetilde{X}_{\mathcal{D},q})}\right]\end{aligned}$$

holds with $\rho_1(\sigma,d) = (\sigma+5/2)(d-1)+1$ and $\rho_2(d) = 2d-1$. The constant $C > 0$ depends only on d and σ.

The exponents ρ_1 and ρ_2 might for specific approximation processes not be optimal but suffice for our purposes. If *oversampling* is used, it is possible to improve the exponents further, up to the optimal case $\rho_2 = 0$. However, there is always a non-zero exponent ρ_1.
We are now in the position to formulate the result corresponding to Theorem 4.5 for mixed regularity Sobolev spaces.

THEOREM 5.4. *Let $\Omega = [-1,1]^d$. Let $\sigma > 1/2$. Let $\Lambda \subseteq \mathcal{P}(d)$ be downward closed and $n = \max_{\mathfrak{u}\in\Lambda}\#\mathfrak{u}$. Assume there is a $q \in \mathbb{N}$ such that the elements of $\boldsymbol{q} = \{q_{\mathfrak{u}} : \mathfrak{u} \in \Lambda\}$ have the form $q_{\mathfrak{u}} = \#\mathfrak{u} + q$. Let $X^{(d)}_{\Lambda,\boldsymbol{q}} \subseteq \Omega$ be a sampling point set as in Definition 5.2. Then, there is a constant $C_{\Lambda,n} > 0$, such that*

$$\|f\|_{L_\infty(\Omega)} \leq C_{\Lambda,n}\left[q^{3n-2}2^{-q(\sigma-1/2)}\|f\|_{H^{\sigma}_{\mathrm{mix}}(\Omega)} + q^{2n-1}\|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right]$$

for all $f \in H^{\sigma}_{\mathrm{mix},\Lambda}(\Omega)$. The constant $C_{\Lambda,n}$ has the form

$$C_{\Lambda,n} = \max_{\mathfrak{u}\in\Lambda} C_{\mathfrak{u}} \sum_{\mathfrak{u}\in\Lambda} 2^{\#\mathfrak{u}}(1+n)^{3n-2},$$

where $C_{\mathfrak{u}}$ is the constant from the sampling inequality, Theorem 5.3, when applied to $\Omega_{\mathfrak{u}}$. In the case of $\Lambda = \{\mathfrak{u} \subseteq \mathfrak{D} : \#\mathfrak{u} \leq n\}$, i.e., in the case of order-n-functions, the constant $C_{\Lambda,n}$ becomes

$$C_{\Lambda,n} = \max_{\#\mathfrak{u}\leq n} C_{\mathfrak{u}}\frac{(2d)^n}{n!}.$$

Proof. Let us set $D := 2^{-(\sigma-1/2)}$. With this, we have

$$\begin{aligned}
\|f\|_{L_\infty(\Omega)} &\le \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} \|f((\cdot;\boldsymbol{a})_{\mathfrak{v}})\|_{L_\infty(\Omega_{\mathfrak{v}})} \\
&\le \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} C_{\mathfrak{v}} \left[q_{\mathfrak{v}}^{3\#\mathfrak{v}-2} D^{q_{\mathfrak{v}}-\#\mathfrak{v}} \|f((\cdot;\boldsymbol{a})_{\mathfrak{v}})\|_{L_\infty(\Omega_{\mathfrak{v}})} + q_{\mathfrak{v}}^{2\#\mathfrak{v}-1}\|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right] \\
&\le \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} C_{\mathfrak{v}} \left[(q+\#\mathfrak{v})^{3\#\mathfrak{v}-2} D^{q} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + (q+\#\mathfrak{v})^{2\#\mathfrak{v}-1}\|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right] \\
&\le C_\Lambda \left[(q+n)^{3n-2} D^{q} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + (q+n)^{2n-1}\|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right] \\
&\le C_{\Lambda,n} \left[q^{3n-2} D^{q} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + q^{2n-1}\|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right]
\end{aligned}$$

with $C_\Lambda = \sum_{\mathfrak{u}\in\Lambda}\sum_{\mathfrak{v}\subseteq\mathfrak{u}} C_{\mathfrak{v}}$ and $C_{\Lambda,n} = C_\Lambda(1+n)^{3n-2}$. In the last step we used the bound

$$(q+n)^k = \sum_{j=0}^{k}\binom{k}{j} q^j n^{k-j} \le q^k \sum_{j=0}^{n}\binom{k}{j} n^{k-j} = q^k(1+n)^k. \tag{5.1}$$

The explicit constant in the case of order-n-functions follows as in the proof of Theorem 4.6. In particular, we can use the representation (4.3) and the estimate (4.4), which then easily yields the desired bound. ∎

5.2. Curse of dimensionality. Next, we want to express these bounds in the number of points of the sampling set $X^{(d)}_{\Lambda,\boldsymbol{q}}$. To this end, we will use the estimate

$$2^{q_{\mathfrak{u}}-2\#\mathfrak{u}+1} \le \#X_{\mathfrak{u},q_{\mathfrak{u}}} \le 2^{q_{\mathfrak{u}}-\#\mathfrak{u}+1}\frac{q_{\mathfrak{u}}^{\#\mathfrak{u}-1}}{(\#\mathfrak{u}-1)!}, \qquad \emptyset \ne \mathfrak{u} \subseteq \mathfrak{D}, \tag{5.2}$$

which can be found in [17, Corollary 1] and is a coarse bound on the asymptotic behaviour described in (1.4). For $\mathfrak{u} = \emptyset$ we have a univariate grid with m_q points.

PROPOSITION 5.5. *Let $\Lambda \subseteq \mathcal{P}(d)$ be downward closed with $\#\Lambda \ge 2$ and $n = \max_{\mathfrak{u}\in\Lambda}\#\mathfrak{u}$. Assume there is a $q \in \mathbb{N}$ such that the elements of $\boldsymbol{q} = \{q_{\mathfrak{u}} : \mathfrak{u} \in \Lambda\}$ have the form $q_{\mathfrak{u}} = \#\mathfrak{u} + q$. Then, the number N of points in the sampling point set $X^{(d)}_{\Lambda,\boldsymbol{q}} \subseteq \Omega$ satisfies*

$$2^{q-n}\#\Lambda \le N \le 2^{q+1} q^{n-1}(1+n)^{n-1}\#\Lambda.$$

Proof. The upper bound follows with the upper bound from (5.2). We have

$$\begin{aligned}
N &\le \sum_{\emptyset\ne\mathfrak{u}\in\Lambda} 2^{q+1}\frac{(q+\#\mathfrak{u})^{\#\mathfrak{u}-1}}{(\#\mathfrak{u}-1)!} + m_q \\
&\le 2^{q+1}q^{n-1}(1+n)^{n-1}(\#\Lambda-1) + 2^{q-1} + 1 \\
&\le 2^{q+1}q^{n-1}(1-n)^{n-1}\#\Lambda,
\end{aligned}$$

where we also used (5.1). The lower bound follows in the same way. Here, we have

$$N \ge \sum_{\emptyset\ne\mathfrak{u}\in\Lambda} 2^{q-\#\mathfrak{u}+1} + m_q \ge 2^{q-n+1}(\#\Lambda-1) \ge 2^{q-n}\#\Lambda.$$

∎

In the case of order-n-functions, which we will further pursue in the rest of this section, this becomes

$$\left(\frac{d}{n}\right)^n 2^{q-n} \leq N \leq 2^{q+1} q^{n-1} (1+n)^{n-1} \left(\frac{ed}{n}\right)^n, \tag{5.3}$$

using also Lemma 2.9. As, in contrast to standard sparse grids, the number q does not depend on the space dimension d but only on n, the number of points in our sampling point set grows only polynomially in the space dimension.
Next, let us set once again $\tau := \sigma - 1/2$. From the right-hand side of (5.3) we can conclude

$$2^{-\tau q} \leq N^{-\tau} 2^\tau q^{\tau(n-1)} (1+n)^{\tau(n-1)} \left(\frac{ed}{n}\right)^n =: c_{\tau,n} N^{-\tau} q^{\tau(n-1)} d^{\tau n} \tag{5.4}$$

From the left-hand side of (5.3) we can conclude

$$n \log\left(\frac{d}{n}\right) + (q-n)\log 2 = q \log 2 + n \left[\log\left(\frac{d}{n}\right) - \log 2\right] \leq \log N,$$

which gives for $d \geq 2n$ the bound

$$q \leq \frac{1}{\log 2} \log N. \tag{5.5}$$

These two bounds allow us now to express the error bound of Theorem 5.4 in terms of the number of points N instead of the parameter q. We will do this only in the situation of order-n-functions. However, as this is the worst case scenario, a similar result holds for more general sets Λ.

THEOREM 5.6. *Let $\Omega = [-1,1]^d$ and $\sigma > 1/2$. Let $1 \leq n \leq d/2$ and $\Lambda = \{\mathfrak{u} \subseteq \mathcal{D} : \#\mathfrak{u} \leq n\}$. Assume there is a $q \in \mathbb{N}$ such that the elements of $\boldsymbol{q} = \{q_\mathfrak{u} : \mathfrak{u} \in \Lambda\}$ have the form $q_\mathfrak{u} = \#\mathfrak{u} + q$. Let $X^{(d)}_{\Lambda,\boldsymbol{q}} \subseteq \Omega$ be a sampling point set as in Definition 5.2. Then, there is a constant $C = C_{\sigma,n} > 0$, such that, with $\tau = \sigma - 1/2$,*

$$\|f\|_{L_\infty(\Omega)} \leq C d^n \left[d^{\tau n} (\log N)^{\rho_1(\sigma,n)} N^{-\tau} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + (\log N)^{\rho_2(n)} \|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right],$$

where ρ_1 and ρ_2 are the same functions as in Theorem 5.3. for all $f \in H^\sigma_{\mathrm{mix},\Lambda}(\Omega)$.

Proof. We start with the bound from Theorem 5.4 and use first (5.4) and then (5.5) to eliminate the q terms. This yields with $C_n := \max_{\#\mathfrak{u} \leq n} C_\mathfrak{u}$ and $\tau = \sigma - 1/2$,

$$\begin{aligned}
\|f\|_{L_\infty(\Omega)} &\leq C_n \frac{(2d)^n}{n!} \left[q^{3n-2} 2^{-\tau q} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + q^{2n-1} \|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right] \\
&\leq C_n \frac{(2d)^n}{n!} \left[c_{\tau,n} d^{\tau n} q^{3n-2+\tau(n-1)} N^{-\tau} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + q^{2n-1} \|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right] \\
&= C_n \frac{(2d)^n}{n!} \left[c_{\tau,n} d^{\tau n} q^{\rho_1(\sigma,n)} N^{-\tau} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} + q^{\rho_2(n)} \|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right] \\
&\leq C_n \frac{(2d)^n}{n!} \left[\frac{c_{\tau,n}}{(\log 2)^{\rho_1(\sigma,n)}} d^{\tau n} (\log N)^{\rho_1(\sigma,n)} N^{-\tau} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)} \right. \\
&\qquad \left. + \left(\frac{\log N}{\log 2}\right)^{\rho_2(n)} \|f\|_{\ell_\infty(X^{(d)}_{\Lambda,\boldsymbol{q}})}\right].
\end{aligned}$$

We can now use this result to show that once again, the curse of dimensionality does not apply in this situation. We proceed in the same fashion as we did in the standard Sobolev case, noting that also $H^\sigma_{\mathrm{mix}}(\Omega)$ with $\sigma > 1/2$ is a reproducing kernel Hilbert space with positive definite reproducing kernel $K_\sigma : \Omega \times \Omega \to \mathbb{R}$ given by the tensor product of the reproducing kernel $k_\sigma : [-1,1] \times [-1,1] \to \mathbb{R}$ of $H^\sigma([-1,1])$ with itself, i.e.

$$K_\sigma(\boldsymbol{x}, \boldsymbol{y}) = \prod_{j=1}^d k_\sigma(x_j, y_j).$$

As in the case of standard Sobolev spaces we have the following result, which is proven in the same way as Theorem 4.7.

PROPOSITION 5.7. *Let $\Omega = [-1,1]^d$. Let $\Lambda \subseteq \mathcal{P}(d)$ be a downward closed set and let $\sigma > 1/2$. Then, $H^\sigma_{\mathrm{mix},\Lambda}(\Omega)$ is a closed subspace of $H^\sigma_{\mathrm{mix}}(\Omega)$ and hence a reproducing kernel Hilbert space. If $K_\sigma : \Omega \times \Omega \to \mathbb{R}$ is the reproducing kernel of $H^\sigma_{\mathrm{mix}}(\Omega)$ and $P : H^\sigma_{\mathrm{mix}}(\Omega) \to H^\sigma_{\mathrm{mix},\Lambda}(\Omega)$ is the orthogonal projection, then the reproducing kernel $K_{\sigma,\Lambda} : \Omega \times \Omega \to \mathbb{R}$ of $H^\sigma_{\mathrm{mix},\Lambda}(\Omega)$ is given by*

$$K_{\sigma,\Lambda}(\cdot, \boldsymbol{x}) = PK_\sigma(\cdot, \boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega.$$

The kernel is positive definite.

Hence, we can once again compute an interpolant of the form (4.7) in $\mathcal{O}(N^3)$ time. Using the stability of kernel-based interpolants, Theorem 5.6 yields the following result.

THEOREM 5.8. *Let $\Omega = [-1,1]^d$ and $\sigma > 1/2$. Let $1 \le n \le d/2$ and $\Lambda = \{\mathfrak{u} \subseteq \mathcal{D} : \#\mathfrak{u} \le n\}$. Assume there is a $q \in \mathbb{N}$ such that the elements of $\boldsymbol{q} = \{q_\mathfrak{u} : \mathfrak{u} \in \Lambda\}$ have the form $q_\mathfrak{u} = \#\mathfrak{u} + q$. Let $X^{(d)}_{\Lambda,\boldsymbol{q}} \subseteq \Omega$ be a sampling point set as in Definition 5.2. Let $s_f \in H^\sigma_{\mathrm{mix},n}(\Omega)$ be the interpolant to $f \in H^\sigma_{\mathrm{mix},n}(\Omega)$ using the reproducing kernel of $H^\sigma_{\mathrm{mix},n}(\Omega)$ and the set $X^{(d)}_{\Lambda,\boldsymbol{q}}$. Then,*

$$\|f - s_f\|_{L_\infty(\Omega)} \le C_{\sigma,n} d^{(\sigma+1/2)n} (\log N)^{\rho_1(\sigma,n)} N^{-(\sigma-1/2)} \|f\|_{H^\sigma_{\mathrm{mix}}(\Omega)}$$

with a constant $C_{\sigma,n}$ depending only on n and σ.

This indeed also implies that for such function classes, the curse of dimensionality does not apply. To see this, we use that $\sigma > 1/2$ implies that there is a $\delta > 0$ such that $\sigma + 2\delta > 1/2$. Since the logarithm of N is growing slower than any power of N, we can use the coarse estimate $(\log N)^{\rho_1(\sigma,n)} \le CN^\delta$. For $f \in H^\sigma_{\mathrm{mix},n}(\Omega)$ with $\|f\|_{H^\sigma_{\mathrm{mix},m}(\Omega)} = 1$ we see that we have accuracy $\epsilon > 0$ in the approximation if

$$C_{\sigma,n} d^{(\sigma+1/2)n} (\log N)^{\rho_1(\sigma,n)} N^{-(\sigma-1/2)} \le C_{\sigma,n} d^{(\sigma+1/2)n} N^{-\sigma+1/2+\delta} \le \epsilon.$$

Setting $\tau := \sigma - 1/2 - \delta$, this is obviously the case if

$$N \ge \epsilon^{-1/\tau} C_{\sigma,n}^{-1/\tau} d^{(\sigma+1/2)n/\tau}.$$

The right-hand side obviously grows only polynomially in d. Hence, we have the following result.

COROLLARY 5.9. *For $H^\sigma_{\mathrm{mix},n}(\Omega)$ with $\sigma > d/2$ and a fixed $n < d$, the curse of dimensionality does not apply.*

6. An Example: Anchored Sobolev Spaces. In this short section, we want to show that the construction of point sets and the corresponding error analysis indeed lead to the desired results. To this end, we use the theory of weighted, anchored Sobolev spaces over $[0,1]$, see for example [10, Example 4.2]. However, for simplicity, we will use only weights equal to one and use $\boldsymbol{c} = \boldsymbol{0}$ as the anchor. With this simplificiation, the general theory yields the following.

Given $m \in \mathbb{N}$, the Sobolev space $H^m([0,1])$ can be equipped with the inner product

$$\langle f, g \rangle_{H^m([0,1])} = \sum_{r=0}^{m-1} f^{(r)}(0) g^{(r)}(0) + \int_0^1 f^{(m)}(y) g^{(m)}(y) dy,$$

which induces a norm, which is equivalent to the standard norm on $H^m([0,1])$. With this inner product, $H^m([0,1])$ is of course also a reproducing kernel Hilbert space and the kernel is explicitly given by $(x,y) \mapsto 1 + k_m(x,y)$ with

$$k_m(x,y) = \sum_{r=1}^{m-1} \frac{x^r}{r!}\frac{y^r}{r!} + \int_0^1 \frac{(x-z)_+^{m-1}}{(m-1)!}\frac{(y-z)_+^{m-1}}{(m-1)!}dz. \tag{6.1}$$

Hence, according to the general theory, the reproducing kernel of $H^m_{\text{mix}}([0,1]^d)$ can be written by

$$K_m(\boldsymbol{x},\boldsymbol{y}) = \sum_{\mathfrak{u}\subseteq\mathcal{D}} K_{\mathfrak{u}}(\boldsymbol{x},\boldsymbol{y}), \qquad K_{\mathfrak{u}}(\boldsymbol{x},\boldsymbol{y}) = \prod_{j\in\mathfrak{u}} k_j(x_j,y_j),$$

where the product over the empty index set is defined to be 1.
As we have $k_m(\cdot,0) = 0$ by definition, the kernels $K_{\mathfrak{u}}$ satisfy the projection property that for all $\mathfrak{u}\subseteq\mathcal{D}$, $\mathfrak{u}\neq\emptyset$ and $j\in\mathfrak{u}$ we have

$$K_{\mathfrak{u}}(\cdot,\boldsymbol{y}) = 0, \qquad \text{if } y_j = 0.$$

Thus, according to [10, Theorem 4.1], we have the orthogonal decomposition

$$H^m_{\text{mix}}([0,1]^d) = \sum_{\mathfrak{u}\subseteq\mathcal{D}} H_{\mathfrak{u}},$$

where $H_{\mathfrak{u}}$ is the reproducing kernel Hilbert space with reproducing kernel $K_{\mathfrak{u}}$. These considerations lead to the following result.

PROPOSITION 6.1. *Let $\Lambda\subseteq\mathcal{P}(d)$ be downwards closed. Let $m\in\mathbb{N}$. Then, the Sobolev space $H^m_{\text{mix}}([0,1]^d)$ has the orthogonal decomposition*

$$H^m_{\text{mix}}([0,1]^d) = H^m_{\text{mix},\Lambda}([0,1])^d \oplus H^m_{\text{mix},\Lambda}([0,1]^d)^\perp,$$

where

$$H^m_{\text{mix},\Lambda}([0,1]^d) = \sum_{\mathfrak{u}\in\Lambda} H_{\mathfrak{u}}.$$

Moreover, the reproduing kernel of $H^m_{\text{mix},\Lambda}([0,1]^d)$ is given by

$$K_{m,\Lambda}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\mathfrak{u}\in\Lambda} K_{\mathfrak{u}}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\mathfrak{u}\in\Lambda}\prod_{j\in\mathfrak{u}} k_m(x_j,y_j), \qquad \boldsymbol{x},\boldsymbol{y}\in[0,1]^d,$$

where k_m is defined in (6.1).

Proof. It only remains to prove the statement on the reproducing kernel. However, this immidiately follows from standard results on sums of reproducing kernel Hilbert spaces, see for example [2, Part I.6]. ∎

In the rest of this section, we will only look at the two-dimensional case, i.e. $d=2$ and at order-1-functions, i.e. we will consider $H^m_{\text{mix},1}([0,1]^d)$. In other words, we will use the set of subsets $\Lambda = \{\emptyset,\{1\},\{2\}\}$. According to Proposition 6.1, the reproducing kernel of $H^m_{\text{mix},1}([0,1]^2)$ is given by

$$K_{m,\Lambda}(\boldsymbol{x},\boldsymbol{y}) = 1 + k_m(x_1,y_1) + k_m(x_2,y_2), \qquad \boldsymbol{x},\boldsymbol{y}\in[0,1]^2,$$

where k_m is the kernel from (6.1).

Next, we follow Definition 4.3 to define our sampling points. As we consider the approximation of a function of order one, this leads to a discrete set of points of the form

$$X_N = \left\{(x_1^{(N)}, 0), \ldots, (x_n^{(N)}, 0), (0, 0), (0, y_1^{(N)}), \ldots, (0, y_n^{(N)})\right\}$$

with $N = 2n + 1$ and $0 < x_i^{(N)}, y_i^{(N)} \leq 1$ for $1 \leq i \leq n$.
With this kernel and these interpolation points, we have computed the interpolant

$$s_f = \sum_{\boldsymbol{x} \in X_N}^{N} \alpha_{\boldsymbol{x}} K_{m,\Lambda}(\cdot, \boldsymbol{x})$$

by solving the linear system $A_{N,N}\boldsymbol{\alpha} = \boldsymbol{f}|X_N$ with the system matrix given by

$$A_{N,N} := (K_{m,\Lambda}(\boldsymbol{x}, \boldsymbol{y}))_{\boldsymbol{x},\boldsymbol{y} \in X_N} \in \mathbb{R}^{N \times N}.$$

To measure the error $f - s_f$, we have chosen a regular grid $Y_M \subseteq [0, 1]^2$ of $M = (2n)^2$, where n is the number of points on one axis in X_N and computed the discrete maximum norm of the error over Y_M, i.e. $e_N := \|f - s_f\|_{\ell_\infty(Y_M)}$.
As a test function, we used the smooth function $f(x_1, x_2) = x_1^2 + x_2^2 + 13x_2^4$. The smoothness of the Sobolev space was chosen as $m = 4$. The error is shown in Figure 1. According to Theorem 5.8, we expect the error to behave as $N^{-3.5}$ if we ignore the $\log N$ terms. However, the plot in Figure 1 indicates that we have an even better behaviour of $N^{-4.45}$, which is most certainly due to the fact that the target function is much smoother than required.

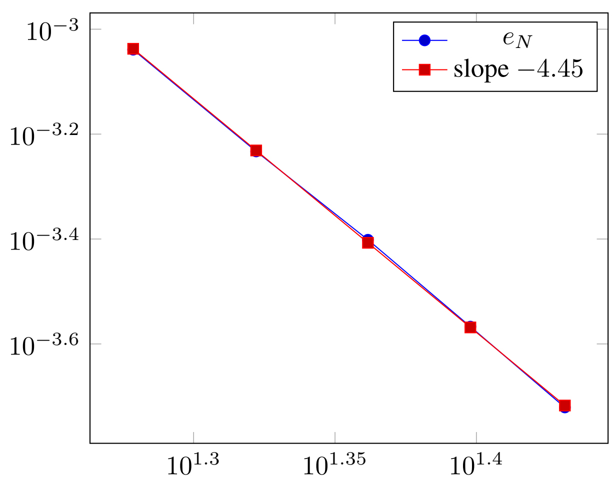

Fig. 1: Plot of the error for the example. The x-axis represents the number of points N, the y-axis represents the error.

REFERENCES

[1] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Academic Press, 2 edition, 2003.
[2] N. Aronszajn. Theory of reproducing kernels. *Trans. Am. Math. Soc.*, 68:337–404, 1950.
[3] C. D. Bastian and H. Rabitz. High dimensional model representation as a glass box in supervised machine learning. arXiv preprint, 2018.

[4] R. Bellmann. *Adaptive Control Processes: A guided Tour*. Princton University Press, 1961.

[5] R. A. DeVore and R. C. Sharpley. Besov spaces on domains in $\mathbb{R}^d$. *Trans. AMS*, 335:843–864, 1993.

[6] P. Giesl and H. Wendland. Meshless collocation: Error estimates with application to dynamical systems. *SIAM J. Numer. Anal.*, 45:1723–1741, 2007.

[7] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, New York, sixth edition, 2000.

[8] M. Griebel. Sparse grids for higher dimensional problems. In L. M. Pardo, A. Pinkus, E. Süli, and M. J. Todd, editors, *Foundations of Computational Mathematics, Santander 2005*, pages 106–161. Cambridge University Press, 2006.

[9] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer, Berlin, 2012.

[10] F. Y Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Wozniakowski. On decompositions of multivariate functions. *Math. Comput.*, 79:953–966, 2010.

[11] G. Li, X. Xing, W. Welsh, and H. Rabitz. High dimensional model representation constructed by support vector regression. i. independent variables with known probability distributions. *Journal of Mathematical Chemistry*, 55:278–303, 2017.

[12] T. Müller-Gronbach. Hyperbolic cross desings for approximation of random fields. *Journal of Statistical Planning and Inference*, 66:321–344, 1998.

[13] F. J. Narcowich, J. D. Ward, and H. Wendland. Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. *Math. Comput.*, 74:643–763, 2005.

[14] E. Novak. High dimensional numerical problems. *Nonlinear Anal., Theory Methods Appl.*, 30:1439–1446, 1997.

[15] D. Potts and M. Schmischke. Approximation of high-dimensional periodic functions with fourier-based methods. *SIAM Journal on Numerical Analysis*, 59(5):2393–2429, 2021.

[16] H. Rabitz and Ö. F. Aliş. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25:197–233, 1999.

[17] C. Rieger and H. Wendland. Sampling inequalities for sparse grids. *Numer. Math.*, 136:439 – 466, 2017.

[18] J. Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48(4):1875 –1897, 2020.

[19] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton, New Jersey, 1971.

[20] F. H. Szafraniec. The reproducing kernel property and its space: The basics. In Alpay D., editor, *Operator Theory*. Springer, Basel, 2015.

[21] H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, UK, 2005.

[22] H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101:643–662, 2005.

[23] J. Wloka. *Partial Differential Equations*. Cambridgen University Press, Cambridge, 1987.

[24] T. Ziehn and A. S. Tomlin. GUI-HDMR - a software tool for global sensitivity analysis of complex models. *Environmental Modelling & Software*, 24(7):775–785, 2009.