APPROXIMATION OF HIGH-DIMENSIONAL PERIODIC FUNCTIONS WITH FOURIER-BASED METHODS

DANIEL POTTS* AND MICHAEL SCHMISCHKE†

Abstract. In this paper we propose an approximation method for high-dimensional 1-periodic functions based on the multivariate ANOVA decomposition. We provide an analysis on the classical ANOVA decomposition on the torus and prove some important properties such as the inheritance of smoothness for Sobolev type spaces and the weighted Wiener algebra. We exploit special kinds of sparsity in the ANOVA decomposition with the aim to approximate a function in a scattered data or black-box approximation scenario. This method allows us to simultaneously achieve an importance ranking on dimensions and dimension interactions which is referred to as attribute ranking in some applications. In scattered data approximation we rely on a special algorithm based on the non-equispaced fast Fourier transform (or NFFT) for fast multiplication with arising Fourier matrices. For black-box approximation we choose the well-known rank-1 lattices as sampling schemes and show properties of the appearing special lattices.

Key words. ANOVA decomposition, high-dimensional approximation, Fourier approximation

AMS subject classifications. 65T, 42B05

1. Introduction. The approximation of high-dimensional functions is an important and current topic with great interest in many applications. We consider a setting of periodic functions $f: \mathbb{T}^d \to \mathbb{C}$, $d \in \mathbb{N}$, over the torus $\mathbb{T}$ where certain data about the function is known. Here, we distinguish between a black-box setting, i.e., f can be evaluated at points $\boldsymbol{x} \in \mathbb{T}^d$ at a certain cost, and a scattered data setting, i.e., sampling points $X \subseteq \mathbb{T}^d$ and function values $(f(\boldsymbol{x}))_{\boldsymbol{x} \in X}$ are given. Besides the natural question of wanting to find an approximation for f, we want to consider the question of interpretability, i.e., analyzing the importance of the dimensions and dimension interactions of the function. In applications this is sometimes referred to as an *attribute ranking*.

The main tool to achieve our goals is the **analysis of variance (ANOVA)** decomposition [6, 45, 38, 23] which is an important model in the analysis of dimension interactions of multivariate, high-dimensional functions. It has proved useful in understanding the reason behind the success of certain quadrature methods for high-dimensional integration [40, 4, 17] and also infinite-dimensional integration [1, 19, 34]. The ANOVA decomposition decomposes a d-variate function in 2^d ANOVA terms where each term belongs to a subset of $\mathcal{D} := \{1, 2, \ldots, d\}$. The single term depends only on the variables in the corresponding subset and the number of these variables is the order of the ANOVA term. In this paper we study the classical ANOVA decomposition for periodic functions and how it acts on the frequency domain. The decomposition is referred to as *classical* since it is based on an integral projection operator. In this setting we find relationships between ANOVA terms and the support of the frequencies as subsets of $\mathbb{Z}^d$. Moreover, we prove formulas for the representation of ANOVA terms and projections.

Classical approximation methods cannot be applied for high-dimensional functions in general since the data required increases exponentially because of the curse of dimensionality. However, the observation has been made that in practical applications

---

*Chemnitz University of Technology, Germany (potts@math.tu-chemnitz.de, http://www.tu-chemnitz.de/~potts/).

†Chemnitz University of Technology, Germany (michael.schmischke@math.tu-chemnitz.de, http://www.tu-chemnitz.de/~mischmi/).

with multivariate, high-dimensional functions, often only the ANOVA terms of a low order are enough to describe a function, see e.g. [6]. This leads to the notion of a superposition dimension $d_s \in \mathcal{D}$ that limits the order of the ANOVA terms involved. Using this as a sparsity assumption to circumvent the curse of dimensionality, we consider functions where the ANOVA decomposition is mostly supported on terms of a low order, i.e., the norm of the remaining decomposition weighed by the norm of f is small. This leads to a truncation of the decomposition through a superposition threshold. We consider how the previously described error can be related to the decay of Fourier coefficients and specifically the smoothness of f.

We present and analyze an approximation method that uses sensitivity analysis, cf. [47, 48, 38], on the truncated ANOVA decomposition which is able to identify important ANOVA terms and incorporate this information in finding an approximation. The goal is to simplify the approximation model which yields benefits in reducing the influence of overfitting regarding the amount of data. We determine approximations of the Fourier coefficients of the function (or *learn* them) by solving least-squares problems. This is done through exploiting the special structure of the system matrix by identifying submatrices with the corresponding ANOVA terms. In the case of black-box approximations we are using rank-1 lattices as a spatial discretization, see e.g. [24, 25, 26, 27], and for scattered data approximation we rely on the iterative LSQR method [42] and the fast matrix-vector multiplications for Fourier matrices provided by the non-equispaced fast Fourier transform (or NFFT) introduced in [31].

The paper is organized as follows. In Section 3 we introduce the classical ANOVA decomposition and study its behavior for periodic functions with regard to the Fourier system. We prove new formulas for the Fourier coefficients of projections in Lemma 3.1 and ANOVA terms in Lemma 3.5. Moreover, we prove that functions in Sobolev type spaces and the weighted Wiener algebra inherit their smoothness to the ANOVA terms, see Theorem 3.10 and Theorem 3.11. In Section 4 we consider the truncated ANOVA decomposition and prove formulas for their Fourier coefficients, see Lemma 4.1 and Corollary 4.2. We also give direct formulas for the truncated decomposition using the projections in Lemma 4.4 and Corollary 4.5. Furthermore, we relate Sobolev type spaces and the weighted Wiener algebra to the previously introduced functions of low-dimensional structure and compute the errors in Theorem 4.6 and Theorem 4.7. Specifically, we consider a class of product and order-dependent weights, see [35, 13, 33, 14], of functions with isotropic and dominating-mixed smoothness, cf. [16, 26, 5], to obtain specific error bounds, see Corollary 4.8 and Corollary 4.10. In Section 5 we present an approximation method for functions that are of an low-dimensional structure, cf. Algorithm 5.1. We start by considering a black-box approximation scenario with rank-1 lattices as sampling schemes and show properties of the arising lattices in Lemma 5.3, Lemma 5.5, and Corollary 5.6. Furthermore, we discuss scattered data approximation in Subsection 5.2.2. The arising approximation errors are considered in Section 6 with main results being Theorem 6.1, Theorem 6.2, Theorem 6.5, and Theorem 6.7. In Section 7 we perform numerical experiments with a specific benchmark function.

2. Prerequisites and Notation. We consider multivariate 1-periodic functions $f \colon \mathbb{T}^d \to \mathbb{C}$ with spatial dimension $d \in \mathbb{N}$, which are square-integrable, i.e., elements of $\mathrm{L}_2(\mathbb{T}^d)$. Those functions have a unique representation with regard to the Fourier

system $\{e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}\}_{\boldsymbol{k} \in \mathbb{Z}^d}$ as Fourier series

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}},$$

where $c_{\boldsymbol{k}}(f) := \int_{\mathbb{T}^d} f(\boldsymbol{x}) \, e^{-2\pi i \boldsymbol{k} \cdot \boldsymbol{x}} \mathrm{d}\boldsymbol{x} \in \mathbb{C}$, $\boldsymbol{k} \in \mathbb{Z}^d$, are the Fourier coefficients of f. Given a finite index set $I \subseteq \mathbb{Z}^d$, we call the trigonometric polynomial

(2.1) $$S_I f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in I} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}$$

the Fourier partial sum of f with respect to the index set I.

In this paper we make use of indexing with sets. First, for a given spatial dimension d we denote with $\mathcal{D} = \{1, 2, \ldots, d\}$ the set of coordinate indices and subsets as bold small letters, e.g., $\boldsymbol{u} \subseteq \mathcal{D}$. The complement of those subsets are always with respect to $\mathcal{D}$, i.e., $\boldsymbol{u}^c = \mathcal{D} \setminus \boldsymbol{u}$. For a vector $\boldsymbol{x} \in \mathbb{C}^d$ we define $\boldsymbol{x}_{\boldsymbol{u}} = (x_i)_{i \in \boldsymbol{u}} \in \mathbb{C}^{|\boldsymbol{u}|}$. There remains a small ambiguity regarding the order of the components of $\boldsymbol{x}_{\boldsymbol{u}}$ which can be clarified if one chooses a consistent ordering, e.g., ascending order which would be a natural choice.

Furthermore, we use the p-norm which is defined as

$$\|\boldsymbol{x}\|_p = \begin{cases} |\{i \in \mathcal{D} \colon x_i \neq 0\}| & \colon p = 0 \\ \left(\sum_{i=1}^d |x_i|^p\right)^{1/p} & \colon 1 \leq p < \infty \\ \max_{i \in \mathcal{D}} |x_i| & \colon p = \infty \end{cases}$$

for $\boldsymbol{x} \in \mathbb{R}^d$. Note that the case $1 \leq p < \infty$ can be expanded to $0 < p < 1$, but then $\|\cdot\|_p$ would only be a quasi-norm. In the case $p = 0$, $\|\cdot\|_p$ is not a norm at all.

2.1. Rank-1 lattice. In the case of black-box approximation we are going to rely on rank-1 lattice as sampling schemes, see e.g. [24, 25, 26, 27]. For a given lattice size $M \in \mathbb{N}$ and a generating vector $\boldsymbol{z} \in \mathbb{Z}^d$ we define a rank-1 lattice

$$\Lambda(\boldsymbol{z}, M) := \left\{ \boldsymbol{x}_j := \frac{j}{M} \boldsymbol{z} \mod \boldsymbol{1} \colon j = 0, 1, \ldots, M - 1 \right\}.$$

These lattices are useful in the evaluation of trigonometric polynomials

$$p \in \Pi_I := \mathrm{span}\left\{ e^{2\pi i \boldsymbol{k} \cdot \circ} \colon \boldsymbol{k} \in I \right\}$$

over a finite index set $I \subseteq \mathbb{Z}^d$ for given Fourier coefficients $c_{\boldsymbol{k}}(p)$. As discussed in [37], we have

$$p(\boldsymbol{x}_j) = p\left(\frac{j}{M} \boldsymbol{z} \mod \boldsymbol{1}\right) = \sum_{l=0}^{M-1} \left(\sum_{\substack{\boldsymbol{k} \in I \\ \boldsymbol{k} \cdot \boldsymbol{z} \equiv l \mod M}} c_{\boldsymbol{k}}(p) \right) e^{2\pi i \frac{jl}{M}}.$$

The computation of the sum over l can be realized trough a one-dimensional FFT and therefore the evaluation of p at all lattice nodes can be done using only this single FFT. The arithmetic complexity of this evaluation is in $\mathcal{O}(M \log M + d |I|)$.

However, using a special kind of rank-1 lattice, it is possible to reconstruct the Fourier coefficients $c_{\boldsymbol{k}}(p)$ by sampling p at the nodes in $\Lambda(\boldsymbol{z}, M)$ in an exact and stable

way. For an index set $I \subseteq \mathbb{Z}^d$ we define a reconstructing rank-1 lattice $\Lambda(\boldsymbol{z}, M, I)$ as a rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ such that the condition

$$\boldsymbol{m} \cdot \boldsymbol{z} \not\equiv 0 \mod M \quad \forall \boldsymbol{m} \in \mathcal{D}(I) \setminus \{\boldsymbol{0}\}$$

is fulfilled with

$$(2.2) \qquad \mathcal{D}(I) := \{\boldsymbol{k} - \boldsymbol{h} \colon \boldsymbol{k}, \boldsymbol{h} \in I\}$$

being the difference set for I. Using the nodes of a reconstructing rank-1 lattice $\Lambda(\boldsymbol{z}, M, I)$, the Fourier coefficients can be calculated as

$$c_{\boldsymbol{k}}(p) = \frac{1}{M} \sum_{j=0}^{M} p\left(\frac{j}{M}\boldsymbol{z} \mod \boldsymbol{1}\right) e^{-2\pi i j \frac{\boldsymbol{k} \cdot \boldsymbol{z}}{M}}.$$

The calculation of all Fourier coefficients $c_{\boldsymbol{k}}(p)$, $\boldsymbol{k} \in I$, can then be realized trough a one-dimensional FFT and the computation of the products $\boldsymbol{k} \cdot \boldsymbol{z}$. Consequently, the arithmetic complexity of this evaluation is again in $\mathcal{O}(M \log M + d |I|)$.

The principle of this reconstruction can be generalized to functions $f \in \mathcal{A}^w(\mathbb{T}^d) := \{f \in L_1(\mathbb{T}^d) \colon \|f\|_{\mathcal{A}^w(\mathbb{T}^d)} := \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w(\boldsymbol{k}) |c_{\boldsymbol{k}}(f)| < \infty\}$, $w \colon \mathbb{Z}^d \to [1, \infty)$, by taking the Fourier partial sum $S_I f$ for a suitable index set $I \subseteq \mathbb{Z}^d$ and treating the evaluations of f as the evaluations of the trigonometric polynomial $S_I f$. Using the same idea as before, we get

$$c_{\boldsymbol{k}}(f) \approx \hat{f}_{\boldsymbol{k}} := \frac{1}{M} \sum_{j=0}^{M} f\left(\frac{j}{M}\boldsymbol{z} \mod \boldsymbol{1}\right) e^{-2\pi i j \frac{\boldsymbol{k} \cdot \boldsymbol{z}}{M}}$$

with a a reconstructing rank-1 lattice $\Lambda(\boldsymbol{z}, M, I)$. The error for each coefficient is

$$(2.3) \qquad \hat{f}_{\boldsymbol{k}} = c_{\boldsymbol{k}}(f) + \sum_{\boldsymbol{h} \in \Lambda^\perp(\boldsymbol{z}, M) \setminus \{\boldsymbol{0}\}} c_{\boldsymbol{k}+\boldsymbol{h}}(f)$$

with the integer dual lattice

$$\Lambda^\perp(\boldsymbol{z}, M) := \left\{\boldsymbol{k} \in \mathbb{Z}^d \colon \boldsymbol{k} \cdot \boldsymbol{z} \equiv 0 \mod M\right\}.$$

Consequently, if $\sum_{\boldsymbol{h} \in \Lambda^\perp(\boldsymbol{z}, M) \setminus \{\boldsymbol{0}\}} |c_{\boldsymbol{k}+\boldsymbol{h}}(f)|$ is small, the approximations $\hat{f}_{\boldsymbol{k}}$ are close to the Fourier coefficients $c_{\boldsymbol{k}}(f)$. For further details on this topic we refer to [26, 27] and [43, Chapter 8].

3. The classical ANOVA decomposition of 1-periodic functions. In this section we introduce the ANOVA decomposition, see e.g. [6, 38, 23], and derive new results for the periodic setting specifically with regard to the decomposition acting on the frequency domain.

We start by defining the projection operator

$$(3.1) \qquad \mathrm{P}_{\boldsymbol{u}} f(\boldsymbol{x}_{\boldsymbol{u}}) := \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}^c}$$

that integrates over the variables $\boldsymbol{x}_{\boldsymbol{u}^c}$. For $|\boldsymbol{u}| > 0$ this operator maps a function from $L_2(\mathbb{T}^d)$ to $L_2(\mathbb{T}^{|\boldsymbol{u}|})$ by the Cauchy-Schwarz inequality and the image $\mathrm{P}_{\boldsymbol{u}} f$ depends

only on the variables $\boldsymbol{x_u} \in \mathbb{T}^{|\boldsymbol{u}|}$. In the case of $\boldsymbol{u} = \emptyset$, the projection gives us the integral of f. We define the index set

$$(3.2) \qquad \mathbb{P}_{\boldsymbol{u}}^{(d)} := \left\{ \boldsymbol{k} \in \mathbb{Z}^d : \boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0} \right\}$$

which can be identified with $\mathbb{Z}^{|\boldsymbol{u}|}$ using the mapping $\boldsymbol{k} \mapsto \boldsymbol{k_u}$. Note that we use the convention $\mathbb{Z}^{|\emptyset|} = \{0\}$. We now prove a relationship between the Fourier coefficients of $\mathrm{P}_{\boldsymbol{u}} f$ and f.

LEMMA 3.1. *Let $f \in \mathrm{L}_2(\mathbb{T}^d)$ and $\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}$. Then*

$$c_{\boldsymbol{\ell}}(\mathrm{P}_{\boldsymbol{u}} f) = c_{\boldsymbol{k}}(f)$$

for $\boldsymbol{k} \in \mathbb{Z}^d$ with $\boldsymbol{k_u} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0}$.

Proof. We consolidate the two integrals and derive

$$
\begin{aligned}
c_{\boldsymbol{\ell}}(\mathrm{P}_{\boldsymbol{u}} f) &= \int_{\mathbb{T}^{|\boldsymbol{u}|}} \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}^c} \, \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x_u}} \mathrm{d}\boldsymbol{x_u} \\
&= \int_{\mathbb{T}^d} f(\boldsymbol{x}) \, \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x_u}} \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{T}^d} f(\boldsymbol{x}) \, \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \mathrm{d}\boldsymbol{x} = c_{\boldsymbol{k}}(f) \, . \qquad \square
\end{aligned}
$$

Using Lemma 3.1, we are able to write $\mathrm{P}_{\boldsymbol{u}} f$ as both, a d-dimensional Fourier series $\mathrm{P}_{\boldsymbol{u}} f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{k}}(f) \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}$ and a $|\boldsymbol{u}|$-dimensional Fourier series $\mathrm{P}_{\boldsymbol{u}} f(\boldsymbol{x_u}) = \sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} c_{\boldsymbol{\ell}}(\mathrm{P}_{\boldsymbol{u}} f) \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x_u}}$.

Now, we recursively define the **ANOVA term** for $\boldsymbol{u} \subseteq \mathcal{D}$

$$(3.3) \qquad f_{\boldsymbol{u}} := \mathrm{P}_{\boldsymbol{u}} f - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} f_{\boldsymbol{v}}.$$

There exists a direct formula for the ANOVA terms $f_{\boldsymbol{u}}$ defined in (3.3).

LEMMA 3.2. *Let $a \in \mathbb{N}_0$ and $b \in \mathbb{N}$ with $b > a$. Then*

$$\sum_{n=a}^{b-1} (-1)^{n-a+1} \binom{b-a}{n-a} = (-1)^{b-a}.$$

Proof. We prove an equivalent form obtained through multiplication with $(-1)^a$ and an index shift

$$\sum_{n=0}^{b-a-1} (-1)^{n+a+1} \binom{b-a}{n} = (-1)^b.$$

Splitting the sum and applying the Binomial theorem yields

$$
\begin{aligned}
\sum_{n=0}^{b-a-1} (-1)^{n+a+1} \binom{b-a}{n} &= \sum_{n=0}^{b-a} (-1)^{n+a+1} \binom{b-a}{n} - (-1)^{b+1} \\
&= (-1)^{a+1} \underbrace{\sum_{n=0}^{b-a} (-1)^n \binom{b-a}{n}}_{=(-1+1)^{b-a}=0} + (-1)^b \\
&= (-1)^b. \qquad \square
\end{aligned}
$$

LEMMA 3.3. *Let* $f \in \mathrm{L}_2(\mathbb{T}^d)$ *with* $\boldsymbol{u} \subseteq \mathcal{D}$. *Then*

$$(3.4) \qquad\qquad f_{\boldsymbol{u}} = \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{P}_{\boldsymbol{v}} f.$$

Proof. A proof based on properties of projection operators was given in [36] while we use combinatorial arguments. We prove this statement through structural induction over the cardinality of $\boldsymbol{u}$. For $|\boldsymbol{u}| = 0$, i.e., $\boldsymbol{u} = \emptyset$, we have

$$(-1)^{0-0} \left(P_\emptyset\right)(\boldsymbol{x}) = \left(P_\emptyset\right)(\boldsymbol{x}) = \left(P_\emptyset\right)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \emptyset} f_{\boldsymbol{v}}(\boldsymbol{x}).$$

Now, let (3.4) be true for $\boldsymbol{v} \subseteq \mathcal{D}$, $|\boldsymbol{v}| = 0, 1, \ldots, m-1$, $m \in \{1, 2, \ldots, d\}$, and take a subset $\boldsymbol{u} \subseteq \mathcal{D}$ with $|\boldsymbol{u}| = m$. We use the notation

$$\delta_{\boldsymbol{w} \subseteq \boldsymbol{v}} = \begin{cases} 1 & : \boldsymbol{w} \subseteq \boldsymbol{v} \\ 0 & : \text{otherwise.} \end{cases}$$

and start from the recursive expression in (3.3) to obtain

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \left(P_{\boldsymbol{u}} f\right)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} f_{\boldsymbol{v}}(\boldsymbol{x}) = \left(P_{\boldsymbol{u}} f\right)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \sum_{\boldsymbol{w} \subseteq \boldsymbol{v}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \left(P_{\boldsymbol{w}} f\right)(\boldsymbol{x})$$

$$= \left(P_{\boldsymbol{u}} f\right)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \left(P_{\boldsymbol{w}} f\right)(\boldsymbol{x}) \delta_{\boldsymbol{w} \subseteq \boldsymbol{v}}.$$

We exchange the two sums and sum over the order of the ANOVA terms

$$\sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \left(P_{\boldsymbol{w}} f\right)(\boldsymbol{x}) \delta_{\boldsymbol{w} \subseteq \boldsymbol{v}} = \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} \left(P_{\boldsymbol{w}} f\right)(\boldsymbol{x}) \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \delta_{\boldsymbol{w} \subseteq \boldsymbol{v}}$$

$$= \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} \left(P_{\boldsymbol{w}} f\right)(\boldsymbol{x}) \sum_{n=|\boldsymbol{w}|}^{m-1} \sum_{\substack{\boldsymbol{v} \subseteq \boldsymbol{u} \\ |\boldsymbol{v}|=n}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \delta_{\boldsymbol{w} \subseteq \boldsymbol{v}}$$

$$= \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} \left(P_{\boldsymbol{w}} f\right)(\boldsymbol{x}) \sum_{n=|\boldsymbol{w}|}^{m-1} (-1)^{n-|\boldsymbol{w}|} \sum_{\substack{\boldsymbol{v} \subseteq \boldsymbol{u} \\ |\boldsymbol{v}|=n}} \delta_{\boldsymbol{w} \subseteq \boldsymbol{v}}.$$

Applying Lemma 3.2 yields the formula. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We proceed to present a relationship between the Fourier coefficients of $f_{\boldsymbol{u}}$ and f. Furthermore, we prove $f_{\boldsymbol{u}} \in \mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})$. Therefore, we define the index set

$$\mathbb{F}_{\boldsymbol{u}}^{(d)} := \left\{ \boldsymbol{k} \in \mathbb{Z}^d : \boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0}, k_j \neq 0 \, \forall j \in \boldsymbol{u} \right\}$$

which can be identified with $(\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}$ using the mapping $\boldsymbol{k} \mapsto \boldsymbol{k}_{\boldsymbol{u}}$. Here, we use the convention $(\mathbb{Z} \setminus \{0\})^{|\emptyset|} = \{0\}$.

LEMMA 3.4. *Let* $\boldsymbol{u}, \boldsymbol{v} \subseteq \mathcal{D}$ *with* $\boldsymbol{u} \neq \boldsymbol{v}$. *Then* $\mathbb{F}_{\boldsymbol{u}}^{(d)} \cap \mathbb{F}_{\boldsymbol{v}}^{(d)} = \emptyset$. *Moreover, we have*

$$\mathbb{Z}^d = \bigcup_{\boldsymbol{u} \subseteq \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)}.$$

Proof. Let $\boldsymbol{u}, \boldsymbol{v} \subseteq \mathcal{D}$, $\boldsymbol{u} \neq \boldsymbol{v}$, and w.l.o.g. $|\boldsymbol{u}| \geq |\boldsymbol{v}|$. We assume there exists a $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{u}}^{(d)} \cap \mathbb{F}_{\boldsymbol{v}}^{(d)}$ and first consider the case $\boldsymbol{u} \cap \boldsymbol{v} = \emptyset$. Since $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}$ we have $\tilde{\boldsymbol{k}}_{\boldsymbol{u}^{\mathrm{c}}} = \boldsymbol{0}$ and therefore $\tilde{\boldsymbol{k}}_{\boldsymbol{v}} = \boldsymbol{0}$. This contradicts $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{v}}^{(d)}$. In the case of $\boldsymbol{u} \cap \boldsymbol{v} \neq \emptyset$ there exists a $j \in \boldsymbol{v}^{\mathrm{c}} \cap \boldsymbol{u}$. Then $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{v}}^{(d)}$ implies that $\tilde{k}_j = 0$ which contradicts $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}$.

The inclusion $\bigcup_{\boldsymbol{u} \subseteq \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)} \subseteq \mathbb{Z}^d$ is trivial since $\mathbb{F}_{\boldsymbol{u}}^{(d)} \subseteq \mathbb{Z}^d$ for every $\boldsymbol{u} \subseteq \mathcal{D}$. To prove $\mathbb{Z}^d \subseteq \bigcup_{\boldsymbol{u} \subseteq \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)}$ we take a $\boldsymbol{k} \in \mathbb{Z}^d$ and define $\boldsymbol{u} = \{j \in \mathcal{D} \colon k_j \neq 0\}$. Then $\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}$ and therefore $\boldsymbol{k} \in \bigcup_{\boldsymbol{u} \subseteq \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)}$. $\qquad\square$

LEMMA 3.5. *Let $f \in \mathrm{L}_2(\mathbb{T}^d)$ with $\boldsymbol{u} \subseteq \mathcal{D}$ and $\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}$. Then*

$$c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = \begin{cases} c_{\boldsymbol{k}}(f) & : \boldsymbol{\ell} \in (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|} \\ \delta_{\boldsymbol{u}, \emptyset} \cdot c_0(f) & : \boldsymbol{\ell} = \boldsymbol{0} \\ 0 & : otherwise \end{cases}$$

for $\boldsymbol{k} \in \mathbb{Z}^d$ with $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\boldsymbol{u}^{\mathrm{c}}} = \boldsymbol{0}$. Furthermore, $f_{\boldsymbol{u}} \in \mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})$.

Proof. We begin by employing the direct formula (3.4) to obtain

$$c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = \int_{\mathbb{T}^{|\boldsymbol{u}|}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) \, \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}} \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}}$$

$$= \int_{\mathbb{T}^{|\boldsymbol{u}|}} \left[\sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} \mathrm{P}_{\boldsymbol{v}} f(\boldsymbol{x}_{\boldsymbol{v}}) \right] \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}} \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}}$$

$$= \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} \int_{\mathbb{T}^{|\boldsymbol{u}|}} \mathrm{P}_{\boldsymbol{v}} f(\boldsymbol{x}_{\boldsymbol{v}}) \, \mathrm{e}^{-2\pi\mathrm{i}\boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}} \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}}$$

$$= \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} c_{\boldsymbol{k}_{\boldsymbol{v}}} (\mathrm{P}_{\boldsymbol{v}} f) \, \delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}}.$$

We go on to prove $c_0(f_{\boldsymbol{u}}) = \delta_{\boldsymbol{u}, \emptyset} \cdot c_0(f)$. In this case, $\boldsymbol{k}_{\boldsymbol{v}} = \boldsymbol{0}$ and $\delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}} = 1$ for every $\boldsymbol{v} \subseteq \boldsymbol{u}$. By the Binomial Theorem, we have

$$c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} c_{\boldsymbol{k}_{\boldsymbol{v}}} (\mathrm{P}_{\boldsymbol{v}} f) \, \delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}} = c_0(f) \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|}$$

$$= c_0(f) \sum_{n=0}^{|\boldsymbol{u}|} \binom{|\boldsymbol{u}|}{n} (-1)^{|\boldsymbol{u}| - n} = c_0(f) \cdot \delta_{\boldsymbol{u}, \emptyset}.$$

For the second case, we consider an $\boldsymbol{\ell}$ and with a set $\overline{\boldsymbol{v}} \subseteq \boldsymbol{u}$ such that $\emptyset \neq \overline{\boldsymbol{v}} := \{i \in \boldsymbol{u} \colon k_i = 0\} \neq \boldsymbol{u}$. Then $\delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}} = 1 \iff \overline{\boldsymbol{v}}^{\mathrm{c}} := \boldsymbol{u} \setminus \overline{\boldsymbol{v}} \subseteq \boldsymbol{v}$ and with the Binomial Theorem we get

$$c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} c_{\boldsymbol{k}_{\boldsymbol{v}}} (\mathrm{P}_{\boldsymbol{v}} f) \, \delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}} = \sum_{\overline{\boldsymbol{v}}^{\mathrm{c}} \subseteq \boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} c_{\boldsymbol{k}_{\boldsymbol{v}}} (\mathrm{P}_{\boldsymbol{v}} f)$$

$$= c_{\boldsymbol{k}}(f) \sum_{\overline{\boldsymbol{v}}^{\mathrm{c}} \subseteq \boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} = c_{\boldsymbol{k}}(f) \sum_{n=|\overline{\boldsymbol{v}}^{\mathrm{c}}|}^{|\boldsymbol{u}|} \binom{|\boldsymbol{u}| - |\overline{\boldsymbol{v}}^{\mathrm{c}}|}{n - |\overline{\boldsymbol{v}}^{\mathrm{c}}|} (-1)^{|\boldsymbol{u}| - n}$$

$$= c_{\boldsymbol{k}}(f) \sum_{m=0}^{|\boldsymbol{u}| - |\overline{\boldsymbol{v}}^{\mathrm{c}}|} \binom{|\boldsymbol{u}| - |\overline{\boldsymbol{v}}^{\mathrm{c}}|}{m} (-1)^{|\boldsymbol{u}| - |\overline{\boldsymbol{v}}^{\mathrm{c}}| - m} = 0.$$

For the case were the entries of $\boldsymbol{\ell}$ are all nonzero, only the addend where $\boldsymbol{v} = \boldsymbol{u}$ is nonzero, i.e., $c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = c_{\boldsymbol{k}}(f)$. $\qquad \square$

With Lemma 3.5 we have two equivalent series representations for the ANOVA term $f_{\boldsymbol{u}}$, the d-dimensional Fourier series $f_{\boldsymbol{u}}(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}$ and the $|\boldsymbol{u}|$-dimensional Fourier series $f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) = \sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) \, e^{2\pi i \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}}$ with $c_{\boldsymbol{\ell}}(f_{\boldsymbol{u}})$ as in Lemma 3.5. The ANOVA terms have the following important property.

COROLLARY 3.6. *Let $f \in L_2(\mathbb{T}^d)$ and $\boldsymbol{u}, \boldsymbol{v} \subseteq \mathcal{D}$ with $\boldsymbol{u} \neq \boldsymbol{v}$. Then the ANOVA terms $f_{\boldsymbol{u}}$ and $f_{\boldsymbol{v}}$ are orthogonal, i.e.,*

$$\langle f_{\boldsymbol{u}}, f_{\boldsymbol{v}} \rangle_{L_2(\mathbb{T}^d)} = 0.$$

Proof. We employ Lemma 3.4 and Lemma 3.5 to deduce

$$\langle f_{\boldsymbol{u}}, f_{\boldsymbol{v}} \rangle_{L_2(\mathbb{T}^d)} = \langle \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}, \sum_{\boldsymbol{\ell} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{\ell}}(f) \, e^{2\pi i \boldsymbol{\ell} \cdot \boldsymbol{x}} \rangle_{L_2(\mathbb{T}^d)}$$

$$= \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} \sum_{\boldsymbol{\ell} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{k}}(f) \, \overline{c_{\boldsymbol{\ell}}(f)} \, \delta_{\boldsymbol{k}, \boldsymbol{\ell}} = 0. \qquad \square$$

Having defined the ANOVA terms, we now go on to the **ANOVA decomposition**, cf. [6, 38].

THEOREM 3.7. *Let $f \in L_2(\mathbb{T}^d)$, the ANOVA terms $f_{\boldsymbol{u}}$ as in (3.3) and the set of coordinate indices $\mathcal{D} = \{1, 2, \ldots, d\}$. Then f can be uniquely decomposed as*

$$(3.5) \quad f(\boldsymbol{x}) = f_{\emptyset} + \sum_{i=1}^{d} f_{\{i\}}(x_i) + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} f_{\{i,j\}}(\boldsymbol{x}_{\{i,j\}}) + \cdots + f_{\mathcal{D}}(\boldsymbol{x}) = \sum_{\boldsymbol{u} \subseteq \mathcal{D}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$$

*which we call **analysis of variance (ANOVA) decomposition**.*

Proof. We use that $\mathbb{Z}^d$ is the disjoint union of the sets $\mathbb{F}_{\boldsymbol{u}}^{(d)}$ for $\boldsymbol{u} \subseteq \mathcal{D}$ and obtain

$$\sum_{\boldsymbol{u} \subseteq \mathcal{D}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) = \sum_{\boldsymbol{u} \subseteq \mathcal{D}} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}} = \sum_{\boldsymbol{k} \in \bigcup_{\boldsymbol{u} \subseteq \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)}} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}$$

$$= \sum_{\boldsymbol{k} \in \mathbb{Z}^d} c_{\boldsymbol{k}}(f) \, e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}} = f(\boldsymbol{x}).$$

Since the union is disjoint, the decomposition is unique. $\qquad \square$

Remark 3.8. The ANOVA decomposition (3.5) depends strongly on the projection operator $P_{\boldsymbol{u}} f$, see (3.1). The integral operator considered in this paper leads to the so called classical ANOVA decomposition. Another important variant is the anchored decomposition where one chooses an anchor point $\boldsymbol{c} \in \mathbb{T}^d$ and the projection operator is then defined as

$$P_{\boldsymbol{u}} f(\boldsymbol{x}_{\boldsymbol{u}}) = f(\boldsymbol{y}), \, \boldsymbol{y}_{\boldsymbol{u}} = \boldsymbol{x}_{\boldsymbol{u}}, \boldsymbol{y}_{\boldsymbol{u}^c} = \boldsymbol{c}_{\boldsymbol{u}^c}.$$

This decomposition can for example be used in methods for the integration of high-dimensional functions such as the multivariate decomposition method, see e.g. [34, 11]. However, the error analysis may again be based on the classical ANOVA decomposition, see e.g. [12].

In Figure 1 we have visualized the different frequency index sets $\mathbb{F}_{\boldsymbol{u}}^{(d)}$, $\boldsymbol{u} \subseteq \mathcal{D}$, for a 3-dimensional example.
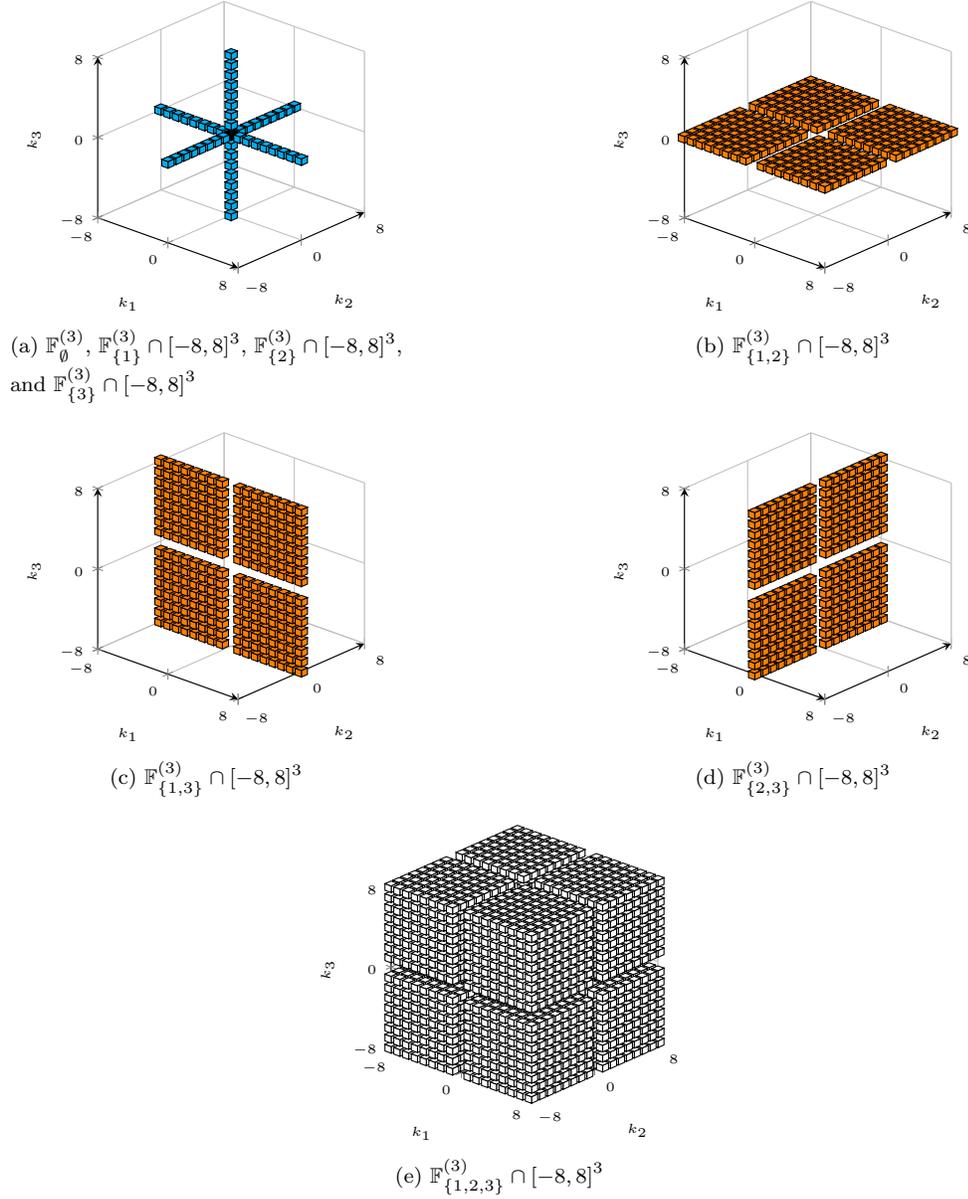
(a) $\mathbb{F}_{\emptyset}^{(3)}$, $\mathbb{F}_{\{1\}}^{(3)} \cap [-8,8]^3$, $\mathbb{F}_{\{2\}}^{(3)} \cap [-8,8]^3$, and $\mathbb{F}_{\{3\}}^{(3)} \cap [-8,8]^3$

(b) $\mathbb{F}_{\{1,2\}}^{(3)} \cap [-8,8]^3$

(c) $\mathbb{F}_{\{1,3\}}^{(3)} \cap [-8,8]^3$

(d) $\mathbb{F}_{\{2,3\}}^{(3)} \cap [-8,8]^3$

(e) $\mathbb{F}_{\{1,2,3\}}^{(3)} \cap [-8,8]^3$

Fig. 1: The ANOVA decomposition working on the hypercube $[-8,8]^3$ as a part of the 3-dimensional index set $\mathbb{Z}^3$.

3.1. Variance and Sensitivity. In order to get a notion of the importance of single terms compared to the entire function, we define the **variance of a function**

$$\sigma^2(f) := \int_{\mathbb{T}^d} \left(f(\boldsymbol{x}) - c_{\boldsymbol{0}}(f) \right)^2 \, d\boldsymbol{x}$$

for real-valued f. In this case, we have the equivalent formulation

$$\sigma^2(f) = \|f\|_{\mathrm{L}_2(\mathbb{T}^d)}^2 - |c_{\mathbf{0}}(f)|^2$$

which yields a sensible definition for complex-valued functions f. For the ANOVA terms $f_{\boldsymbol{u}}$ with $\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}$ we have $c_{\mathbf{0}}(f_{\boldsymbol{u}}) = 0$ and therefore

$$\sigma^2(f_{\boldsymbol{u}}) = \|f_{\boldsymbol{u}}\|_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})}^2.$$

LEMMA 3.9. *Let $f \in \mathrm{L}_2(\mathbb{T}^d)$. Then we obtain for the variance*

$$\sigma^2(f) = \sum_{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}} \sigma^2(f_{\boldsymbol{u}}).$$

Proof. We show that the right-hand side equals the left-hand side by employing Lemma 3.4 and Lemma 3.5

$$\sum_{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}} \sigma^2(f_{\boldsymbol{u}}) = \sum_{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} |c_{\boldsymbol{k}}(f)|^2 = \sum_{\boldsymbol{k} \in \bigcup_{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)}} |c_{\boldsymbol{k}}(f)|^2$$

$$= \sum_{\boldsymbol{k} \in \mathbb{Z}^d} |c_{\boldsymbol{k}}(f)|^2 - |c_{\mathbf{0}}(f)|^2 = \|f\|_{\mathrm{L}_2(\mathbb{T}^d)}^2 - |c_{\mathbf{0}}(f)|^2. \qquad \square$$

The global sensitivity indices

$$(3.6) \qquad\qquad \varrho(\boldsymbol{u}, f) := \frac{\sigma^2(f_{\boldsymbol{u}})}{\sigma^2(f)} \in [0, 1]$$

for $\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}$ provide a comparable score to rank the importance of ANOVA terms against each other, cf. [47, 48, 38]. Clearly, we have $\sum_{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}} \varrho(\boldsymbol{u}, f) = 1$ by Lemma 3.9.

We now introduce one notion of **effective dimensions** as proposed in [6]. Given a fixed $\delta \in (0, 1]$, the general notion of **superposition dimension** is defined as the minimum

$$\min \left\{ s \in \mathcal{D} : \sum_{\substack{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| \leq s}} \sigma^2(f_{\boldsymbol{u}}) \geq \delta \sigma^2(f) \right\}.$$

If we consider a particular Hilbert space $H \subseteq \mathrm{L}_2(\mathbb{T}^d)$ with norm $\|\cdot\|_H$, we modify the superposition dimension in the sense of this space, see e.g. [41]. For $f \in H$ and $\delta \in (0, 1]$ we define the modified superposition dimension as

$$(3.7) \qquad \mathrm{d}^{(\mathrm{sp})} := \min \left\{ s \in \mathcal{D} : \sup_{\|f\|_H \leq 1} \sum_{|\boldsymbol{u}| > s} \|f_{\boldsymbol{u}}\|_{\mathrm{L}_2(\mathbb{T}^d)}^2 \leq 1 - \delta \right\}.$$

Finally, we investigate how the smoothness of f translates to projections $\mathrm{P}_{\boldsymbol{u}} f$ and ANOVA terms $f_{\boldsymbol{u}}$. For a different setting this has been discussed in [38, 18, 19] and therein called *inheritance of smoothness.* In our setting, we express smoothness

through special subspaces of $L_2(\mathbb{T}^d)$ and how f being an element of those spaces translates to the projections $P_{\boldsymbol{u}}f$ and ANOVA terms $f_{\boldsymbol{u}}$. In particular, we look at **Sobolev type spaces**, cf. [32],

$$H^w(\mathbb{T}^d) := \left\{ f \in L_2(\mathbb{T}^d) \colon \|f\|_{H^w(\mathbb{T}^d)} := \left(\sum_{\boldsymbol{k} \in \mathbb{Z}^d} w^2(\boldsymbol{k}) \, |c_{\boldsymbol{k}}(f)|^2 \right)^{\frac{1}{2}} < \infty \right\}$$

and the **weighted Wiener algebra**

$$\mathcal{A}^w(\mathbb{T}^d) := \left\{ f \in L_1(\mathbb{T}^d) \colon \|f\|_{\mathcal{A}^w(\mathbb{T}^d)} := \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w(\boldsymbol{k}) \, |c_{\boldsymbol{k}}(f)| < \infty \right\}$$

with a weight function $w \colon \mathbb{Z}^d \to [1, \infty)$ for both cases.

THEOREM 3.10 (Inheritance of smoothness for Sobolev type spaces). *Let $f \in H^w(\mathbb{T}^d)$ with weight function $w \colon \mathbb{Z}^d \to [1, \infty)$. Then for any weight function $w_{\boldsymbol{u}} \colon \mathbb{Z}^{|\boldsymbol{u}|} \to [1, \infty)$ with*

$$w_{\boldsymbol{u}}(\boldsymbol{k}_{\boldsymbol{u}}) \leq w(\boldsymbol{k}) \, \forall \boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}$$

we have $P_{\boldsymbol{u}}f \in H^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})$ and $f_{\boldsymbol{u}} \in H^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})$.

Proof. We show that the norm $\|P_{\boldsymbol{u}}f\|_{H^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})}$ is finite by using Lemma 3.1

$$\sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} w_{\boldsymbol{u}}^2(\boldsymbol{\ell}) \, |c_{\ell}(P_{\boldsymbol{u}}f)|^2 = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} w_{\boldsymbol{u}}^2(\boldsymbol{k}_{\boldsymbol{u}}) \, |c_{\boldsymbol{k}}(f)|^2 \leq \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} w^2(\boldsymbol{k}) \, |c_{\boldsymbol{k}}(f)|^2$$

$$\leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w^2(\boldsymbol{k}) \, |c_{\boldsymbol{k}}(f)|^2 = \|f\|_{H^w(\mathbb{T}^d)}^2 < \infty.$$

Analogously, we employ Lemma 3.5 to prove $f_{\boldsymbol{u}} \in H^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})$

$$\sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} w_{\boldsymbol{u}}^2(\boldsymbol{\ell}) \, |c_{\ell}(f_{\boldsymbol{u}})|^2 = \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} w_{\boldsymbol{u}}^2(\boldsymbol{k}_{\boldsymbol{u}}) \, |c_{\boldsymbol{k}}(f)|^2 \leq \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} w^2(\boldsymbol{k}) \, |c_{\boldsymbol{k}}(f)|^2$$

$$\leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w^2(\boldsymbol{k}) \, |c_{\boldsymbol{k}}(f)|^2 = \|f\|_{H^w(\mathbb{T}^d)}^2 < \infty. \qquad \square$$

THEOREM 3.11 (Inheritance of smoothness for the weighted Wiener algebra). *Let $f \in \mathcal{A}^w(\mathbb{T}^d)$ with weight function $w \colon \mathbb{Z}^d \to [1, \infty)$. Then for any weight function $w_{\boldsymbol{u}} \colon \mathbb{Z}^{|\boldsymbol{u}|} \to [1, \infty)$ with*

$$w_{\boldsymbol{u}}(\boldsymbol{k}_{\boldsymbol{u}}) \leq w(\boldsymbol{k}) \, \forall \boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}$$

we have $P_{\boldsymbol{u}}f \in \mathcal{A}^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})$ and $f_{\boldsymbol{u}} \in \mathcal{A}^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})$.

Proof. We use Lemma 3.1 to show that $P_{\boldsymbol{u}}f \in \mathcal{A}^w(\mathbb{T}^{|\boldsymbol{u}|})$

$$\sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} w_{\boldsymbol{u}}(\boldsymbol{\ell}) \, |c_{\ell}(f_{\boldsymbol{u}})| = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} w_{\boldsymbol{u}}(\boldsymbol{k}_{\boldsymbol{u}}) \, |c_{k}(f)| \leq \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} w(\boldsymbol{k}) \, |c_{k}(f)|$$

$$\leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w(\boldsymbol{k}) \, |c_{k}(f)| = \|f\|_{\mathcal{A}^w(\mathbb{T}^d)} < \infty.$$

We utilize Lemma 3.5 to prove $f_{\boldsymbol{u}} \in \mathcal{A}^{w_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})$

$$
\sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} w_{\boldsymbol{u}}(\boldsymbol{\ell}) \, |c_{\ell}(f_{\boldsymbol{u}})| = \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} w_{\boldsymbol{u}}(\boldsymbol{k_u}) \, |c_k(f)| \leq \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} w(\boldsymbol{k}) \, |c_k(f)|
$$

$$
\leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w(\boldsymbol{k}) \, |c_k(f)| = \|f\|_{\mathcal{A}^w(\mathbb{T}^d)} < \infty. \qquad \square
$$

The inheritance of smoothness has special significance with regard to the numerical realization of the method presented in Section 5. It ensures that the ANOVA terms $f_{\boldsymbol{u}}$ are at least as smooth as the function f in consideration which is relevant for the quality of the approximation produced by the method.

4. Truncated ANOVA decomposition. The number of ANOVA terms of a function is equal to the cardinality of $\mathcal{P}(\mathcal{D}) = 2^d$ and therefore grows exponentially in the dimension. This reflects the curse of dimensionality in a certain way and poses a problem for the approximation of a function. In this section we consider truncating the ANOVA decomposition, i.e., removing certain terms $f_{\boldsymbol{u}}$, and therefore creating a certain form of sparsity. We define a **subset of ANOVA terms** as a subset of the power set of $\mathcal{D}$, i.e., $U \subseteq \mathcal{P}(D)$, such that the inclusion condition

$$(4.1) \qquad\qquad \boldsymbol{u} \in U \implies \forall \boldsymbol{v} \subseteq \boldsymbol{u} \colon \boldsymbol{v} \in U$$

holds, cf. [23, Chapter 3.2]. This is necessary due to the recursive definition of the ANOVA terms, see (3.3).

For any subset of ANOVA terms U we then define the truncated ANOVA decomposition as

$$
\mathrm{T}_U f := \sum_{\boldsymbol{u} \in U} f_{\boldsymbol{u}}.
$$

A specific truncation idea can be obtained by relating to the superposition dimension $\mathrm{d}^{(\mathrm{sp})}$, see (3.7). For a chosen superposition threshold $d_s \in \mathcal{D}$ (that may or may not be equal to the superposition dimension $\mathrm{d}^{(\mathrm{sp})}$), we define $U_{d_s} := \{\boldsymbol{u} \subseteq \mathcal{D} \colon |\boldsymbol{u}| \leq d_s\}$ and $\mathrm{T}_{d_s} := \mathrm{T}_{U_{d_s}}$. We subsequently prove properties of both T_U in general and T_{d_s} in particular.

LEMMA 4.1. *Let $f \in \mathrm{L}_2(\mathbb{T}^d)$ and $U \subseteq \mathcal{P}(D)$ be a subset of ANOVA terms. Then $\mathrm{T}_U f \in \mathrm{L}_2(\mathbb{T}^d)$ and for $\boldsymbol{k} \in \mathbb{Z}^d$ the Fourier coefficient is*

$$
c_{\boldsymbol{k}}(\mathrm{T}_U f) = \begin{cases} c_{\boldsymbol{k}}(f) & : \exists \boldsymbol{u} \in U \colon \boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)} \\ 0 & : otherwise. \end{cases}
$$

Proof. Clearly, we have $\mathrm{T}_U f \in \mathrm{L}_2(\mathbb{T}^d)$. Let now $\boldsymbol{k} \in \mathbb{Z}^d$. Then there exists a $\boldsymbol{u}_0 \subseteq \mathcal{D}$ such that $\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}_0}^{(d)}$. We employ Lemma 3.5 and obtain

$$
c_{\boldsymbol{k}}(\mathrm{T}_U f) = \int_{\mathbb{T}^d} \left(\sum_{\boldsymbol{u} \in U} f_{\boldsymbol{u}}(\boldsymbol{x_u}) \right) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \mathrm{d}\boldsymbol{x} = \sum_{\boldsymbol{u} \in U} \int_{\mathbb{T}^d} f_{\boldsymbol{u}}(\boldsymbol{x_u}) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \mathrm{d}\boldsymbol{x}
$$

$$
= \sum_{\boldsymbol{u} \in U} \int_{\mathbb{T}^{|\boldsymbol{u}|}} f_{\boldsymbol{u}}(\boldsymbol{x_u}) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k_u} \cdot \boldsymbol{x_u}} \mathrm{d}\boldsymbol{x_u} \, \delta_{\boldsymbol{k_u},\boldsymbol{0}} = \sum_{\boldsymbol{u} \in U} c_{\boldsymbol{k}}(f) \, \delta_{\boldsymbol{k_{u^c}},\boldsymbol{0}} \, (1 - \delta_{\boldsymbol{k_u},\boldsymbol{0}})
$$

$$
= \begin{cases} c_{\boldsymbol{k}}(f) & : \boldsymbol{u}_0 \in U \\ 0 & : \text{otherwise.} \end{cases} \qquad\qquad \square
$$

COROLLARY 4.2. *Let $f \in L_2(\mathbb{T}^d)$ and $d_s \in \mathcal{D}$ a superposition threshold. Then $T_{d_s} f \in L_2(\mathbb{T}^d)$ and only the Fourier coefficients corresponding to d_s-sparse frequencies are nonzero, i.e.,*

$$c_{\boldsymbol{k}}(T_{d_s} f) = \begin{cases} c_{\boldsymbol{k}}(f) & : \|\boldsymbol{k}\|_0 \leq d_s \\ 0 & : otherwise. \end{cases}$$

Proof. Since U_{d_s} is a subset of ANOVA terms, $T_{d_s} f \in L_2(\mathbb{T}^d)$ follows directly from Lemma 4.1. Moreover, $\exists \boldsymbol{u} \in U_{d_s} : \boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)} \iff \|\boldsymbol{k}\|_0 \leq d_s$. □

The following lemma shows that the number of terms in U_{d_s} is polynomial in d for a fixed d_s and therefore allows us to circumvent the curse of dimensionality in terms of the number of sets.

LEMMA 4.3. *We estimate the cardinality of $|U_{d_s}|$ as follows*

$$|U_{d_s}| < \left(\frac{ed}{d_s}\right)^{d_s},$$

i.e., the number of terms in U_{d_s} has polynomial growth in d for fixed $d_s \in \mathcal{D} \setminus \{d\}$.

Proof. We estimate the sum as follows

$$|U_{d_s}| = \sum_{n=0}^{d_s} \binom{d}{n} \leq \sum_{n=0}^{d_s} \frac{d^n d_s^n}{n! \, d_s^n} = \sum_{n=0}^{d_s} \left(\frac{d}{d_s}\right)^n \frac{d_s^n}{n!} \leq \left(\frac{d}{d_s}\right)^{d_s} \sum_{n=0}^{d_s} \frac{d_s^n}{n!}.$$

Estimating the sum by the Taylor series for e^{d_s} yields the statement. □

In the following we show direct formulas for the truncated ANOVA decomposition based on the projections similarly as for the ANOVA terms, see (3.4).

LEMMA 4.4. *Let $f \in L_2(\mathbb{T}^d)$ and $U \subseteq \mathcal{P}(D)$ a subset of ANOVA terms. Then we have the direct formula*

$$T_U f = \sum_{\boldsymbol{u} \in U} \sum_{\substack{\boldsymbol{v} \in U \\ \boldsymbol{u} \subseteq \boldsymbol{v}}} (-1)^{|\boldsymbol{v}| - |\boldsymbol{u}|} P_{\boldsymbol{u}} f.$$

Proof. We apply equation (3.4) and obtain immediately

$$T_U f = \sum_{\boldsymbol{u} \in U} f_{\boldsymbol{u}} = \sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} P_{\boldsymbol{v}} f = \sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{v} \in U} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} P_{\boldsymbol{v}} f \, \delta_{\boldsymbol{v} \subseteq \boldsymbol{u}}$$

$$= \sum_{\boldsymbol{v} \in U} \sum_{\substack{\boldsymbol{u} \in U \\ \boldsymbol{v} \subseteq \boldsymbol{u}}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} P_{\boldsymbol{v}} f. \qquad\qquad □$$

COROLLARY 4.5. *Let $f \in L_2(\mathbb{T}^d)$ and $d_s \in \mathcal{D}$ a superposition threshold. Then we have the direct formula*

$$T_{d_s} f = \sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} \left[\sum_{n=|\boldsymbol{u}|}^{d_s} (-1)^{n - |\boldsymbol{u}|} \binom{d - |\boldsymbol{u}|}{n - |\boldsymbol{u}|} \right] P_{\boldsymbol{u}} f.$$

Proof. Since the equality

$$\sum_{\substack{\boldsymbol{v}\in U_{d_s}\\ \boldsymbol{u}\subseteq\boldsymbol{v}}}(-1)^{|\boldsymbol{v}|-|\boldsymbol{u}|} = \sum_{n=|\boldsymbol{u}|}^{d_s}(-1)^{n-|\boldsymbol{u}|}\binom{d-|\boldsymbol{u}|}{n-|\boldsymbol{u}|},$$

holds, we employ Lemma 4.4 and the formula is proven. □

The truncated ANOVA decomposition plays a major role in our approximation approach presented in Section 5. Therefore we are interested in functions that can be approximated well by a truncated ANOVA decomposition. Specifically, we are looking to characterize functions such that the truncation operation by $\mathrm{T}_U f$ for different sets U retains most of the function, i.e., we have a relative error

$$(4.2) \qquad \frac{\|f-\mathrm{T}_U f\|_{H_1}}{\|f\|_{H_2}} < \varepsilon$$

with $\varepsilon > 0$, and H_1, H_2 certain subspaces of $\mathrm{L}_2(\mathbb{T}^d)$. It is especially interesting to characterize these functions by properties like the smoothness. To this end, we start by proving general bounds for Sobolev type spaces $\mathrm{H}^w(\mathbb{T}^d)$ and the weighted Wiener algebra $\mathcal{A}^w(\mathbb{T}^d)$ to later relate this to weight functions w defined by specific kinds of smoothness.

Moreover, this can be related to the superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ for a $\delta \in (0,1]$, see (3.7). Let $H_1 = \mathrm{L}_2(\mathbb{T}^d)$ and $H_2 \in \{\mathrm{H}^w(\mathbb{T}^d), \mathcal{A}^w(\mathbb{T}^d)\}$ for a weight function w. If we choose truncation by a superposition threshold $d_s \in \mathcal{D}$ then the bound on the right-hand side $\varepsilon(d_s) \in (0,1)$ depends on d_s. Moreover, we have

$$(4.3) \qquad \sup_{f\neq 0}\frac{\|f-\mathrm{T}_{d_s}f\|^2_{\mathrm{L}_2(\mathbb{T}^d)}}{\|f\|^2_{H_2}} = \sup_{\|f\|_{H_2}\leq 1}\sum_{|\boldsymbol{u}|>d_s}\|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^d)} < \varepsilon(d_s)$$

which follows from $\|f-\mathrm{T}_{d_s}f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \sum_{|\boldsymbol{u}|>d_s}\|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^d)}$. The modified superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ will now be smaller or equal to $\min\{d_s \in \mathcal{D}: \varepsilon(d_s) \leq 1-\delta\}$, i.e., truncation by this minimum as superposition threshold is guaranteed to be effective in relation to δ.

THEOREM 4.6. *Let $f \in \mathrm{H}^w(\mathbb{T}^d)$ with weight function $w\colon \mathbb{Z}^d \to [1,\infty)$. Then*

$$\frac{\|f-\mathrm{T}_U f\|_{\mathrm{L}_2(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}} \leq \frac{1}{\min_{\substack{\boldsymbol{k}\in\bigcup_{\substack{\boldsymbol{u}\subseteq\mathcal{D}\\\boldsymbol{u}\notin U}}\mathbb{F}_{\boldsymbol{u}}^{(d)}}}w(\boldsymbol{k})}.$$

Proof. We employ Parseval's identity and Lemma 4.1 to derive

$$\|f-\mathrm{T}_U f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\mathrm{c}_{\boldsymbol{k}}(f)-\mathrm{c}_k(\mathrm{T}_U f)|^2 = \sum_{\substack{\boldsymbol{k}\in\bigcup_{\substack{\boldsymbol{u}\subseteq\mathcal{D}\\\boldsymbol{u}\notin U}}\mathbb{F}_{\boldsymbol{u}}^{(d)}}}|\mathrm{c}_{\boldsymbol{k}}(f)|^2$$

$$= \sum_{\substack{\boldsymbol{k}\in\bigcup_{\substack{\boldsymbol{u}\subseteq\mathcal{D}\\\boldsymbol{u}\notin U}}\mathbb{F}_{\boldsymbol{u}}^{(d)}}}\frac{w^2(\boldsymbol{k})}{w^2(\boldsymbol{k})}|\mathrm{c}_{\boldsymbol{k}}(f)|^2$$

$$\leq \frac{1}{\min_{\substack{\boldsymbol{k}\in\bigcup_{\substack{\boldsymbol{u}\subseteq\mathcal{D}\\\boldsymbol{u}\notin U}}\mathbb{F}_{\boldsymbol{u}}^{(d)}}}w^2(\boldsymbol{k})}\|f\|^2_{\mathrm{H}^w(\mathbb{T}^d)}. \qquad \square$$

THEOREM 4.7. *Let $f \in \mathcal{A}^w(\mathbb{T}^d)$ with weight function $w \colon \mathbb{Z}^d \to [1, \infty)$. Then*

$$\frac{\|f - \mathrm{T}_U f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathcal{A}^w(\mathbb{T}^d)}} \leq \frac{1}{\min_{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ \boldsymbol{u} \notin U}} \mathbb{F}_{\boldsymbol{u}}^{(d)}} w(\boldsymbol{k})}.$$

For $f \in \mathrm{H}^w(\mathbb{T}^d)$ with a weight function $w \colon \mathbb{Z}^d \to [1, \infty)$ such that $\{1/w(\boldsymbol{k})\}_{\boldsymbol{k} \in \mathbb{Z}^d} \in \ell_2$ we have

$$\|f - \mathrm{T}_U f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq \sqrt{\sum_{\substack{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ \boldsymbol{u} \notin U}} \mathbb{F}_{\boldsymbol{u}}^{(d)}}} \frac{1}{w^2(\boldsymbol{k})}} \, \|f\|_{\mathrm{H}^w(\mathbb{T}^d)}.$$

Proof. We estimate the L_∞-norm by the sum of the absolute values of the Fourier coefficients and then use Lemma 4.1

$$\|f - \mathrm{T}_U f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} |c_{\boldsymbol{k}}(f) - c_k(\mathrm{T}_U f)| = \sum_{\substack{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ \boldsymbol{u} \notin U}} \mathbb{F}_{\boldsymbol{u}}^{(d)}}} |c_{\boldsymbol{k}}(f)|$$

$$(4.4) \qquad\qquad = \sum_{\substack{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ \boldsymbol{u} \notin U}} \mathbb{F}_{\boldsymbol{u}}^{(d)}}} \frac{w(\boldsymbol{k})}{w(\boldsymbol{k})} |c_{\boldsymbol{k}}(f)|$$

$$\leq \frac{1}{\min_{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ \boldsymbol{u} \notin U}} \mathbb{F}_{\boldsymbol{u}}^{(d)}} w(\boldsymbol{k})} \|f\|_{\mathcal{A}^w(\mathbb{T}^d)}.$$

Employing the Cauchy-Schwarz inequality in (4.4) instead of extracting the minimum yields

$$\|f - \mathrm{T}_U f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq \sqrt{\sum_{\substack{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ \boldsymbol{u} \notin U}} \mathbb{F}_{\boldsymbol{u}}^{(d)}}} \frac{1}{w^2(\boldsymbol{k})}} \, \|f\|_{\mathrm{H}^w(\mathbb{T}^d)}.$$

The condition $\{1/w(\boldsymbol{k})\}_{\boldsymbol{k} \in \mathbb{Z}^d} \in \ell_2$ assures that the sum which appears in the bound is finite. □

In the following, we relate the truncation of f by the operator T_{d_s} with the smoothness of f. To this end, we introduce the weights

$$(4.5) \qquad\qquad w^{\alpha,\beta}(\boldsymbol{k}) := \gamma_{\mathrm{supp}\,\boldsymbol{k}}^{-1} (1 + \|\boldsymbol{k}\|_1)^\alpha \prod_{s \in \mathrm{supp}\,\boldsymbol{k}} (1 + |k_s|)^\beta$$

with $\mathrm{supp}\,\boldsymbol{k} = \{i \in \mathcal{D} \colon k_i \neq 0\}$ and parameters $\beta \geq 0$, and $\alpha > -\beta$. The parameters α, β, and the weight $\gamma_{\boldsymbol{u}}$, $\boldsymbol{u} \subseteq \mathcal{D}$, regulate the decay of the Fourier coefficients. Specifically, the parameter α is regulating the isotropic smoothness and β the dominating mixed smoothness, cf. [7]. Moreover, $\boldsymbol{\gamma}$ controls the influence of the different dimensions. We choose a POD (*product and order-dependent*) structure for $\gamma_{\boldsymbol{u}}$ such that

$$(4.6) \qquad\qquad \gamma_{\boldsymbol{u}} = \Gamma_{|\boldsymbol{u}|} \prod_{s \in \boldsymbol{u}} \gamma_s,$$

where $\mathbf{\Gamma} \in (0,1]^d$ is nonincreasing and $\boldsymbol{\gamma} = (\gamma_i)_{i=1}^d \in (0,1]^d$. The POD structure is motivated by the application of quasi-Monte Carlo methods for PDEs with random coefficients, cf. [35, 13, 33, 14]. Similar weights for isotropic and dominating mixed smoothness have been considered in [16, 26, 5]. Moreover, the Sobolev type spaces may also be referred to as weighted Korobov spaces, cf. [46] for product weights and [9] for general weights.

We now use the previously obtained bounds for general weight functions w and derive results for the weights $w^{\alpha,\beta}$ from (4.5). We focus on the subsets of ANOVA terms U_{d_s} defined by a superposition threshold $d_s \in \mathcal{D}$.

COROLLARY 4.8. *Let* $f \in \mathcal{A}^{w^{\alpha,\beta}}(\mathbb{T}^d)$ *with weight function from* (4.5) *with POD structure* (4.6), $\beta \geq 0$, $\alpha > -\beta$, $\mathbf{\Gamma} \in (0,1]^d$, *and* $\boldsymbol{\gamma} \in (0,1]^d$. *Then*

$$(4.7) \qquad \frac{\|f - \mathrm{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathcal{A}^{w^{\alpha,\beta}}(\mathbb{T}^d)}} \leq \Gamma_{d_s+1} (2 + d_s)^{-\alpha} 2^{-\beta(d_s+1)} \prod_{s=1}^{d_s+1} \gamma_s^*$$

where $\boldsymbol{\gamma}^*$ *is the non-increasing rearrangement of* $\boldsymbol{\gamma}$.

Proof. We use Theorem 4.7 and calculate the bound for the weight function $w^{\alpha,\beta,\boldsymbol{\gamma}}$ by computing the minimum

$$M := \min_{\substack{\boldsymbol{k} \in \mathbb{Z}^d \\ \|\boldsymbol{k}\|_0 > d_s}} \Gamma_{\|\boldsymbol{k}\|_0}^{-1} (1 + \|\boldsymbol{k}\|_1)^\alpha \prod_{s=1}^d (1 + |k_s|)^\beta \prod_{s \in \mathrm{supp}\,\boldsymbol{k}} \gamma_s^{-1}.$$

Since $\mathbf{\Gamma}$ is non-increasing by definition, $\Gamma_{d_s+1}^{-1}$ has to be equal to the smallest value. The frequencies in $\mathbb{F}_{\boldsymbol{u}}^{(d)}$ have exactly $|\boldsymbol{u}|$ nonzero entries, therefore we get

$$M = \Gamma_{d_s+1}^{-1} (1 + d_s + 1)^\alpha (1+1)^{\beta(d_s+1)} \min_{\substack{\boldsymbol{k} \in \mathbb{Z}^d \\ \|\boldsymbol{k}\|_0 > d_s}} \prod_{s \in \boldsymbol{u}} \gamma_s^{-1}.$$

The remaining product becomes minimal for the product of the $d_s + 1$ smallest entries in $\boldsymbol{\gamma}$ which yields the statement. □

LEMMA 4.9. *Let* $n \in \mathcal{D}$ *and* $\boldsymbol{\gamma} \in (0,1]^d$. *Then*

$$\sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| = n}} \prod_{s \in \boldsymbol{u}} \gamma_s^2 \leq \|\boldsymbol{\gamma}\|_2^{2n}.$$

Proof. We rewrite the sum as follows

$$\sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| = n}} \prod_{s \in \boldsymbol{u}} \gamma_s^2 = \sum_{i_1=1}^d \gamma_{i_1}^2 \sum_{i_2=i_1+1}^d \gamma_{i_2}^2 \cdots \sum_{i_n=i_{n-1}+1}^d \gamma_{i_n}^2.$$

Then every single sum can be estimated by $\|\boldsymbol{\gamma}\|_2^2$, i.e.,

$$\sum_{i_j=i_{j-1}+1}^d \gamma_{i_j}^2 \leq \sum_{i_j=1}^d \gamma_{i_j}^2 = \|\boldsymbol{\gamma}\|_2^2$$

for $j \in \{2, 3, \ldots, d\}$ with equality for $j = 1$. □

COROLLARY 4.10. *Let $f \in \mathrm{H}^{w^{\alpha,\beta}}(\mathbb{T}^d)$ with weight function from* (4.5) *with POD structure* (4.6), $\beta \geq 0$, $\alpha > -\beta$, $\boldsymbol{\Gamma} \in (0,1]^d$, *and* $\boldsymbol{\gamma} \in (0,1]^d$. *Then*

$$
\text{(4.8)} \qquad \frac{\|f - \mathrm{T}_{d_s} f\|_{\mathrm{L}_2(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^{w^{\alpha,\beta}}(\mathbb{T}^d)}} \leq \Gamma_{d_s+1} (2 + d_s)^{-\alpha} 2^{-\beta(d_s+1)} \prod_{s=1}^{d_s+1} \gamma_s^*
$$

where $\boldsymbol{\gamma}^* = (\gamma_s^*)_{s=1}^d$ *is the non-increasing rearrangement of* $\boldsymbol{\gamma}$. *For functions with isotropic smoothness* $\alpha = 0$ *and dominating mixed smoothness* $\beta > 1/2$ *we have*

$$
\frac{\|f - \mathrm{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^{w^{\alpha,\beta}}(\mathbb{T}^d)}} \leq \sqrt{\sum_{n=d_s+1}^{d} 2^n \Gamma_n^2 (\zeta(2\beta) - 1)^n \|\boldsymbol{\gamma}\|_2^{2n}}
$$

where ζ *is the Riemann zeta function. Exponential decay for* Γ_s, *i.e.,* $\Gamma_s = c^s$, $0 < c \leq 1$, *such that the condition*

$$
\text{(4.9)} \qquad \|\boldsymbol{\gamma}\|_2 < \frac{1}{c\sqrt{2\zeta(2\beta) - 2}}
$$

holds, yields the bound

$$
\text{(4.10)} \qquad \frac{\|f - \mathrm{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^{w^{\alpha,\beta}}(\mathbb{T}^d)}} \leq \frac{\left(c \|\boldsymbol{\gamma}\|_2 \sqrt{2\zeta(2\beta) - 2}\right)^{d_s+1}}{\sqrt{1 - 2c^2 \|\boldsymbol{\gamma}\|_2^2 (\zeta(2\beta) - 1)}}.
$$

Proof. The bound from statement (4.8) is a consequence of Theorem 4.6 and can be calculated analogously to the proof of Corollary 4.8. For the second statement, we calculate the constant in the bound from Theorem 4.7. We use Lemma 3.4 and the product structure of the weights $w^{\alpha,\beta}(\boldsymbol{k})$ to obtain
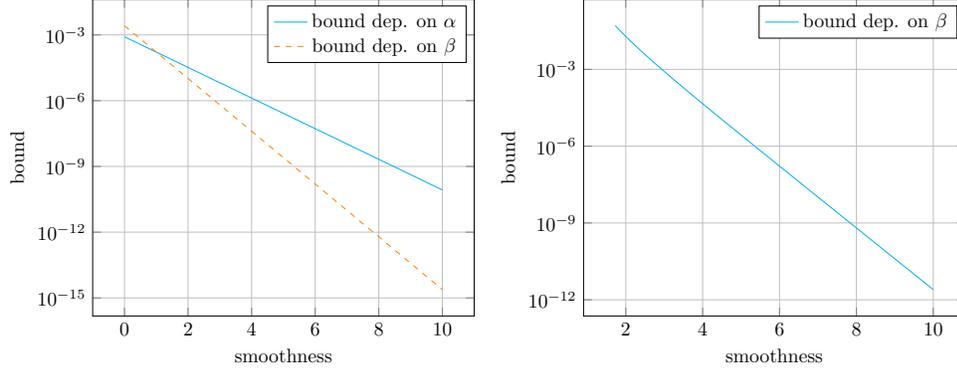
$$
\sum_{\substack{\boldsymbol{k} \in \bigcup_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| > d_s}} \mathbb{F}_{\boldsymbol{u}}^{(d)}}} \frac{1}{w^2(\boldsymbol{k})} = \sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| > d_s}} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} \frac{1}{\Gamma_{|\boldsymbol{u}|}^{-2} (1 + |k_s|)^{2\beta} \prod_{s \in \boldsymbol{u}} \gamma_s^{-2}}
$$

$$
= \sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| > d_s}} \Gamma_{\boldsymbol{u}}^2 \sum_{\boldsymbol{k} \in (\mathbb{Z}\backslash\{0\})^{|\boldsymbol{u}|}} \frac{1}{\left(\prod_{s \in \boldsymbol{u}} \gamma_s^{-2}\right) \left(\prod_{s=1}^{|\boldsymbol{u}|} (1 + |k_s|)^{2\beta}\right)}
$$

$$
= \sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| > d_s}} \Gamma_{|\boldsymbol{u}|}^2 \prod_{s \in \boldsymbol{u}} \gamma_s^2 \sum_{k \in \mathbb{Z}\backslash\{0\}} \frac{1}{(1 + |k|)^{2\beta}}.
$$

We find an explicit form by replacing the sums with the Riemann zeta function

$$
\prod_{s \in \boldsymbol{u}} \gamma_s^2 \sum_{k \in \mathbb{Z}\backslash\{0\}} \frac{1}{(1 + k)^{2\beta}} = \prod_{s \in \boldsymbol{u}} 2\gamma_s^2 \sum_{k \in \mathbb{N}} \frac{1}{(1 + k)^{2\beta}} = 2^{|\boldsymbol{u}|} (\zeta(2\beta) - 1)^{|\boldsymbol{u}|} \prod_{s \in \boldsymbol{u}} \gamma_s^2.
$$

Applying Lemma 4.9 then gives us the upper bound

$$
\sum_{n=d_s+1}^{d} 2^n \Gamma_n^2 (\zeta(2\beta) - 1)^n \sum_{\substack{\boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| = n}} \prod_{s \in \boldsymbol{u}} \gamma_s^2 \leq \sum_{n=d_s+1}^{d} 2^n \Gamma_n^2 (\zeta(2\beta) - 1)^n \|\boldsymbol{\gamma}\|_2^{2n}.
$$

(a) bounds (4.8), (4.7) for $0 \leq \alpha \leq 10$, $\beta = 1$ (b) bound (4.10) with $\alpha = 0$, $1.7287 \leq \beta \leq 10$
(solid), and for $\alpha = 1$, $0 \leq \beta \leq 10$ (dashed)

Fig. 2: Decay of errors from (4.8), (4.10), and (4.7) in relation to their isotropic smoothness α and dominating-mixed smoothness β with $d = 9$, $d_s = 3$, dimension dependent coefficients $\boldsymbol{\gamma} = (1/s)_{s=1}^{9}$ and order dependent coefficients $\boldsymbol{\Gamma} = (\pi^{-s}\sqrt{3}^{s})_{s=1}^{9}$.

If we choose an exponential decay for Γ_n, i.e., $\Gamma_n := c^n, 0 < c \leq 1$, the explicit upper bound becomes

$$\sum_{n=d_s+1}^{d} 2^n c^{2n} \left(\zeta(2\beta) - 1\right)^n \|\boldsymbol{\gamma}\|_2^{2n} = \frac{q^{d_s+1}}{1-q}(1 - q^{d-d_s})$$

where $q := 2c^2 \left(\zeta(2\beta) - 1\right) \|\boldsymbol{\gamma}\|_2^2$ with $0 < q < 1$ because of the condition (4.9).   □

The bound in Corollary 4.8 and (4.8) in Corollary 4.10 are independent of the spatial dimensions d of the functions f as long as they have the same superposition threshold and the norm stays the same. This allows us to circumvent the curse of dimensionality here and use the ANOVA terms in U_{d_s} for a superposition threshold $d_s \in \mathcal{D}$. The bound (4.10) can also be considered for $d \to \infty$. The dependence on the dimension d is contained within the norm $\|\boldsymbol{\gamma}\|_2^2$. Choosing a square-summable sequence $\{\gamma_\ell\}_{\ell \in \mathbb{N}}$ results in an upper bound for $\|\boldsymbol{\gamma}\|_2$ for any $d \to \infty$. In this case the bound can be made independent of d by the condition (4.9).

Figure 2 shows the different bounds for weights $w^{\alpha,\beta}$ with $\boldsymbol{\gamma} = (1/s)_{s=1}^{9}$ and $\boldsymbol{\Gamma} = (\pi^{-s}\sqrt{3}^{s})_{s=1}^{9}$, see (4.5). With regard to the superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ for $\mathrm{H}^{w^{\alpha,\beta}}(\mathbb{T}^d)$, cf. (3.7), one may interpret this as follows: Given $f \in \mathrm{H}^{w^{\alpha,\beta}}(\mathbb{T}^d)$, the value $\varepsilon(\alpha,\beta) \in (0,1)$ of the bound in part (a) of Figure 2 tells us that for $\delta = 1 - \varepsilon(\alpha,\beta)^2$ the superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ is smaller or equal to the superposition threshold $d_s = 3$, e.g., $\varepsilon(0,1) \approx 0.0008$ and therefore $\delta = 0.99999936$.

5. ANOVA approximation method. We consider the general problem of approximating a periodic function $f \colon \mathbb{T}^d \to \mathbb{C}$ given certain function evaluations of f. Specifically, we distinguish two approximation scenarios – black-box approximation and scattered data approximation. In the case of black-box approximation, we are able to evaluate f at any given point $\boldsymbol{x} \in \mathbb{T}^d$. Since the evaluations come at a certain cost, we aim to keep them minimal or require a certain trade-off. For scattered data approximation we have a finite set of nodes $X \subseteq \mathbb{T}^d$ and know the function values

$\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x} \in X}$. Here, one cannot add more nodes to X or choose the locations of the nodes. Both scenarios have a high relevance for problems in various applications.

In this section, we consider an approximation scheme for high-dimensional, periodic functions of a low-dimensional structure, i.e., functions with a small superposition dimension $\mathrm{d}^{(\mathrm{sp})} \in \mathcal{D}$ for a $\delta \in (0,1]$ that is close to one, cf. (3.7). In this case the truncation by T_{d_s} with a small superposition threshold $d_s \in \mathcal{D}$ will be effective. It has been observed that functions in many practical applications belong to such a class, see e.g. [6]. In Section 4 we have considered errors for functions of dominating-mixed and isotropic smoothness defined trough the decay of the Fourier coefficients and therefore obtained an upper bound for the modified superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ from (3.7). Considering Figure 2, we know that e.g. POD weights lead to a decay such that the functions are of a low-dimensional structure.

The approximation scheme can be viewed in both approximation scenarios although the details are different. We work for now with the node set X as well as function evaluations $\boldsymbol{y}$ and keep in mind that X may also be chosen if we are in the black-box case. The first step is to reduce the ANOVA decomposition to the terms in U_{d_s}, i.e., we approximate

$$f \approx \mathrm{T}_{d_s} f = \sum_{\boldsymbol{u} \in U_{d_s}} f_{\boldsymbol{u}}.$$

The Fourier coefficients $c_{\boldsymbol{k}}(\mathrm{T}_{d_s} f)$ can only be nonzero if the frequency $\boldsymbol{k}$ is at most d_s-sparse, i.e., $\|\boldsymbol{k}\|_0 \le d_s$, see Corollary 4.2. Based on this, we aim to approximate f by a Fourier partial sum $S_I f$ with a finite index set

$$(5.1) \qquad\qquad I \subseteq \left\{ \boldsymbol{k} \in \mathbb{Z}^d \colon \operatorname{supp} \boldsymbol{k} \in U_{d_s} \right\}.$$

The challenge is to determine an appropriate index set I. To this end, we employ a special scheme to determine frequency locations based on the ANOVA terms and an importance ranking on them.

We call the first step *active set detection* and its aim is to determine an importance ranking on the terms $f_{\boldsymbol{u}}$ with $\boldsymbol{u} \in U_{d_s}$ based on the global sensitivity indices $\varrho(\boldsymbol{u}, f)$, cf. (3.6). This information is also highly relevant to interpret relations in our data X and $\boldsymbol{y}$.

Based on the sensitivity indices we build an active set of ANOVA terms $U \subseteq U_{d_s}$. This relates to the importance of frequencies and therefore information on how to choose the index set I from (5.1). Reducing the number of ANOVA terms and in turn the number of frequencies leads to a reduction of the model complexity. The effects of overfitting are therefore lessened. In Subsection 5.1 we consider the details of the active set detection and in Subsection 5.3 the approximation with an active set as well as approximation errors.

5.1. Active set detection. The method assumes that the underlying function f is of a low-dimensional structure, i.e., $f \approx \mathrm{T}_{d_s} f$ for some superposition threshold $d_s \in \mathcal{D}$. The goal in the active set detection step is to determine an importance ranking for the ANOVA terms. In order to do this, we choose an appropriate search index set. Since we have no a-priori knowledge about the importance of the ANOVA terms or the smoothness of the function f, we work with order-dependent finite index sets $I_0 = \{0\}, I_1 \subseteq (\mathbb{Z} \setminus \{0\}), \dots, I_{d_s} \subseteq (\mathbb{Z} \setminus \{0\})^{d_s}$. This achieves that two ANOVA terms $f_{\boldsymbol{u}}$ and $f_{\boldsymbol{v}}$ with $|\boldsymbol{u}| = |\boldsymbol{v}|$ are supported on equivalent index sets. We then use

the projection operator

$$(5.2) \qquad \mathrm{P}_{\boldsymbol{u}} I := \{ \boldsymbol{k} \in \mathbb{Z}^d \colon \boldsymbol{k}_{\boldsymbol{u}} \in I, \boldsymbol{k}_{\boldsymbol{u}^\mathrm{c}} = \boldsymbol{0} \}$$

to project the index sets and obtain

$$(5.3) \qquad I(U_{d_s}) = \bigcup_{\boldsymbol{u} \in U_{d_s}} \mathrm{P}_{\boldsymbol{u}} I_{|\boldsymbol{u}|}.$$

This leads to the approximation by a Fourier partial sum

$$(5.4) \qquad f(\boldsymbol{x}) \approx \mathrm{T}_{d_s} f(\boldsymbol{x}) \approx S_{I(U_{d_s})} f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in I(U_{d_s})} \mathrm{c}_{\boldsymbol{k}}(f)\, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}.$$

The Fourier coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$ in (5.4) are unknown and we aim to determine approximations for them from the data X and $\boldsymbol{y}$. To this end, we consider the least-squares problem

$$(5.5) \qquad \hat{\boldsymbol{f}}_{\mathrm{sol}} = \underset{\hat{\boldsymbol{f}} \in \mathbb{C}^{\left| I(U_{d_s}) \right|}}{\arg\min} \left\| \boldsymbol{y} - \boldsymbol{F}_{I(U_{d_s})} \hat{\boldsymbol{f}} \right\|_2^2$$

with Fourier matrix $\boldsymbol{F}_{I(U_{d_s})} = \left(\mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \right)_{\boldsymbol{x} \in X, \boldsymbol{k} \in I(U_{d_s})}$. If the Fourier matrix has full rank, the elements of the solution vector $\hat{\boldsymbol{f}}_{\mathrm{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k} \in I(U_{d_s})}$ are the unique least-squares approximation to the Fourier coefficients, i.e., $\hat{f}_{\boldsymbol{k}} \approx \mathrm{c}_{\boldsymbol{k}}(f)$, with respect to X and $\boldsymbol{y}$. Depending on the approximation scenario, there are different methods of solving least-squares problems of the type (5.5). We refer to Subsection 5.2 for details.

We use the approximate Fourier coefficients $\hat{f}_{\boldsymbol{k}}$ to build the approximate Fourier partial sum

$$(5.6) \qquad S_{I(U_{d_s})} f(\boldsymbol{x}) \approx S_{I(U_{d_s})}^X f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in I(U_{d_s})} \hat{f}_{\boldsymbol{k}}\, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}$$

which provides an initial approximation to the function f. In order to achieve a Fourier matrix $\boldsymbol{F}_{I(U_{d_s})}$ with full rank and combat the effects of overfitting, we may need to severely limit the number of frequencies in the order-dependent sets I_1, I_2, $\ldots$, I_{d_s}. Details on this will be considered in the following subsections for the specific approximation scenarios.

In order to determine an importance ranking on the ANOVA terms, we assume that the global sensitivity indices of $S_{I(U_{d_s})}^X f$ and f behave similarly, i.e., it holds that

$$(5.7) \qquad \varrho(\boldsymbol{u}_1, S_{I(U_{d_s})}^X f) \le \varrho(\boldsymbol{u}_2, S_{I(U_{d_s})}^X f) \implies \varrho(\boldsymbol{u}_1, f) \le \varrho(\boldsymbol{u}_2, f)$$

for $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U_{d_s}$. This allows us to use a threshold vector $\boldsymbol{\varepsilon} \in [0,1]^{d_s}$ to define an active set of ANOVA terms that only contains the *important* terms with respect to $\boldsymbol{\varepsilon}$

$$(5.8) \qquad U_{X,\boldsymbol{y}}^{(\boldsymbol{\varepsilon})} := \{ \boldsymbol{v} \subseteq \mathcal{D} \colon \exists \boldsymbol{u} \in U_{d_s} \colon \boldsymbol{v} \subseteq \boldsymbol{u} \text{ and } \varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f) > \varepsilon_{|\boldsymbol{u}|} \}.$$

The inclusion condition (4.1) is fulfilled by definition. We reduce the ANOVA decomposition to this set of terms to determine an approximation for f in Subsection 5.3.

5.2. Least-squares approximation. In this section, we discuss the solution of least-squares problems of the form

$$(5.9) \qquad \min_{\hat{\boldsymbol{f}} \in \mathbb{C}^{|I(U)|}} \left\| \boldsymbol{y} - \boldsymbol{F}_{I(U)} \hat{\boldsymbol{f}} \right\|_2^2$$

with a Fourier matrix $\boldsymbol{F}_{I(U)} = \left(\mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \right)_{\boldsymbol{x} \in X, \boldsymbol{k} \in I(U)}$. Here, U is an arbitrary subset of ANOVA terms and for each term we have a given finite frequency index set $I_{\boldsymbol{u}} \subseteq (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}$. The set

$$(5.10) \qquad I(U) = \bigcup_{\boldsymbol{u} \in U} \mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}}$$

is obtained through the projections (5.2).

The following remark shows that the Fourier matrix can be structured with respect to the ANOVA terms. Moreover, we can decompose the matrix-vector multiplications with both, $\boldsymbol{F}_{I(U)}$ and its adjoint $\boldsymbol{F}_{I(U)}^*$.

Remark 5.1. Let $\boldsymbol{F}_{I(U)}$ be a Fourier matrix with respect to a node set X and an index set $I(U)$ with a subset of ANOVA terms $U \subseteq \mathcal{P}(D)$ and index sets $I_{\boldsymbol{u}} \subseteq (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}$, $\boldsymbol{u} \in U$. Then

$$\boldsymbol{F} \hat{\boldsymbol{f}} = \left(\boldsymbol{F}_{\boldsymbol{u}_1} \ \boldsymbol{F}_{\boldsymbol{u}_2} \ \cdots \ \boldsymbol{F}_{\boldsymbol{u}_n} \right) \hat{\boldsymbol{f}}$$

where $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ with $n = |U|$ is a numbering of the subsets of coordinate indices in U such that $\hat{\boldsymbol{f}} = \left(\hat{\boldsymbol{f}}_{\boldsymbol{u}_1} \ \hat{\boldsymbol{f}}_{\boldsymbol{u}_2} \ \cdots \ \hat{\boldsymbol{f}}_{\boldsymbol{u}_n} \right)^{\top}$. The Fourier matrices are $\boldsymbol{F}_{\boldsymbol{u}} = \left(\mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}} \right)_{\boldsymbol{x} \in X, \boldsymbol{\ell} \in I_{\boldsymbol{u}}}$. The matrix-vector product with $\boldsymbol{F}$ can therefore be decomposed as

$$\boldsymbol{F} \hat{\boldsymbol{f}} = \sum_{\boldsymbol{u} \in U} \boldsymbol{F}_{\boldsymbol{u}} \hat{\boldsymbol{f}}_{\boldsymbol{u}}$$

with vector components $\hat{\boldsymbol{f}}_{\boldsymbol{u}}$. For the adjoint product $\boldsymbol{F}^* \boldsymbol{f}$ with a vector $\boldsymbol{f} \in \mathbb{C}^{|X|}$ we obtain the result $\hat{\boldsymbol{a}} \in \mathbb{C}^{|I(U)|}$ by computing the products

$$\hat{\boldsymbol{a}}_{\boldsymbol{u}} = \boldsymbol{F}_{\boldsymbol{u}}^* \boldsymbol{f}, \quad \forall \boldsymbol{u} \in U.$$

Then we have the result vector $\hat{\boldsymbol{a}} = \left(\hat{\boldsymbol{a}}_{\boldsymbol{u}_1} \ \hat{\boldsymbol{a}}_{\boldsymbol{u}_2} \ \cdots \ \hat{\boldsymbol{a}}_{\boldsymbol{u}_n} \right)^{\top}$.

5.2.1. Black-box scenario. In the case of black-box approximation, i.e., the set X can be chosen, we have to determine an appropriate special discretization for index sets of the type $I(U)$. Here, we have different possibilities. One might think of rank-1 lattices that have been used for integration before, see e.g. [8], and approximation, see e.g. [26, 30]. For a general introduction to lattice rules, we refer to Subsection 2.1. Sparse grid sampling related to the Smolyak algorithm is a further possibility, cf. [15, 21, 22, 23].

In the following, we focus on using reconstructing single rank-1 lattice for function approximation. If we have a reconstructing single rank-1 lattice $\Lambda(\boldsymbol{z}, M, I(U)) \subseteq \mathbb{Z}^d$ for a generating vector $\boldsymbol{z} \in \mathbb{Z}^d$ and size $M \in \mathbb{N}$ with respect to an index set $I(U)$, then

$$(5.11) \qquad \boldsymbol{F}_{I(U)}^* \boldsymbol{F}_{I(U)} = M \cdot \mathrm{I}$$

with I the identity matrix, see [43, Chapter 8.2]. Then the solution to problem (5.9) is unique and given by the multiplication of the Moore-Penrose inverse $\boldsymbol{F}_{I(U)}^{\dagger}$ with $\boldsymbol{y}$, see e.g. [3]. Through the property (5.11) the Moore-Penrose inverse is simplified to

$$(5.12) \qquad \boldsymbol{F}_{I(U)}^{\dagger} = \frac{1}{M} \boldsymbol{F}_{I(U)}^{*},$$

i.e., a multiplication with the adjoint matrix. This allows us to efficiently compute approximations for the Fourier coefficients of f if the nodes form a reconstructing rank-1 lattice.

It remains the issue of determining such a reconstructing rank-1 lattice given an index set of type $I(U)$. In [43, Theorem 8.16] it was shown that reconstructing lattices exist if the lattice size M is sufficiently large. Since the evaluations of f come at a certain cost, it is necessary to consider the lattice size for our special types of index sets which we do in the following.

An important quantity to get estimations on the lattice size is the difference set $\mathcal{D}(I(U))$ from (2.2) since the result [43, Theorem 8.16] tells us that there exists a reconstructing rank-1 lattice with prime cardinality

$$|I(U)| \leq M \leq |\mathcal{D}(I(U))|.$$

In the following, we proof properties and show estimates on the cardinality of both $I(U)$ and $\mathcal{D}(I(U))$.

LEMMA 5.2. *Let $U \subseteq \mathcal{P}(D)$ be a subset of ANOVA terms and $I_{\boldsymbol{u}} \subseteq \mathbb{Z}^{|\boldsymbol{u}|}$, $\boldsymbol{u} \in U$, finite symmetric frequency sets. Then we have*

$$\mathcal{D}(I(U)) = \bigcup_{\substack{\boldsymbol{u} \in U \\ \boldsymbol{v} \subseteq \boldsymbol{u}}} \{\boldsymbol{k} - \boldsymbol{h} \colon \boldsymbol{k} \in \mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}}, \boldsymbol{h} \in \mathrm{P}_{\boldsymbol{v}} I_{\boldsymbol{v}}\}.$$

Proof. It is easy to see that $\bigcup_{\substack{\boldsymbol{u} \in U \\ \boldsymbol{v} \subseteq \boldsymbol{u}}} \{\boldsymbol{k} - \boldsymbol{h} \colon \boldsymbol{k} \in \mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}}, \boldsymbol{h} \in \mathrm{P}_{\boldsymbol{v}} I_{\boldsymbol{v}}\} \subseteq \mathcal{D}(I(U))$ since $\mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}} \subseteq I(U)$ for every $\boldsymbol{u} \in U$ and $\boldsymbol{v} \in U$ for all $\boldsymbol{v} \subseteq \boldsymbol{u} \in U$ due to (4.1). In order to show the other inclusion we take an element $\boldsymbol{\ell} \in \mathcal{D}(I(U))$. By the uniqueness property of the ANOVA decomposition we know that there exists $\boldsymbol{u}, \boldsymbol{v} \in U$ such that $\boldsymbol{\ell} = \boldsymbol{k} - \boldsymbol{h}$ with $\boldsymbol{k} \in \mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}}$ and $\boldsymbol{h} \in \mathrm{P}_{\boldsymbol{v}} I_{\boldsymbol{v}}$. Taking the symmetry of the index sets $I_{\boldsymbol{u}}$ into account, we have proven the statement.   □

The following lemma gives an estimate for the size of the difference set of index sets of type $I(U)$ if there exists an upper bound on the cardinality of the term dependent sets $I_{\boldsymbol{u}}$.

LEMMA 5.3. *Let U be a subset of ANOVA terms and $I_{\boldsymbol{u}} \subseteq (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}, \boldsymbol{u} \in U$, symmetric frequency sets. Then the cardinality of the difference set of $I(U)$ is bounded by*

$$(5.13) \qquad |\mathcal{D}(I(U))| \leq \sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} |I_{\boldsymbol{u}}| \, |I_{\boldsymbol{v}}| \leq 2^{\max_{\boldsymbol{u} \in U} |\boldsymbol{u}|} |U| \max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}|^2 \, .$$

Proof. We estimate the cardinality of the difference set by applying Lemma 5.2

$$|\mathcal{D}(I(U))| \leq \sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} |I_{\boldsymbol{u}}| \, |I_{\boldsymbol{v}}| \, .$$

Here, we do not have equality since the union in Lemma 5.2 is not necessarily disjoint. Applying the upper bound on the cardinality of the sets $I_{\boldsymbol{u}}$, we arrive at

$$\sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} |I_{\boldsymbol{u}}| \, |I_{\boldsymbol{v}}| \leq \sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} \max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}|^2 \leq \max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}|^2 \, 2^{\max_{\boldsymbol{u} \in U} |\boldsymbol{u}|} \sum_{\boldsymbol{u} \in U} 1$$

$$\leq 2^{\max_{\boldsymbol{u} \in U} |\boldsymbol{u}|} \, |U| \max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}|^2 . \qquad \qquad \square$$

Remark 5.4. The cardinality of U_{d_s} is bounded by $(\mathrm{e} \cdot d/d_s)^{d_s}$, see Lemma 4.3. Therefore the estimate in (5.13) becomes

$$|\mathcal{D}(I(U_{d_s}))| \leq \left(\frac{2\mathrm{e} \cdot d}{d_s} \right)^{d_s} \max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}|^2 .$$

In the following, we consider special term-dependent frequency index sets of the structure

$$(5.14) \qquad I_{\boldsymbol{u}} := \left\{ \boldsymbol{\ell} \in (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|} : w(\boldsymbol{k}) \leq N_{\boldsymbol{u}} \text{ for } \boldsymbol{k} \in \mathbb{Z}^d \text{ with } \boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}, \, \boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0} \right\}$$

with a subset of coordinate indices $\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}$, a weight function $w : \mathbb{Z}^d \to [1, \infty)$ and cut-off $N_{\boldsymbol{u}} \in \mathbb{N}$. For a given subset of ANOVA terms $U \subseteq \mathcal{P}(\mathcal{D})$ we estimate the cardinalities of both, $I(U)$ and the difference set $\mathcal{D}(I(U))$.

LEMMA 5.5. *Let $U \subseteq \mathcal{P}(\mathcal{D})$ be a subset of ANOVA terms, $I_{\emptyset} = \{\boldsymbol{0}\}$, and $I_{\boldsymbol{u}}$, $\emptyset \neq \boldsymbol{u} \in U$, finite frequency sets as in (5.14) for a weight function $w : \mathbb{Z}^d \to [1, \infty)$ and $N_{\boldsymbol{u}} \in \mathbb{N}$. Moreover, let $h_{\min} : \mathbb{N} \to [1, \infty)$ and $h_{\max} : \mathbb{N} \to [1, \infty)$ be functions such that*

$$c \, h_{\min}(N_{\boldsymbol{u}_{\min}}) \leq \min_{\boldsymbol{u} \in U \setminus \{\emptyset\}} |I_{\boldsymbol{u}}| \quad and \quad \max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}| \leq C \, h_{\max}(N_{\boldsymbol{u}_{\max}})$$

with $\boldsymbol{u}_{\min} = \arg\min_{\boldsymbol{u} \in U \setminus \{\emptyset\}} |I_{\boldsymbol{u}}|$, $\boldsymbol{u}_{\max} = \arg\max_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}|$, and $0 < c \leq C$. Then we have for the asymptotic behavior of the cardinality of $I(U)$

$$c \, h_{\min}(N_{\boldsymbol{u}_{\min}}) \leq \frac{|I(U)|}{|U|} \leq C \, h_{\max}(N_{\boldsymbol{u}_{\max}}).$$

The constants do not depend on the spatial dimension d.

Proof. Since the projected sets $\mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}}$, $\boldsymbol{u} \in U$, are disjoint, we have

$$|I(U)| = \sum_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}| .$$

In order to show the upper bound, we estimate the cardinality of each index set by $h_{\max}$

$$\sum_{\boldsymbol{u} \in U} |I_{\boldsymbol{u}}| \leq \sum_{\boldsymbol{u} \in U} C \, h_{\max}(N_{\boldsymbol{u}_{\max}}) \leq C \, h_{\max}(N_{\boldsymbol{u}_{\max}}) \sum_{\boldsymbol{u} \in U} 1 = |U| \, C \, h_{\max}(N_{\boldsymbol{u}_{\max}}).$$

The lower bound follows with similar arguments. $\qquad \square$

COROLLARY 5.6. *Let $U \subseteq \mathcal{P}(\mathcal{D})$ be a subset of ANOVA terms, $I_{\emptyset} = \{\boldsymbol{0}\}$, and $I_{\boldsymbol{u}}$, $\emptyset \neq \boldsymbol{u} \in U$, finite symmetric frequency sets as in (5.14) for a weight function*

$w \colon \mathbb{Z}^d \to [1, \infty)$ *and* $N_{\boldsymbol{u}} \in \mathbb{N}$. *Moreover, let* $h_{\max}$ *be a function as in* *Lemma* 5.5. *Then*

$$|\mathcal{D}(I(U))| \leq C^2 \, 2^{\max_{\boldsymbol{u} \in U} |\boldsymbol{u}|} \, |U| \, h_{\max}^2(N_{\boldsymbol{u}_{\max}}).$$

Proof. The corollary is a direct consequence of Lemma 5.3 and Lemma 5.5. □

We may apply [43, Algorithm 8.17] to construct the reconstructing rank-1 lattice $\Lambda(\boldsymbol{z}, M, I(U))$ via a component-by-component approach. Choosing the set $X = \Lambda(\boldsymbol{z}, M, I(U))$ as sampling nodes yields a Moore-Penrose inverse of type (5.12) and we are able to compute the solution to (5.9) by multiplying with the adjoint Fourier matrix. This computation can be done efficiently using a lattice fast Fourier transform or LFFT, see [43, Section 8.2.2].

5.2.2. Scattered data scenario. In this section, we consider the scenario of scattered data approximation, i.e., we have a fixed set of nodes $X \subseteq \mathbb{T}^d$. Here, we aim to solve the least-squares problem (5.9) with the iterative LSQR method [42]. Specifically, we are interested in the matrix-free variant, i.e., we do not have to construct the system matrix $\boldsymbol{F}_{I(U)} \in \mathbb{C}^{|X|, |I(U)|}$ explicitly. The curse of dimensionality would quickly lead to the size of the matrix becoming intractable. The matrix-free variant requires two algorithms, one which takes a vector $\boldsymbol{a} \in \mathbb{C}^{|I(U)|}$ as an input and returns the result of the matrix-vector multiplication $\boldsymbol{F}_{I(U)} \boldsymbol{a}$ and one that takes $\hat{\boldsymbol{a}} \in \mathbb{C}^{|X|}$ as an input and returns the result of $\boldsymbol{F}_{I(U)}^* \hat{\boldsymbol{a}}$. If we take Remark 5.1 into account, it is only necessary to provide algorithms for fast multiplication with Fourier matrices $\boldsymbol{F}_{I_{\boldsymbol{u}}} \in \mathbb{C}^{|X|, |I_{\boldsymbol{u}}|}$, $\boldsymbol{u} \in U$.

The existence of such algorithms depends on the choice of the specific index sets $I_{\boldsymbol{u}}$. For full grids, i.e., frequency sets of the type

$$I_{\boldsymbol{u}} = G_N^{\boldsymbol{u}} = \left\{ \boldsymbol{k} \in \mathbb{Z}^{|\boldsymbol{u}|} : -\frac{N_{\boldsymbol{u}}}{2} \leq k_i \leq \frac{N_{\boldsymbol{u}}}{2} - 1, i = 1, 2, \ldots, |\boldsymbol{u}| \right\}, N_{\boldsymbol{u}} \in 2\mathbb{N},$$

the non-equispaced fast Fourier transform (NFFT) was introduced in [31]. Moreover, for hyperbolic cross index sets of the form

$$I_{\boldsymbol{u}} = H_n^{|\boldsymbol{u}|} = \bigcup_{\substack{\boldsymbol{j} \in \mathbb{N}_0^{|\boldsymbol{u}|} \\ \|\boldsymbol{j}\|_1 = n}} \hat{G}_{\boldsymbol{j}}$$

with $\hat{G}_{\boldsymbol{n}} = \times_{s=1}^{|\boldsymbol{u}|} \hat{G}_{n_s}$ and $\hat{G}_{n_s} = (-2^{n_s-1}, 2^{n_s-1}]^{|\boldsymbol{u}|} \cap \mathbb{Z}$, we have the non-equispaced hyperbolic cross fast Fourier transform (NHCFFT), cf. [10].

5.3. Approximation with active set. Now that we have obtained the active set $U_{X, \boldsymbol{y}}^{(\boldsymbol{\varepsilon})}$ from (5.8), we aim to construct an approximation using only these ANOVA terms. The global sensitivity indices $\varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f)$ calculated from the approximation $S_{I(U_{d_s})}^X f$ in (5.6) provide us with a basis to choose term-dependent frequency index sets $I_{\boldsymbol{u}} \subseteq (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}$, $\emptyset \neq \boldsymbol{u} \in U_{X, \boldsymbol{y}}^{(\boldsymbol{\varepsilon})}$. A higher sensitivity index suggests that the term is more important to the function and therefore a larger corresponding index set could be advisable.

We project the index sets as before to obtain $I(U_{X, \boldsymbol{y}}^{(\boldsymbol{\varepsilon})})$, see (5.10). Note that in general and depending on the threshold $\boldsymbol{\varepsilon}$, we have reduced the number of frequencies significantly. This is a sensible measure to reduce the effects of overfitting. Now, we

approximate f by the Fourier partial sum

$$f(\boldsymbol{x}) \approx \mathrm{T}_{U_{X,\boldsymbol{y}}^{(\varepsilon)}} f(\boldsymbol{x}) \approx S_{I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} \mathrm{c}_{\boldsymbol{k}}(f)\, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}.$$

The Fourier coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$ are again unknown and we determine them by least-squares approximation from X and $\boldsymbol{y}$. The unique solution is given by

$$(5.15) \qquad \hat{\boldsymbol{f}}_{\mathrm{sol}} = \underset{\hat{\boldsymbol{f}} \in \mathbb{C}^{\left| I(U_{X,\boldsymbol{y}}^{(\varepsilon)}) \right|}}{\arg\min} \left\| \boldsymbol{y} - \boldsymbol{F}_{I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} \hat{\boldsymbol{f}} \right\|_2^2$$

if the Fourier matrix $\boldsymbol{F}_{I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} = \left(\mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \right)_{\boldsymbol{x} \in X, \boldsymbol{k} \in I(U_{X,\boldsymbol{y}}^{(\varepsilon)})}$ has full rank. Details on how to solve this system for scattered data and black-box approximation can be found in Subsection 5.2. We use the elements of the solutions vector $\hat{\boldsymbol{f}}_{\mathrm{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k} \in I(U_{X,\boldsymbol{y}}^{(\varepsilon)})}$ to form the approximate Fourier partial sum and our solution

$$f(\boldsymbol{x}) \approx S_{I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} f(\boldsymbol{x}) \approx S_{I(U_{X,\boldsymbol{y}}^{(\varepsilon)})}^X f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} \hat{f}_{\boldsymbol{k}}\, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}.$$

The following algorithm summarizes the proposed method.

Algorithm 5.1 ANOVA Approximation Method

Input: $X \subseteq \mathbb{T}^d$ finite node set
          $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x} \in X}$ function values
          $d_s \in \mathcal{D}$ superposition threshold

1: Choose finite order-dependent search sets $I_1 \subseteq \mathbb{Z} \setminus \{0\}, \ldots, I_{d_s} \subseteq (\mathbb{Z} \setminus \{0\})^{d_s}$.
2: Compute solution of least-squares problem (5.5).
3: $\hat{\boldsymbol{f}}_{\mathrm{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k} \in I(U_{d_s})} \leftarrow \arg\min_{\hat{\boldsymbol{f}} \in \mathbb{C}^{|I(U_{d_s})|}} \left\| \boldsymbol{y} - \boldsymbol{F}_{I(U_{d_s})} \hat{\boldsymbol{f}} \right\|_2^2$
4: Compute global sensitivity indices for approximation $S_{I(U_{d_s})}^X f$ using (3.6).
5: $\varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f) \leftarrow \dfrac{\left\| (S_{I(U_{d_s})}^X f)_{\boldsymbol{u}} \right\|_{\mathrm{L}_2(\mathbb{T}^d)}^2}{\left\| S_{I(U_{d_s})}^X f \right\|_{\mathrm{L}_2(\mathbb{T}^d)}^2 - \left| \mathrm{c}_0 \left(S_{I(U_{d_s})}^X f \right) \right|^2}, \ \boldsymbol{u} \in U_{d_s}$
6: Choose threshold vector $\boldsymbol{\varepsilon} \in [0,1]^{d_s}$ and build active set.
7: $U_{X,\boldsymbol{y}}^{(\boldsymbol{\varepsilon})} \leftarrow \left\{ \boldsymbol{v} \subseteq \mathcal{D} \colon \exists \boldsymbol{u} \in U_{d_s} \colon \boldsymbol{v} \subseteq \boldsymbol{u} \quad \text{and} \quad \varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f) > \varepsilon_{|\boldsymbol{u}|} \right\}$
8: Use information from global sensitivity indices to choose finite index sets $I_{\boldsymbol{u}} \subseteq (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}$ per ANOVA term in $U_{X,\boldsymbol{y}}^{(\varepsilon)}$.
9: Compute solution of least-squares problem (5.15).
10: $\hat{\boldsymbol{f}}_{\mathrm{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k} \in I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} \leftarrow \arg\min_{\hat{\boldsymbol{f}} \in \mathbb{C}^{|I(U_{X,\boldsymbol{y}}^{(\varepsilon)})|}} \left\| \boldsymbol{y} - \boldsymbol{F}_{I(U_{X,\boldsymbol{y}}^{(\varepsilon)})} \hat{\boldsymbol{f}} \right\|_2^2$

Output: $\hat{f}_{\boldsymbol{k}} \in \mathbb{C}, \boldsymbol{k} \in I(U_{X,\boldsymbol{y}}^{(\varepsilon)})$ approximations to Fourier
                                         coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$
            $\varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f) \in [0,1], \boldsymbol{u} \in U_{d_s}$ global sensitivity indices of $S_{I(U_{d_s})}^X f$
                                         or *importance ranking* on the terms

---

6. Error analysis. The error of our approximation method measured in the norm of some space $H \subseteq L_2(\mathbb{T}^d)$ can be decomposed into multiple components by the triangle inequality

$$\left\| f - S_{I(U)}^X f \right\|_H \leq \underbrace{\left\| f - T_U f \right\|_H}_{\text{ANOVA truncation error}} + \underbrace{\left\| T_U f - S_{I(U)}^X f \right\|_H}_{\text{approximation error}}$$

for an active set of ANOVA terms $U \subseteq U_{d_s}$ with superposition threshold $d_s \in \mathcal{D}$. We distinguish between the ANOVA truncation error and the approximation error. Here, the analysis of the ANOVA truncation error is independent of the concrete approximation problem (5.15) and the scenario (scattered data or black-box).

6.1. ANOVA truncation error. The ANOVA truncation error is related to the truncation of the ANOVA decomposition to the set U_{d_s} with superposition threshold $d_s \in \mathcal{D}$ and the active set $U \subseteq U_{d_s}$. We can separate the ANOVA truncation error as follows

$$(6.1) \qquad \| f - T_U f \|_H \leq \underbrace{\| f - T_{d_s} f \|_H}_{\text{truncation by } d_s} + \underbrace{\| T_{d_s} f - T_U f \|_H}_{\text{active set truncation}} .$$

Here, we bring T_{d_s} in with the aim to relate the error to our function class of low-order interactions, see (4.2). To control the second term, we require assumptions on the sensitivity indices of the ANOVA terms in $U_{d_s} \setminus U$. Since the error is only related to the structure of the function it can be considered independently of any specific approximation scenario like black-box or scattered data approximation. We show bounds for this error in the case that f is an element of a Sobolev type space $H^w(\mathbb{T}^d)$ or a Wiener algebra $\mathcal{A}^w(\mathbb{T}^d)$ and H is $L_2(\mathbb{T}^d)$ or $L_\infty(\mathbb{T}^d)$.

THEOREM 6.1. *Let $f \in H^w(\mathbb{T}^d)$ with a weight function $w \colon \mathbb{Z}^d \to [1, \infty)$ and superposition dimension $d^{(\mathrm{sp})}$, see (3.7), for a $\delta \in (0,1)$. If there exists a subset of ANOVA terms $U \subseteq U_{d^{(\mathrm{sp})}}$ such that*

$$\varrho(\boldsymbol{u}, f) = \frac{\sigma^2(f_{\boldsymbol{u}})}{\sigma^2(f)} < \varepsilon, \, \varepsilon > 0,$$

for every $\boldsymbol{u} \in U_{d^{(\mathrm{sp})}} \setminus U$ then

$$\frac{\| f - T_U f \|_{L_2(\mathbb{T}^d)}}{\| f \|_{H^w(\mathbb{T}^d)}} \leq \sqrt{1 - \delta} + \sqrt{|U_{d^{(\mathrm{sp})}} \setminus U| \varepsilon}.$$

Proof. The ANOVA truncation error can be separated as in (6.1). We prove an upper bound for the active set truncation. With Parseval's equality and the assumption on the global sensitivity indices, we estimate

$$(6.2) \qquad \| T_{d^{(\mathrm{sp})}} f - T_U f \|_{L_2(\mathbb{T}^d)}^2 = \sum_{\boldsymbol{u} \in U_{d^{(\mathrm{sp})}} \setminus U} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} |c_{\boldsymbol{k}}(f)|^2 \leq \sigma^2(f) |U_{d^{(\mathrm{sp})}} \setminus U| \varepsilon.$$

Clearly, we have $\sigma^2(f) \leq \| f \|_{L_2(\mathbb{T}^d)}^2 \leq \| f \|_{H^w(\mathbb{T}^d)}^2.$ □

THEOREM 6.2. *Let $f \in \mathcal{A}^w(\mathbb{T}^d)$ with a weight function $w \colon \mathbb{Z}^d \to [1, \infty)$. If there exsists a subset of ANOVA terms $U \subseteq U_{d_s}$, $d_s \in \mathcal{D}$, such that*

$$(6.3) \qquad \frac{\sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} |c_{\boldsymbol{k}}(f)|}{\sum_{\boldsymbol{k} \in \mathbb{Z}^d} |c_{\boldsymbol{k}}(f)|} < \varepsilon_1, \, \varepsilon_1 > 0$$

for every $\boldsymbol{u} \in U_{d_s} \setminus U$ and we have

$$\frac{\|f - \mathrm{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathcal{A}^w(\mathbb{T}^d)}} < \varepsilon_2, \; \varepsilon_2 > 0,$$

then

$$\frac{\|f - \mathrm{T}_U f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathcal{A}^w(\mathbb{T}^d)}} \le \varepsilon_2 + \sqrt{|U_{d_s} \setminus U| \, \varepsilon_1}.$$

Proof. We split the ANOVA truncation error as in (6.1) and prove an upper bound for the second part. To this end, we estimate the L_∞ norm of f by the absolute values of its Fourier coefficients and apply (6.3) to obtain

$$\|\mathrm{T}_{d_s}f - \mathrm{T}_U f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \le \sum_{\boldsymbol{u} \in U_{d_s} \setminus U} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} |c_{\boldsymbol{k}}(f)| \le |U_{d_s} \setminus U| \, \varepsilon_1 \sum_{\boldsymbol{k} \in \mathbb{Z}^d} |c_{\boldsymbol{k}}(f)|\,.$$

Naturally, it holds that $\sum_{\boldsymbol{k} \in \mathbb{Z}^d} |c_{\boldsymbol{k}}(f)| \le \|f\|_{\mathcal{A}^w(\mathbb{T}^d)}$ which leads to the desired estimate. □

Note that in order to prove a bound for the error in L_∞, we formulated a condition on an ℓ_1 equivalent of the global sensitivity indices $\varrho(\boldsymbol{u}, f)$ in accordance with the Wiener algebra norm.

6.2. Approximation error. In this section, we focus on the approximation error which we separate into two parts as well

$$(6.4) \qquad \left\|\mathrm{T}_U f - S_{I(U)}^X f\right\|_H \le \underbrace{\left\|\mathrm{T}_U f - S_{I(U)} f\right\|_H}_{\text{truncation error}} + \underbrace{\left\|S_{I(U)} f - S_{I(U)}^X f\right\|_H}_{\text{aliasing error}}$$

with $H \in \{\mathrm{L}_2(\mathbb{T}^d), \mathrm{L}_\infty(\mathbb{T}^d)\}$, a subset of ANOVA terms $U \subseteq \mathcal{P}(\mathcal{D})$, and a finite frequency index set $I(U) \subseteq \mathbb{Z}^d$ of structure (5.10) with sets $I_{\boldsymbol{u}}$ as in (5.14). The truncation error remains independent of the approximation scenario and can be estimated by the norms in $\mathcal{A}^w$ and H^w.

LEMMA 6.3. *Let $f \in \mathrm{H}^w(\mathbb{T}^d)$, $w: \mathbb{Z}^d \to [1, \infty)$ a weight function, and $I(U) \subseteq \mathbb{Z}^d$ a finite frequency index set of type (5.10) with $U \subseteq \mathcal{P}(\mathcal{D})$. Then the relative truncation error can be estimated as*

$$(6.5) \qquad \frac{\left\|\mathrm{T}_U f - S_{I(U)} f\right\|_{\mathrm{L}_2(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}} \le \frac{1}{\min_{\boldsymbol{u} \in U} N_{\boldsymbol{u}}}.$$

If in addition we have $\sum_{\boldsymbol{k} \in \mathbb{Z}^d} \frac{1}{w^2(\boldsymbol{k})} < \infty$, we can estimate

$$(6.6) \qquad \frac{\left\|\mathrm{T}_U f - S_{I(U)} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}} \le \sqrt{\sum_{\boldsymbol{u} \in U} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)} \setminus \mathrm{P}_{\boldsymbol{u}} I_{\boldsymbol{u}}} \frac{1}{w^2(\boldsymbol{k})}}.$$

Proof. In order to prove (6.5) we employ Parseval's identity and use the weight

$w(\boldsymbol{k})$

$$\left\|\mathrm{T}_U f - S_{I(U)}f\right\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \sum_{\boldsymbol{u}\in U}\sum_{\boldsymbol{k}\in\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} |c_{\boldsymbol{k}}(f)|^2 = \sum_{\boldsymbol{u}\in U}\sum_{\boldsymbol{k}\in\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} \frac{w^2(\boldsymbol{k})}{w^2(\boldsymbol{k})} |c_{\boldsymbol{k}}(f)|^2$$

$$\leq \sum_{\boldsymbol{u}\in U} \frac{1}{N^2_{\boldsymbol{u}}} \sum_{\boldsymbol{k}\in\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} w^2(\boldsymbol{k}) |c_{\boldsymbol{k}}(f)|^2 \leq \frac{1}{\min_{\boldsymbol{u}\in U}N^2_{\boldsymbol{u}}} \|f\|^2_{\mathrm{H}^w(\mathbb{T}^d)}.$$

For the bound (6.6) we estimate the norm by the absolute sum of the Fourier coefficients and use the Cauchy-Schwarz inequality

$$\left\|\mathrm{T}_U f - S_{I(U)}f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)} = \sum_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\in U}\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash I(U)} |c_{\boldsymbol{k}}(f)| = \sum_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\in U}\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash I(U)} \frac{w(\boldsymbol{k})}{w(\boldsymbol{k})} |c_{\boldsymbol{k}}(f)|$$

$$\leq \|f\|_{\mathrm{H}^w(\mathbb{T}^d)} \sqrt{\sum_{\boldsymbol{u}\in U}\sum_{\boldsymbol{k}\in\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} \frac{1}{w^2(\boldsymbol{k})}}. \qquad \square$$

LEMMA 6.4. *Let $f \in \mathcal{A}^w(\mathbb{T}^d)$ with $w\colon \mathbb{Z}^d \to [1,\infty)$ a weight function such that $\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{1}{w^2(\boldsymbol{k})} < \infty$, and $I(U) \subseteq \mathbb{Z}^d$ a finite frequency index set of type (5.10) with $U \subseteq \mathcal{P}(\mathcal{D})$ and sets $I_{\boldsymbol{u}}$ as in (5.14). Then the relative truncation error can be estimated as*

$$\frac{\left\|\mathrm{T}_U f - S_{I(U)}f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathcal{A}^w(\mathbb{T}^d)}} \leq \min\left\{ \frac{1}{\min_{\boldsymbol{u}\in U}N_{\boldsymbol{u}}}, \max_{\boldsymbol{u}\in U}\sqrt{\sum_{\boldsymbol{k}\in\mathbb{F}^{(d)}_{\boldsymbol{u}}\backslash\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} \frac{1}{w^2(\boldsymbol{k})}}\right\}.$$

Proof. The proof requires similar steps to the proof of Lemma 6.3. $\qquad \square$

For the aliasing error in (6.4), we start by considering the black-box approximation case where we solve the least-squares problem as described in Subsection 5.2.

THEOREM 6.5. *Let $f \in \mathrm{H}^w(\mathbb{T}^d)$ with a weight function $w\colon \mathbb{Z}^d \to [1,\infty)$ such that $\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\frac{1}{w^2(\boldsymbol{k})} < \infty$ and $I(U) \subseteq \mathbb{Z}^d$ a finite frequency index set of type (5.10) with sets $I_{\boldsymbol{u}}$ as in (5.14). Moreover, we have a reconstructing rank-1 lattice $\Lambda(\boldsymbol{z}, M, I(U))$ for a generating vector $\boldsymbol{z} \in \mathbb{Z}^d$ and lattice size $M \in \mathbb{N}$. Then the aliasing error can be estimated as*

$$(6.7) \qquad \frac{\left\|S_{I(U)}f - S^{\Lambda(\boldsymbol{z},M,I(U))}_{I(U)}f\right\|_{\mathrm{L}_2(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}} \leq \sqrt{\sum_{\boldsymbol{k}\in\mathbb{Z}^d\backslash I(U)} \frac{1}{w^2(\boldsymbol{k})}}.$$

Furthermore, if $f \in \mathcal{A}^w(\mathbb{T}^d)$ we get for the L_∞-norm

$$(6.8) \qquad \frac{\left\|S_{I(U)}f - S^{\Lambda(\boldsymbol{z},M,I(U))}_{I(U)}f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)}}{\|f\|_{\mathcal{A}^w(\mathbb{T}^d)}} \leq \frac{1}{\min_{\boldsymbol{u}\in U}N_{\boldsymbol{u}}}.$$

Proof. We show the bound (6.7) by first applying Parseval's identity and (2.3)

$$\left\| S_{I(U)}f - S_{I(U)}^{\Lambda(\boldsymbol{z},M,I(U))} f \right\|_{L_2(\mathbb{T}^d)}^2 = \sum_{\boldsymbol{k}\in I(U)} \left| \hat{f}_{\boldsymbol{k}} - c_{\boldsymbol{k}}(f) \right|^2$$

$$= \sum_{\boldsymbol{k}\in I(U)} \left| \sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} c_{\boldsymbol{k}+\boldsymbol{h}}(f) \right|^2.$$

We then incorporate the weight and utilize the Cauchy-Schwarz inequality to obtain

$$\left\| S_{I(U)}f - S_{I(U)}^{\Lambda(\boldsymbol{z},M,I(U))} f \right\|_{L_2(\mathbb{T}^d)}^2 = \sum_{\boldsymbol{k}\in I(U)} \left| \sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} \frac{w(\boldsymbol{k}+\boldsymbol{h})}{w(\boldsymbol{k}+\boldsymbol{h})} c_{\boldsymbol{k}+\boldsymbol{h}}(f) \right|^2$$

$$\leq \sum_{\boldsymbol{k}\in I(U)} \left(\sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} w^2(\boldsymbol{k}+\boldsymbol{h}) \left| c_{\boldsymbol{k}+\boldsymbol{h}}(f) \right|^2 \right) \times$$

$$\times \left(\sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} \frac{1}{w^2(\boldsymbol{k}+\boldsymbol{h})} \right)$$

From [43, Lemma 8.13] we know that for fixed $\boldsymbol{k}\in I(U)$ we have disjoint sets

$$M_{\boldsymbol{k}} := \left\{ \boldsymbol{k}+\boldsymbol{h} \colon \boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\} \right\} \subseteq \mathbb{Z}^d \setminus I(U).$$

This means we are able to estimate

$$\sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} w^2(\boldsymbol{k}+\boldsymbol{h}) \left| c_{\boldsymbol{k}+\boldsymbol{h}}(f) \right|^2 = \sum_{\boldsymbol{\ell}\in M_{\boldsymbol{k}}} w^2(\boldsymbol{\ell}) \left| c_{\boldsymbol{\ell}}(f) \right|^2$$

$$\leq \sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} w^2(\boldsymbol{\ell}) \left| c_{\boldsymbol{\ell}}(f) \right|^2 = \|f\|_{H^w(\mathbb{T}^d)}^2$$

such that

$$\left\| S_{I(U)}f - S_{I(U)}^{\Lambda(\boldsymbol{z},M,I(U))} f \right\|_{L_2(\mathbb{T}^d)}^2 \leq \|f\|_{H^w(\mathbb{T}^d)}^2 \sum_{\boldsymbol{k}\in I(U)} \sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} \frac{1}{w^2(\boldsymbol{k}+\boldsymbol{h})}.$$

Using that the sets $M_{\boldsymbol{k}}$ are disjoint and $\bigcup_{\boldsymbol{k}\in I(U)} M_{\boldsymbol{k}} \subseteq \mathbb{Z}^d \setminus I(U)$ yields

$$\sum_{\boldsymbol{k}\in I(U)} \sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} \frac{1}{w^2(\boldsymbol{k}+\boldsymbol{h})} = \sum_{\boldsymbol{k}\in I(U)} \sum_{\boldsymbol{\ell}\in M_{\boldsymbol{k}}} \frac{1}{w^2(\boldsymbol{\ell})}$$

$$= \sum_{\boldsymbol{\ell}\in\bigcup_{\boldsymbol{k}\in I(U)} M_{\boldsymbol{k}}} \frac{1}{w^2(\boldsymbol{\ell})} \leq \sum_{\boldsymbol{\ell}\in\mathbb{Z}^d\setminus I(U)} \frac{1}{w^2(\boldsymbol{\ell})}.$$

The L_∞-bound (6.8) is obtained similarly to the method used in the proof of [43,

Theorem 8.14]. We proceed as follows

$$\left\|S_{I(U)}f - S_{I(U)}^{\Lambda(\boldsymbol{z},M,I(U))}f\right\|_{L_\infty(\mathbb{T}^d)} \leq \sum_{\boldsymbol{u}\in U} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}\setminus\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} |c_{\boldsymbol{k}}(f)|$$

$$= \sum_{\boldsymbol{u}\in U} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}\setminus\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} \frac{w(\boldsymbol{k})}{w(\boldsymbol{k})} |c_{\boldsymbol{k}}(f)|$$

$$\leq \frac{1}{\min_{\boldsymbol{u}\in U} N_{\boldsymbol{u}}} \sum_{\boldsymbol{u}\in U} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}\setminus\mathrm{P}_{\boldsymbol{u}}I_{\boldsymbol{u}}} w(\boldsymbol{k}) |c_{\boldsymbol{k}}(f)|.$$

The result is obtained through estimating the sum by $\|f\|_{\mathcal{A}^w(\mathbb{T}^d)}$. $\qquad\square$

In the following we consider the approximation error for scattered data approximation with a fixed node set $X \subseteq \mathbb{T}^d$. Previously, we assumed that the index set $I(U)$ and the node set X are such that the Fourier matrix $\boldsymbol{F}_{I(U)}$ has full rank. In this case the least-squares problem (5.9) has a unique solution. Assuming that the nodes in X are i.i.d. random variables that are uniformly distributed in $\mathbb{T}^d$, it is possible to achieve good bounds on the approximation error, see [2, 20, 29, 39].

LEMMA 6.6. *Let* $f \in \mathrm{H}^w(\mathbb{T}^d)$ *with a weight function* $w\colon \mathbb{Z}^d \to [1,\infty)$ *such that* $\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{1}{w^2(\boldsymbol{k})} < \infty$, $X \subseteq \mathbb{T}^d$ *a finite set of i.i.d. uniformly distributed points,* $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in X}$, *and* $I(U) \subseteq \mathbb{Z}^d$ *a finite frequency index set of type* (5.10) *with* $U \subseteq \mathcal{P}(\mathcal{D})$ *a subset of ANOVA terms and sets* $I_{\boldsymbol{u}}$ *as in* (5.14). *If for the number of frequencies we have* $|I(U)| \leq \frac{|X|}{7r\log|X|}, r > 0$, *then*

$$\sup_{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}\leq 1} \frac{\sum_{\boldsymbol{x}\in X}\left|\left(f - S_{I(U)}f\right)(\boldsymbol{x})\right|^2}{|X|} \leq 5\max\left(\theta_{I(U)}^2, \frac{8r\kappa^2\log|X|}{|X|}\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus I(U)}\frac{1}{w^2(\boldsymbol{k})}\right)$$

with a probability of at least $1 - 3|X|^{1-r}$ *for* $\theta_{I(U)} = \left\|f - S_{I(U)}f\right\|_{L_2(\mathbb{T}^d)}$ *and* $\kappa = \frac{1+\sqrt{5}}{2}$.

Proof. The setting of this lemma is a special case of [39, Theorem 5.1]. $\qquad\square$

The following theorem deals with the actual approximation error by incorporating the previous lemma.

THEOREM 6.7. *Let* $f \in \mathrm{H}^w(\mathbb{T}^d)$ *with a weight function* $w\colon \mathbb{Z}^d \to [1,\infty)$ *such that* $\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{1}{w^2(\boldsymbol{k})} < \infty$, $I(U) \subseteq \mathbb{Z}^d$ *a finite frequency index set of type* (5.10) *with sets* $I_{\boldsymbol{u}}$ *as in* (5.14). *Moreover,* $U \subseteq \mathcal{P}(\mathcal{D})$, *and* $S_{I(U)}^X f$ *are the corresponding approximate Fourier partial sum obtained through the scattered data approximation method described in Subsection* 5.2. *If the elements of* $X \subseteq \mathbb{T}^d$ *are i.i.d. random variables uniformly distributed on* $\mathbb{T}^d$ *and for the number of frequencies we have* $|I(U)| \leq \frac{|X|}{7r\log|X|}, r > 0$, *then*

$$\frac{\left\|S_{I(U)}f - S_{I(U)}^X f\right\|_{L_2(\mathbb{T}^d)}}{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}} \leq \sqrt{8\max\left(\theta_{I(U)}^2, \kappa^2\frac{\log|X|}{|X|}\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus I(U)}\frac{1}{w^2(\boldsymbol{k})}\right)}$$

with a probability of at least $1 - 3|X|^{1-r}$ *for* $\theta_{I(U)} = \left\|f - S_{I(U)}f\right\|_{L_2(\mathbb{T}^d)}$ *and* $\kappa = \frac{1+\sqrt{5}}{2}$.

Proof. We denote the Fourier coefficients with $\hat{\boldsymbol{c}} = (c_{\boldsymbol{k}}(f))_{\boldsymbol{k}\in I(U)}$ and the approximate Fourier coefficients computed by Algorithm 5.1 with $\hat{\boldsymbol{f}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k}\in I(U)}$. With

Parseval's identity as well as the Moore-Penrose inverse we obtain

$$\left\|S_{I(U)}f - S^X_{I(U)}f\right\|_{\mathrm{L}_2(\mathbb{T}^d)} = \sqrt{\sum_{\boldsymbol{k}\in I(U)}\left|\hat{f}_{\boldsymbol{k}} - c_{\boldsymbol{k}}(f)\right|^2} = \left\|\hat{\boldsymbol{f}} - \hat{\boldsymbol{c}}\right\|_2$$

$$= \left\|\left(\boldsymbol{F}^*_{I(U)}\boldsymbol{F}_{I(U)}\right)^{-1}\boldsymbol{F}^*_{I(U)}\boldsymbol{y} - \hat{\boldsymbol{c}}\right\|_2$$

$$= \left\|\left(\boldsymbol{F}^*_{I(U)}\boldsymbol{F}_{I(U)}\right)^{-1}\boldsymbol{F}^*_{I(U)}\left(\boldsymbol{y} - \boldsymbol{F}_{I(U)}\hat{\boldsymbol{c}}\right)\right\|_2.$$

We use the properties of the spectral norm and estimate further

$$\leq \left\|\left(\boldsymbol{F}^*_{I(U)}\boldsymbol{F}_{I(U)}\right)^{-1}\boldsymbol{F}^*_{I(U)}\right\|_2 \left\|\boldsymbol{y} - \boldsymbol{F}_{I(U)}\hat{\boldsymbol{c}}\right\|_2$$

$$= \left\|\left(\boldsymbol{F}^*_{I(U)}\boldsymbol{F}_{I(U)}\right)^{-1}\boldsymbol{F}^*_{I(U)}\right\|_2 \sqrt{\sum_{\boldsymbol{x}\in X}\left|\left(f - S_{I(U)}f\right)(\boldsymbol{x})\right|^2}.$$

Applying [39, Theorem 2.3] yields

$$\sup_{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}\leq 1}\left\|S_{I(U)}f - S^X_{I(U)}f\right\|_{\mathrm{L}_2(\mathbb{T}^d)} \leq \sup_{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}\leq 1}\sqrt{\frac{2}{|X|}\sum_{\boldsymbol{x}\in X}\left|\left(f - S_{I(U)}f\right)(\boldsymbol{x})\right|^2}.$$

Finally, we use Lemma 6.6 to obtain our bound

$$\sup_{\|f\|_{\mathrm{H}^w(\mathbb{T}^d)}\leq 1}\left\|S_{I(U)}f - S^X_{I(U)}f\right\|_{\mathrm{L}_2(\mathbb{T}^d)} \leq \sqrt{8\max\left(\theta^2_{I(U)},\kappa^2\frac{\log|X|}{|X|}\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus I(U)}\frac{1}{w^2(\boldsymbol{k})}\right)}$$

with a probability of at least $1 - 3|X|^{1-r}$.   □

This concludes the consideration of the error of the presented method in both approximation scenarios. We were able to achieve bounds for L_2 and L_∞ for functions in weighted Wiener algebras and Sobolev type spaces.

7. Numerical Results. We present numerical results for the method described in Section 5 for a test function $f\colon [0,1)^9 \to \mathbb{R}$,

(7.1) $f(\boldsymbol{x}) := B_2(x_1)B_4(x_5) + B_2(x_2)B_4(x_6) + B_2(x_3)B_4(x_7) + B_2(x_4)B_4(x_8)B_6(x_9),$

where B_2, B_4 and B_6 are parts of univariate, shifted, scaled and dilated B-splines of order 2, 4, and 6, respectively, see Figure 3 for illustration. Their Fourier series is given by

$$B_j(x) := c_j \sum_{k\in\mathbb{Z}}\mathrm{sinc}^j\left(\frac{\pi\cdot k}{j}\right)\cos(\pi\cdot k)\,\mathrm{e}^{2\pi\mathrm{i}k\cdot x}$$

with $\mathrm{sinc}(x) := \sin(x)/x$ and the constants $c_2 := \sqrt{3/4}$, $c_4 := \sqrt{315/604}$, $c_6 := \sqrt{277200/655177}$ such that $\|B_j\|_{\mathrm{L}_2(\mathbb{T}^d)} = 1$. This allows the direct computation of the Fourier coefficients $c_{\boldsymbol{k}}(f)$ and the norm $\|f\|_{\mathrm{L}_2(\mathbb{T}^d)}$. The ANOVA terms $f_{\boldsymbol{u}}$ are only nonzero for

$$\boldsymbol{u}\in U^* := \mathcal{P}(\{1,5\})\cup\mathcal{P}(\{2,6\})\cup\mathcal{P}(\{3,7\})\cup\mathcal{P}(\{4,8,9\}).$$

(a) B-spline B_2
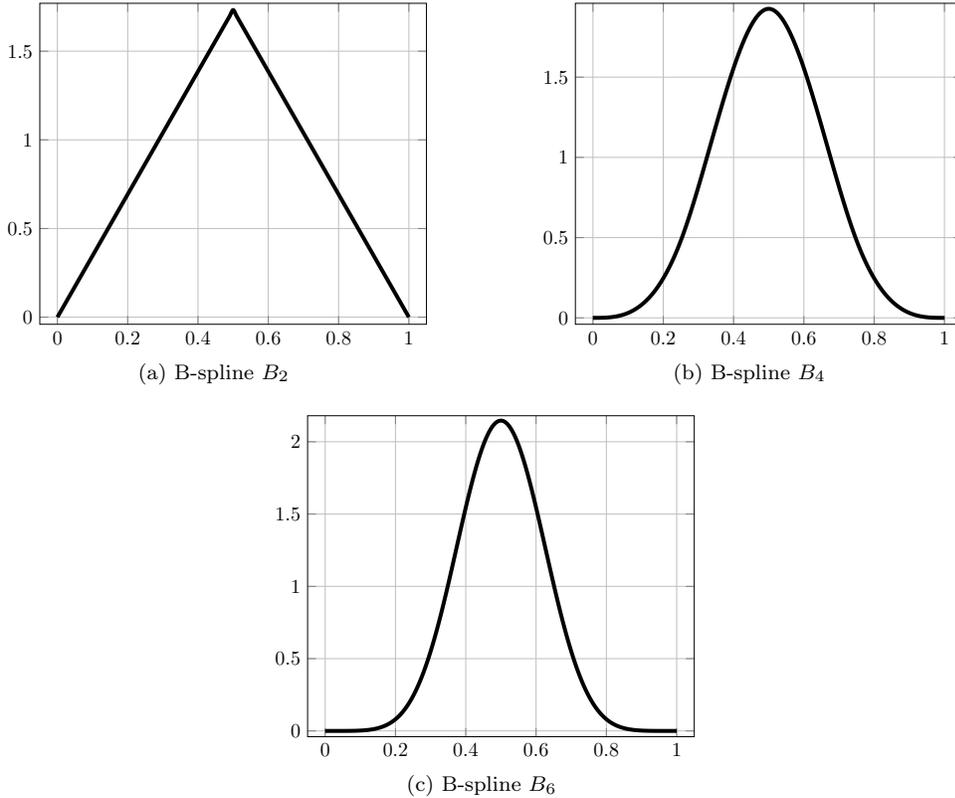

(b) B-spline B_4

(c) B-spline B_6

Fig. 3: B-splines B_2, B_4, and B_6 over $\mathbb{T} \cong [0, 1)$.

The function f therefore has an exact low-dimensional structure for $d_s = 3$, i.e., $\mathrm{T}_3 f = f$. This leads to $d_s = 3$ being the optimal choice for the superposition threshold with no error caused by ANOVA truncation since it corresponds to the superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ for $\delta = 1$, see (4.3). In an approximation scenario with an unknown function f this information is of course not known.

We consider two errors

$$(7.2) \qquad \varepsilon_{\ell_2} = \frac{\left\| \boldsymbol{y} - (S^X_{I(U_{d_s})} f(\boldsymbol{x}))_{\boldsymbol{x} \in X} \right\|_2}{\|\boldsymbol{y}\|_2}, \text{ and } \quad \varepsilon_{\mathrm{L}_2} = \frac{\left\| f - S^X_{I(U_{d_s})} f \right\|_{\mathrm{L}_2(\mathbb{T}^9)}}{\|f\|_{\mathrm{L}_2(\mathbb{T}^9)}}.$$

Here, the error ε_{ℓ_2} can be regarded as a training error since it is taken at the given sampling set X and the error $\varepsilon_{\mathrm{L}_2}$ as a type of generalization error since it measures the error in the Fourier coefficients. Since our goal is to find the important ANOVA terms, i.e., the terms in U^*, we expect to have an interval (or gap) in which to choose the order-dependent threshold $\boldsymbol{\varepsilon} \in [0, 1]^{d_s}$. Therefore, we define

$$I^{(j)} = \begin{cases} \emptyset & : \text{assumption (5.7) is not fulfilled} \\ (a^{(j)}, b^{(j)}) & : \text{assumption (5.7) is fulfilled} \end{cases}$$

with $1 \leq j \leq d_s$ and

$$a^{(j)} := \max \left\{ \varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f) \colon \boldsymbol{u} \in U_{d_s} \setminus U^*, |\boldsymbol{u}| = j \right\},$$

$$b^{(j)} := \min \left\{ \varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f) \colon \boldsymbol{u} \in U^*, |\boldsymbol{u}| = j \right\}.$$

Here, the assumption (5.7) is to be understood for every order of terms, i.e., for $\boldsymbol{u}$ and $\boldsymbol{v}$ with $|\boldsymbol{u}| = |\boldsymbol{v}| = j$.

Remark 7.1. The norm occurring in the error $\varepsilon_{\mathrm{L}_2}$ can be calculated using Parseval's identity

$$\left\| f - S_{I(U_{d_s})}^X f \right\|_{\mathrm{L}_2}^2 = \|f\|_{\mathrm{L}_2}^2 + \sum_{\boldsymbol{k} \in I(U_{d_s})} \left| c_{\boldsymbol{k}}(f) - \hat{f}_{\boldsymbol{k}} \right|^2 - \sum_{\boldsymbol{k} \in I(U_{d_s})} |c_{\boldsymbol{k}}(f)|^2$$

which is possible since we know the exact Fourier coefficients and the norm of the function f. In general, this error cannot be computed.

7.1. Scattered Data Approximation. For our numerical experiments we use one sampling set $X \subseteq \mathbb{T}^9$ of uniformly distributed nodes with $M := |X| = 2.5 \cdot 10^6$, and an evaluation vector $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x} \in X}$. We are going to start by choosing three as the superposition threshold d_s while later reducing it to two which allows us to see the effect of truncating an ANOVA term. Our primary aim for now is to detect the ANOVA terms in U^* which we achieve using the first step of our method, see Subsection 5.1. To this end, we choose a frequency index set $I(U_{d_s}) \subseteq \mathbb{Z}^9$, cf. (5.4), through order-dependent sets $I_0 = \{0\}$, $I_1 = \{-N_1/2, \ldots, N_1/2 - 1\}$, $I_2 = \{-N_2/2, \ldots, N_2/2 - 1\}^2$, and $I_3 = \{-N_3/2, \ldots, N_3/2 - 1\}^3$ with $N_1, N_2, N_3 \in 2\mathbb{N}$. The method gives us an approximation $S_{I(U_{d_s})}^X f$.

Results of numerical experiments with the function f from (7.1) and different choices for the bandwidths N_1, N_2, and N_3 are displayed in Table 1. They show that it is indeed possible to detect the ANOVA terms in U^* using trigonometric polynomials of small degrees. Moreover, both errors are roughly of the same order. Since our number of samples M is fixed, we are looking for values $\boldsymbol{N}$ such that one balances the effects of underfitting and overfitting. The experiments suggest that the choice in examples 5 and 8 is close to optimal. In Figure 4 we depicted the global sensitivity indices $\varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f)$, cf. Algorithm 5.1, for example 8 from Table 1. The one-dimensional sets $\{i\}$, $i = 1, \ldots, 9$, all have large indices as they are all in U^* while the two dimensional sets

$$\{1,5\}, \{2,6\}, \{3,7\}, \{4,8\}, \{4,9\}, \{8,9\} \in U^*$$

are clearly separated from the two dimensional sets in $U_{d_s} \setminus U^*$. The same holds for the one three-dimensional term $\{4,8,9\} \in U^*$. The size of the intervals $I^{(j)}$ suitable to choose the parameters ε_j is especially relevant since it separates *important* from *unimportant* terms.

Since there exists $\boldsymbol{N}$, and $\boldsymbol{\varepsilon}$ such that we are able to recover the set of ANOVA terms U^*, we set $U_{X,\boldsymbol{y}}^{(\boldsymbol{\varepsilon})} = U^*$ from now on. We aim to improve our approximation quality with the given data by solving the minimization problem (5.15). Here, we could choose individual index sets for every ANOVA term in U^* to form $I(U^*)$ based on the global sensitivity indices, but for our function order-dependence can be maintained. Table 2 shows the results of the approximation using the index set $I(U^*)$.

	size of index sets		relative errors		
	$\boldsymbol{N}$	$|I(U_{d_s})|$	ε_{ℓ_2}	$\varepsilon_{\mathrm{L}_2}$	$I^{(1)}, I^{(2)}, I^{(3)}$
1	$[256, 32, 8]$	65704	$4.7 \cdot 10^{-3}$	$4.8 \cdot 10^{-3}$	$(0.0, 0.021)$ $(3.0 \cdot 10^{-8}, 0.019)$ $(1.2 \cdot 10^{-8}, 0.026)$
2	$[256, 32, 16]$	320392	$2.1 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$(0.0, 0.021)$ $(7.2 \cdot 10^{-9}, 0.019)$ $(2.5 \cdot 10^{-8}, 0.026)$
3	$[256, 32, 32]$	2539336	$2.6 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	$(0.0, 0.016)$ $(8.3 \cdot 10^{-5}, 0.015)$ $(2.5 \cdot 10^{-3}, 0.023)$
4	$[256, 64, 8]$	173992	$4.4 \cdot 10^{-3}$	$4.7 \cdot 10^{-3}$	$(0.0, 0.021)$ $(1.1 \cdot 10^{-7}, 0.019)$ $(1.1 \cdot 10^{-8}, 0.026)$
5	$[256, 64, 16]$	428680	$1.6 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$(0.0, 0.021)$ $(1.8 \cdot 10^{-8}, 0.019)$ $(1.6 \cdot 10^{-8}, 0.026)$
6	$[256, 64, 32]$	2647624	$2.5 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$	$(0.0, 0.015)$ $(4.0 \cdot 10^{-4}, 0.015)$ $(2.9 \cdot 10^{-3}, 0.022)$
7	$[512, 64, 8]$	176296	$4.4 \cdot 10^{-3}$	$4.7 \cdot 10^{-3}$	$(0.0, 0.021)$ $(1.1 \cdot 10^{-7}, 0.019)$ $(1.1 \cdot 10^{-8}, 0.026)$
8	$[512, 64, 16]$	430984	$1.6 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$(0.0, 0.021)$ $(1.8 \cdot 10^{-8}, 0.019)$ $(1.6 \cdot 10^{-8}, 0.026)$
9	$[512, 64, 32]$	2649928	$2.5 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$	$(0.0, 0.015)$ $(4.0 \cdot 10^{-4}, 0.015)$ $(2.9 \cdot 10^{-3}, 0.022)$

Table 1: Results of detection step for important ANOVA terms with $M = 2.5 \cdot 10^6$ uniformly distributed nodes ($\boldsymbol{N} = [N_1, N_2, N_3]$).

The number of terms in U^* is significantly smaller than in U_{d_s} such that we are able to increase $\boldsymbol{N}$ while balancing the effects of over- and underfitting. We observe that the reduction of the ANOVA terms to U^* yields benefit with regard to approximation quality due to the reduction in model complexity.

Now that we have experiments with no truncation error in the ANOVA decomposition, i.e., $\mathrm{T}_3 f = f$, we repeat the tests with a superposition threshold $d_s = 2$. In this case, it is not possible to detect the ANOVA term $f_{\{4,8,9\}}$ which results in the set $U^+ := U^* \setminus \{4, 8, 9\}$ being optimal for the detection step. For the following tests, we use the same nodes as we did previously.

The results of the experiments in Table 3 show that it is possible to determine the terms in U^+. Since three-dimensional terms are not included, the term $f_{\{4,8,9\}}$ is not in the approximation which results in the larger errors compared to Table 1.
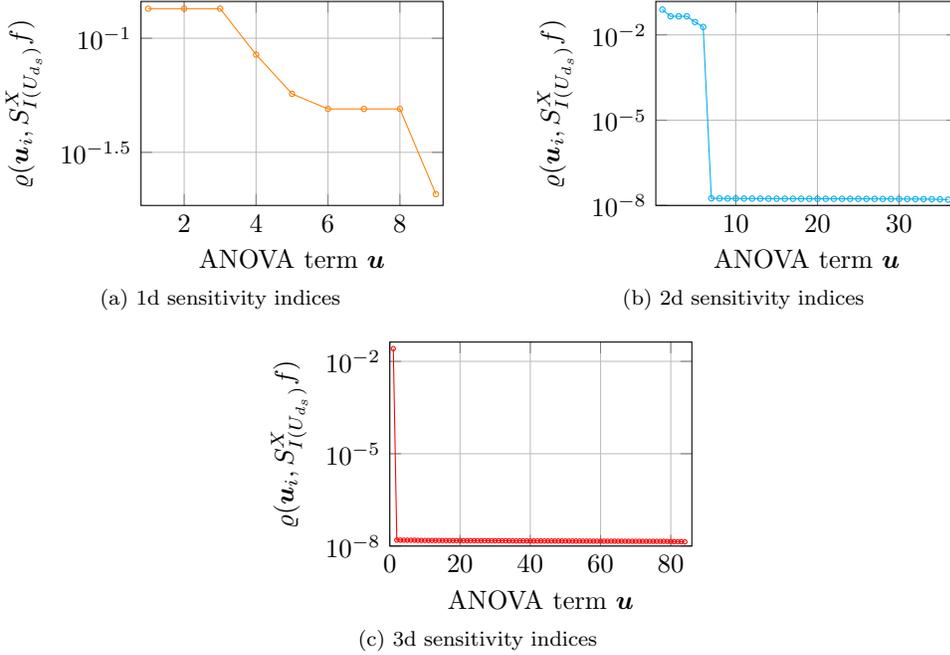
(a) 1d sensitivity indices

(b) 2d sensitivity indices

(c) 3d sensitivity indices

Fig. 4: Behavior of the global sensitivity indices $\varrho(\boldsymbol{u}, S_{I(U_{d_s})}^X f)$ for the example 8 from Table 1.

	size of index sets		relative errors	
	$\boldsymbol{N}$	$|I(U^*)|$	ε_{ℓ_2}	$\varepsilon_{\mathrm{L}_2}$
1	$[1024, 64, 64]$	283069	$5.6 \cdot 10^{-4}$	$6.3 \cdot 10^{-4}$
2	$[1024, 128, 32]$	135773	$6.0 \cdot 10^{-4}$	$6.4 \cdot 10^{-4}$
4	$[1024, 128, 64]$	356029	$2.7 \cdot 10^{-4}$	$3.1 \cdot 10^{-4}$
5	$[1024, 256, 64]$	649405	$2.0 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$

Table 2: Results for approximation with active set U^* and $M = 2.5 \cdot 10^6$ uniformly distributed nodes ($\boldsymbol{N} = [N_1, N_2, N_3]$).

Since there exists $\boldsymbol{N} \in \mathbb{N}^2$ and $\varepsilon > 0$ such that $U_{X,\boldsymbol{y}}^{(\boldsymbol{\varepsilon})} = U^+$, we use U^+ for the next approximation step with suitable index sets $I(U^+)$. The results for different choices of N_1 and N_2 are displayed in Table 4. We are able to achieve better errors with the smaller index sets. Obviously, the influence of the cutoff error is dominating such that a large benefit in taking many additional frequencies cannot be observed.

7.2. Black-Box Approximation. In the following numerical experiments we aim to find reconstructing rank-1 lattice, see Subsection 2.1, for the function f. In the first step, our goal is to determine the set of ANOVA terms U^* and later use it to improve our approximation quality. As discussed in [28], the function f works well with hyperbolic cross index sets of dominating mixed smoothness 3/2. Therefore, we

	size of index sets		relative errors		$I^{(1)}, I^{(2)}$
	$\boldsymbol{N}$	$|I(U_{d_s})|$	ε_{ℓ_2}	$\varepsilon_{\mathrm{L}_2}$	
1	[256, 16]	10396	$9.4 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$(0.0, 0.021)$ $(3.0 \cdot 10^{-6}, 0.020)$
2	[256, 32]	36892	$9.3 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$(0.0, 0.021)$ $(3.0 \cdot 10^{-6}, 0.020)$
3	[256, 64]	145180	$9.1 \cdot 10^{-2}$	$9.6 \cdot 10^{-2}$	$(0.0, 0.021)$ $(4.8 \cdot 10^{-5}, 0.020)$
4	[256, 128]	582940	$8.2 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$(0.0, 0.021)$ $(2.3 \cdot 10^{-4}, 0.020)$

Table 3: Results of detection step for important ANOVA terms with $M = 2.5 \cdot 10^6$ uniformly distributed nodes and superposition threshold $d_s = 2$, $(\boldsymbol{N} = [N_1, N_2])$.

	size of index sets		relative errors	
	$\boldsymbol{N}$	$|I(U^+)|$	ε_{ℓ_2}	$\varepsilon_{\mathrm{L}_2}$
1	[1024, 16]	10558	$9.3 \cdot 10^{-2}$	$9.3 \cdot 10^{-2}$
2	[1024, 32]	14974	$9.3 \cdot 10^{-2}$	$9.3 \cdot 10^{-2}$
3	[1024, 64]	33022	$9.3 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$
4	[1024, 128]	105982	$9.1 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$

Table 4: Approximation results for active set U^+ with $M = 2.5 \cdot 10^6$ uniformly distributed nodes $(\boldsymbol{N} = [N_1, N_2])$.

define

$$(7.3) \qquad \mathcal{H}_j^N = \left\{ \boldsymbol{k} \in \mathbb{Z}^j \colon \prod_{s \in \mathrm{supp}\,\boldsymbol{k}} (1 + |k_s|)^{\frac{3}{2}} \leq N \right\}, N \in \mathbb{N}.$$

We choose as order-dependent index sets $I_0 = \{0\}$, $I_1 = \mathcal{H}_1^{N_1}$, $I_2 = \mathcal{H}_2^{N_2}$, and $I_3 = \mathcal{H}_3^{N_1}$ with $N_1, N_2, N_3 \in \mathbb{N}$ to obtain $I(U_{d_s})$ as in (5.3). The method then gives us a reconstructing rank-1 lattice $X := \Lambda(\boldsymbol{z}, M, I(U_{d_s}))$ with generating vector $\boldsymbol{z} \in \mathbb{Z}^9$ and lattice sizes $M \in \mathbb{N}$ by employing the component-by-component construction from [43, Algorithm 8.17]. The approximation is defined as $S_{I(U_{d_s})}^X f$.

Table 5 shows results of numerical experiments with f, see (7.1), and different choices for the parameters N_1, N_2, and N_3. We can see that there exist an $\boldsymbol{\varepsilon}$ such that it is possible to detect the active set of terms U^* in every test scenario. The lattice size increases with the growing index set as expected. Note that is sufficient to use an index set of 3481 frequencies and a lattice with only 46351 evaluations in order to detect the active set of ANOVA terms.

Now, we set the active set $U_{X,\boldsymbol{y}}^{(\boldsymbol{\varepsilon})} = U^*$. The aim is again to improve our approximation quality by solving the minimization problem (5.15). We also maintain the order-dependence of the set $I(U^*)$ based on the structure of the function. Table 6 shows the results of the approximation using the index set $I(U^*)$. Larger cutoff pa-

rameters N_i become possible such that we are able to achieve a good approximation error with relatively small lattice sizes in relation to our problem dimension. The sizes of our reconstructing lattices stay manageable as well.

size of index sets		relative errors			
$\boldsymbol{N}$	$|I(U_{d_s})|$	ε_{ℓ_2}	$\varepsilon_{\mathrm{L}_2}$	M	$I^{(1)}, I^{(2)}, I^{(3)}$
$[10^2, 10^2, 10^2]$	3481	$2.8 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	47351	$(0.0, 0.021)$
					$(3.4 \cdot 10^{-5}, 0.019)$
					$(5.7 \cdot 10^{-5}, 0.025)$
$[10^3, 10^3, 10^3]$	11203	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	490277	$(0.0, 0.021)$
					$(1.7 \cdot 10^{-7}, 0.019)$
					$(5.7 \cdot 10^{-7}, 0.026)$
$[10^4, 10^4, 10^3]$	16891	$7.0 \cdot 10^{-3}$	$7.1 \cdot 10^{-3}$	1114489	$(0.0, 0.021)$
					$(3.5 \cdot 10^{-10}, 0.019)$
					$(5.3 \cdot 10^{-8}, 0.026)$
$[10^5, 10^4, 10^3]$	17341	$7.0 \cdot 10^{-3}$	$7.0 \cdot 10^{-3}$	2349307	$(0.0, 0.021)$
					$(1.7 \cdot 10^{-9}, 0.019)$
					$(2.8 \cdot 10^{-9}, 0.026)$

Table 5: Results of detection step for important ANOVA terms ($\boldsymbol{N} = [N_1, N_2, N_3]$).

size of index sets		relative errors		
$\boldsymbol{N}$	$|I(U_{d_s})|$	ε_{ℓ_2}	$\varepsilon_{\mathrm{L}_2}$	M
$[10^4, 10^4, 10^4]$	2243	$2.4 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	157243
$[10^5, 10^5, 10^5]$	6565	$8.3 \cdot 10^{-4}$	$8.3 \cdot 10^{-4}$	1346881
$[10^6, 10^5, 10^5]$	7591	$7.7 \cdot 10^{-4}$	$7.7 \cdot 10^{-4}$	883391
$[10^6, 10^6, 10^5]$	13495	$5.0 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	5691109

Table 6: Results of detection step for important ANOVA terms ($\boldsymbol{N} = [N_1, N_2, N_3]$).

8. Summary. In this paper we considered the classical ANOVA decomposition for periodic functions. We studied different index sets $\mathbb{P}_{\boldsymbol{u}}^{(d)}$ and $\mathbb{F}_{\boldsymbol{u}}^{(d)}$ for the projections $\mathrm{P}_{\boldsymbol{u}}f$ and ANOVA terms $f_{\boldsymbol{u}}$, respectively, and proved their properties as well as formulas for the Fourier coefficients. For functions in Sobolev type spaces $\mathrm{H}^w(\mathbb{T}^d)$ and the weighted Wiener algebra $\mathcal{A}^w(\mathbb{T}^d)$ we showed that a function inherits its smoothness to both the projections and ANOVA terms.

Moreover, we related the smoothness of a function characterized by the decay of its Fourier coefficients to the class of functions of a low-dimensional structure and considered relative errors for L_∞ and L_2 weighed by the corresponding Sobolev and Wiener algebra norms. This lead to an upper bound for the modified superposition dimension $\mathrm{d}^{(\mathrm{sp})}$ in those spaces. For product and order-dependent weights $w^{\alpha,\beta}$ we were able to obtain specific bounds.

We introduced an approximation method for high-dimensional functions that are of a low-dimensional structure in Section 5. The method can be employed in black-

box and scattered data approximation. In the former scenario one needs a special discretization for the index sets of type $I(U)$, e.g., rank-1 lattice, and in the latter an algorithm to realize an efficient multiplication with the Fourier matrices. We proved results for the error of the method, see Section 6, in L_2 and L_∞. An L_∞ bound in the scattered data case for the aliasing error in (6.4) is still open. Here, one needs to consider estimating the quantity

$$\sup_{\|f\|_{\mathcal{A}^w(\mathbb{T}^d)} \leq 1} \frac{1}{|X|} \sum_{\boldsymbol{x} \in X} \left| \left(f - S_{I(U)} f \right) (\boldsymbol{x}) \right|.$$

Numerical experiments with a benchmark function were successfully performed in Section 7. The active set detection works well for this function in both approximation scenarios and even for small degrees of trigonometric polynomials. A definite goal is to perform experiments on real-world data sets and try to determine attribute rankings.

Moreover, it is possible to consider a similar analysis of the ANOVA decomposition in weighted Lebesgue spaces with orthogonal polynomials as bases, e.g., the Chebyshev system. This would also allow a generalization of the approximation method to a non-periodic setting, cf. [44].

REFERENCES

[1] J. BALDEAUX AND M. GNEWUCH, *Optimal randomized multilevel algorithms for infinite-dimensional integration on function spaces with ANOVA-type decomposition*, SIAM J. Numer. Anal., 52 (2014), pp. 1128–1155, https://doi.org/10.1137/120896001.

[2] R. F. BASS AND K. GRÖCHENIG, *Random sampling of multivariate trigonometric polynomials*, SIAM J. Math. Anal., 36 (2004), pp. 773–795, https://doi.org/10.1137/S0036141003432316.

[3] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, USA, 1996.

[4] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numer., 13 (2004), pp. 147–269, https://doi.org/10.1017/S0962492904000182.

[5] G. BYRENHEID, L. KÄMMERER, T. ULLRICH, AND T. VOLKMER, *Tight error bounds for rank-1 lattice sampling in spaces of hybrid mixed smoothness*, Numer. Math., 136 (2017), pp. 993–1034, https://doi.org/10.1007/s00211-016-0861-7.

[6] R. CAFLISCH, W. MOROKOFF, AND A. OWEN, *Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension*, J. Comput. Finance, 1 (1997), pp. 27–46, https://doi.org/10.21314/jcf.1997.005.

[7] D. DŨNG, V. N. TEMLYAKOV, AND T. ULLRICH, *Hyperbolic Cross Approximation*, Advanced Courses in Mathematics – CRM Barcelona, Birkhäuser, Cham, 2018.

[8] J. DICK, F. Y. KUO, AND I. H. SLOAN, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer., 22 (2013), pp. 133–288, https://doi.org/10.1017/S0962492913000044.

[9] J. DICK, I. H. SLOAN, X. WANG, AND H. WOŹNIAKOWSKI, *Good lattice rules in weighted Korobov spaces with general weights*, Numer. Math., 103 (2006), pp. 63–97, https://doi.org/10.1007/s00211-005-0674-6.

[10] M. DÖHLER, S. KUNIS, AND D. POTTS, *Nonequispaced hyperbolic cross fast Fourier transform*, SIAM J. Numer. Anal., 47 (2010), pp. 4415 – 4428, https://doi.org/10.1137/090754947.

[11] A. D. GILBERT, F. Y. KUO, D. NUYENS, AND G. W. WASILKOWSKI, *Efficient implementations of the multivariate decomposition method for approximating infinite-variate integrals*, SIAM

J. Sci. Comput., 40 (2018), pp. A3240–A3266, https://doi.org/10.1137/17m1161890.

[12] M. Gnewuch, M. Hefter, A. Hinrichs, and K. Ritter, *Embeddings of weighted Hilbert spaces and applications to multivariate and infinite-dimensional integration*, J. Approx. Theory, 222 (2017), pp. 8–39, https://doi.org/10.1016/j.jat.2017.05.003.

[13] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients*, Numer. Math., 131 (2014), pp. 329–368, https://doi.org/10.1007/s00211-014-0689-y.

[14] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan, *Circulant embedding with QMC: analysis for elliptic PDE with lognormal coefficients*, Numer. Math., 140 (2018), pp. 479–511, https://doi.org/10.1007/s00211-018-0968-0.

[15] M. Griebel, *Sparse grids and related approximation schemes for higher dimensional problems*, in Foundations of Computational Mathematics (FoCM05), Santander, L. Pardo, A. Pinkus, E. Suli, and M. Todd, eds., Cambridge University Press, 2006, pp. 106–161.

[16] M. Griebel and J. Hamaekers, *Fast discrete Fourier transform on generalized sparse grids*, in Sparse Grids and Applications - Munich 2012, J. Garcke and D. Pflüger, eds., vol. 97 of Lect. Notes Comput. Sci. Eng., Springer International Publishing, 2014, pp. 75–107, https://doi.org/10.1007/978-3-319-04537-5_4.

[17] M. Griebel and M. Holtz, *Dimension-wise integration of high-dimensional functions with applications to finance*, J. Complexity, 26 (2010), pp. 455–489, https://doi.org/10.1016/j.jco.2010.06.001.

[18] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The smoothing effect of the ANOVA decomposition*, J. Complexity, 26 (2010), pp. 523–551, https://doi.org/10.1016/j.jco.2010.04.003.

[19] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The ANOVA decomposition of a non-smooth function of infinitely many variables can have every term smooth*, Math. Comp., 86 (2016), pp. 1855–1876, https://doi.org/10.1090/mcom/3171.

[20] K. Groechenig, B. M. Poetscher, and H. Rauhut, *Learning trigonometric polynomials from random samples and exponential inequalities for eigenvalues of random matrices*, Tech. Report math.PR/0701781, 2007, http://cds.cern.ch/record/1013574.

[21] M. Hegland, *Adaptive sparse grids*, in Proc. of 10th Computational Techniques and Applications Conference CTAC-2001, K. Burrage and R. B. Sidje, eds., vol. 44, 2003, pp. C335–C353, https://doi.org/10.21914/anziamj.v44i0.685.

[22] M. Hegland and P. Leopardi, *The rate of convergence of sparse grid quadrature on the torus*, ANZIAM Journal, 52 (2011), https://doi.org/10.21914/anziamj.v52i0.3952.

[23] M. Holtz, *Sparse grid quadrature in high dimensions with applications in finance and insurance*, vol. 77 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 2011, https://doi.org/10.1007/978-3-642-16004-2.

[24] L. Kämmerer, *Reconstructing hyperbolic cross trigonometric polynomials by sampling along rank-1 lattices*, SIAM J. Numer. Anal., 51 (2013), pp. 2773–2796, http://dx.doi.org/10.1137/120871183.

[25] L. Kämmerer, *Reconstructing multivariate trigonometric polynomials from samples along rank-1 lattices*, in Approximation Theory XIV: San Antonio 2013, G. E. Fasshauer and L. L. Schumaker, eds., Springer International Publishing, 2014, pp. 255–271, https://doi.org/10.1007/978-3-319-06404-8_14.

[26] L. Kämmerer, D. Potts, and T. Volkmer, *Approximation of multivariate periodic functions by trigonometric polynomials based on rank-1 lattice sampling*, J. Complexity, 31 (2015), pp. 543–576, https://doi.org/10.1016/j.jco.2015.02.004.

[27] L. Kämmerer, D. Potts, and T. Volkmer, *Approximation of multivariate periodic functions by trigonometric polynomials based on sampling along rank-1 lattice with generating vector of Korobov form*, J. Complexity, 31 (2015), pp. 424–456, https://doi.org/10.1016/j.jco.2014.09.001.

[28] L. Kämmerer, D. Potts, and T. Volkmer, *High-dimensional sparse FFT based on sampling along multiple rank-1 lattices*, Appl. Comput. Harmon. Anal., 51 (2021), pp. 225–257, https://doi.org/10.1016/j.acha.2020.11.002.

[29] L. Kämmerer, T. Ullrich, and T. Volkmer, *Worst-case recovery guarantees for least squares approximation using random samples*, ArXiv e-prints 1911.10111, (2019).

[30] L. Kämmerer and T. Volkmer, *Approximation of multivariate periodic functions based on sampling along multiple rank-1 lattices*, J. Approx. Theory, 246 (2019), pp. 1–27, https://doi.org/10.1016/j.jat.2019.05.001.

[31] J. Keiner, S. Kunis, and D. Potts, *Using NFFT3 - a software library for various nonequispaced fast Fourier transforms*, ACM Trans. Math. Software, 36 (2009), pp. Article 19, 1–30, https://doi.org/10.1145/1555386.1555388.

[32] T. Kühn, S. Mayer, and T. Ullrich, *Counting via entropy: New preasymptotics for the*

approximation numbers of Sobolev embeddings, SIAM J. Numer. Anal., 54 (2016), pp. 3625–3647, https://doi.org/10.1137/16m106580x.

[33] F. Y. KUO AND D. NUYENS, *Application of quasi-Monte Carlo methods to elliptic PDEs with random diffusion coefficients: A survey of analysis and implementation*, Found. Comput. Math, 16 (2016), pp. 1631–1696, https://doi.org/10.1007/s10208-016-9329-5.

[34] F. Y. KUO, D. NUYENS, L. PLASKOTA, I. H. SLOAN, AND G. W. WASILKOWSKI, *Infinite-dimensional integration and the multivariate decomposition method*, J. Comput. Appl. Math., 326 (2017), pp. 217–234, https://doi.org/10.1016/j.cam.2017.05.031.

[35] F. Y. KUO, C. SCHWAB, AND I. H. SLOAN, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 3351 – 3374, https://doi.org/10.1137/110845537.

[36] F. Y. KUO, I. H. SLOAN, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, *On decompositions of multivariate functions*, Math. Comp., 79 (2010), pp. 953–966, https://doi.org/10.1090/s0025-5718-09-02319-9.

[37] D. LI AND F. J. HICKERNELL, *Trigonometric spectral collocation methods on lattices*, in Recent Advances in Scientific Computing and Partial Differential Equations, S. Y. Cheng, C.-W. Shu, and T. Tang, eds., vol. 330 of Contemp. Math., AMS, 2003, pp. 121–132, https://doi.org/10.1090/conm/330/05887.

[38] R. LIU AND A. B. OWEN, *Estimating mean dimensionality of analysis of variance decompositions*, J. Amer. Statist. Assoc., 101 (2006), pp. 712–721, https://doi.org/10.1198/016214505000001410.

[39] M. MOELLER AND T. ULLRICH, L_2-*norm sampling discretization and recovery of functions from RKHS with finite trace*, ArXiv e-prints 2009.11940, (2020).

[40] H. NIEDERREITER, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992, https://doi.org/10.1137/1.9781611970081.

[41] A. OWEN, *Effective dimension of some weighted pre-Sobolev spaces with dominating mixed partial derivatives*, SIAM J. Numer. Anal., 57 (2019), pp. 547–562, https://doi.org/10.1137/17m1158975.

[42] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71, https://doi.org/10.1145/355984.355989.

[43] G. PLONKA, D. POTTS, G. STEIDL, AND M. TASCHE, *Numerical Fourier Analysis*, Applied and Numerical Harmonic Analysis, Birkhäuser, 2018, https://doi.org/10.1007/978-3-030-04306-3.

[44] D. POTTS AND M. SCHMISCHKE, *Learning multivariate functions with low-dimensional structures using polynomial bases*, ArXiv e-prints 1912.03195, (2019).

[45] H. RABITZ AND O. F. ALIS, *General foundations of high dimensional model representations*, J. Math. Chem., 25 (1999), pp. 197–233, https://doi.org/10.1023/A:1019188517934.

[46] I. H. SLOAN AND H. WOŹNIAKOWSKI, *Tractability of multivariate integration for weighted Korobov classes*, J. Complexity, 17 (2001), pp. 697–721, https://doi.org/10.1006/jcom.2001.0599.

[47] I. M. SOBOL, *On sensitivity estimation for nonlinear mathematical models*, Keldysh Applied Mathematics Institute, 1 (1990), pp. 112–118.

[48] I. M. SOBOL, *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Math. Comput. Simulation, 55 (2001), pp. 271–280, https://doi.org/10.1016/s0378-4754(00)00270-6.