TECHNISCHE UNIVERSITÄT
CHEMNITZ

Faculty of Mathematics

Master thesis

A Fourier approach to learning sparse additive models

Michael Schmischke

Born March 1, 1996 in Rodewisch

Chemnitz, April 15, 2019

First advisor: Prof. Dr. Daniel Potts

Second advisor: Dr. Lutz Kämmerer

Contents

List of Figures

List of Tables

1 Introduction

In the Information Age high dimensional data plays an ever increasing role in our society and economy. We are trying to understand more and more complex relationships or, mathematically speaking, functions. In general, working with high dimensional functions, e.g. with quadrature, results in reaching limitations of standard methods due to the curse of dimensionality - an expression that was first introduced by the mathematician Richard Bellman. Simply speaking, this translates to the amount of required data growing exponentially with the dimension. Working around this issue, e.g. with sparsity assumptions, is a central field in approximation theory. Besides approximating functions we are also interested in finding the important dimensions and dimension interactions of a function which is particularly important for applications where a lot of data is collected.

This thesis touches a specific problem in the wide area of approximation of high dimensional functions. We consider periodic functions on the manifold of the d-variate torus. The basis of our approach builds the well-known analysis of variance (ANOVA) decomposition [13, 3, 4, 17]. Analysis of variance is best known for its use in statistics to investigate relationships between groups through the one-way ANOVA, two-way ANOVA or N-way ANOVA. We start by studying the properties of the ANOVA decomposition on the torus and relating it to the Fourier analysis of periodic functions in Section 3. We find representations for the involved projections and ANOVA terms in the frequency domain in Lemma 3.2 and Corollary 3.6. Furthermore, we use the notion of inheritance of smoothness from the function to its ANOVA terms proposed in [3] and prove this for Sobolev type spaces, see Theorem 3.19, and the weighted Wiener algebra, see Theorem 3.22.

The ANOVA decomposition also offers tools to identify the important dimension interactions with regard to the variance of the function. The described notion of effective dimensions leads to a certain type of sparsity, i.e., we assume that a large part of the variance of a function can be explained by considering only low-dimensional interactions of the variables. This leads us to an approximate model, see Definition 3.27, where the number of ANOVA terms grows polynomial with the dimension. We prove error bounds for approximation with this model in L_2 for Sobolev type spaces, see Theorem 3.32, and L_∞ for the weighted Wiener algebra, see Theorem 3.35.

Subsequently, we consider two scenarios for function approximation. In Section 4 we assume to have a function with black-box-access, i.e., we can evaluate the function at any point. This can be a probable scenario in applications where evaluations are cheap. We use the previously introduced approximate model to detect important dimension interactions and use the information we gained from this to calculate an approximation. Here, we apply the well-known theory surrounding rank-1 lattices as sampling schemes [7, 18, Chapter 8] and the high-dimensional FFT. Combining results from Section 3 and [18, Chapter 8], we prove an L_∞ error bound for this method, see Theorem 4.7

In Section 5 we are working with given scattered data. In applications this can be related to the fact that function evaluations are expensive or that data stems only from a certain range or time. The basis of the approach stays similar to the previous scenario, we use the approximate model to detect the important dimension interactions. This is done by solving a least-squares problem with the help of the Nonequispaced Fast Fourier Transform (NFFT) [10]. Again, we take this information to find an approximation for the function.

In Sections 4 and 5 we present numerical examples for a 9-dimensional test function. The results show that the proposed methods work well in both scenarios for the considered test function.

2 Fundamentals

In this section we introduce the important fundamentals for the considerations in this thesis. We start with Fourier series together with Lebesgue spaces and move on to Sobolev spaces where we characterize the smoothness of functions by the decay of Fourier coefficients as considered in [11]. The section is based on the book by Folland [1].

Furthermore, we consider rank-1 lattices as sampling schemes to reconstruct high dimensional functions as proposed in [18]. Specifically, we discuss the fast evaluation of trigonometric polynomials and the efficient reconstruction of Fourier coefficients.

2.1 Fourier Series, Lebesgue Spaces and Sobolev Spaces

In this thesis we are working exclusively with 1-periodic functions

$$f : \mathbb{T}^d \to \mathbb{C}$$

on the manifold of the d-dimensional torus $\mathbb{T}^d$. Depending on the context, we identify $\mathbb{T} \cong [0, 1)$ or $\mathbb{T} \cong [-1/2, 1/2)$. First, we define the Lebesgue function spaces over the torus.

Definition 2.1 The space

$$\mathrm{L}_p(\mathbb{T}^d) := \left\{ f : \mathbb{T}^d \to \mathbb{C}\colon \; \|f\|_{\mathrm{L}_p(\mathbb{T}^d)} < \infty \right\}$$

equipped with the norm

$$\|f\|_{\mathrm{L}_p(\mathbb{T}^d)} = \begin{cases} \left(\int_{\mathbb{T}^d} |f(\boldsymbol{x})|^p \,\mathrm{d}\boldsymbol{x}\right)^{\frac{1}{p}} & : 1 \le p < \infty \\ \operatorname{ess\,sup} |f(\boldsymbol{x})| & : p = \infty \end{cases}$$

is called **Lebesgue function space** (over the torus) with parameter $1 \le p \le \infty$. ◇

Theorem 2.2 *For the Lebesgue function spaces we have the following embeddings*

$$\mathrm{L}_p(\mathbb{T}^d) \subset \mathrm{L}_q(\mathbb{T}^d)$$

for $1 \le q \le p \le \infty$.

Proof. Let $f \in \mathrm{L}_p(\mathbb{T}^d)$. We prove this statement for $1 \le q \le p < \infty$ using Hölder's inequality

$$\|f\|^q_{\mathrm{L}_q(\mathbb{T}^d)} = \int_{\mathbb{T}^d} |f(\boldsymbol{x})|^q \,\mathrm{d}\boldsymbol{x} \le \left(\int_{\mathbb{T}^d} |f(\boldsymbol{x})|^p \,\mathrm{d}\boldsymbol{x}\right)^{\frac{q}{p}} \left(\int_{\mathbb{T}^d} 1 \,\mathrm{d}\boldsymbol{x}\right)^{1-\frac{q}{p}} < \infty.$$

For the case $p = \infty$ we have

$$\|f\|^q_{\mathrm{L}_q(\mathbb{T}^d)} = \int_{\mathbb{T}^d} |f(\boldsymbol{x})|^q \,\mathrm{d}\boldsymbol{x} \le C \cdot \operatorname{ess\,sup} |f(\boldsymbol{x})|^q < \infty.$$

■

If the function f is integrable over $\mathbb{T}^d$ it can be expressed as a Fourier series.

Definition 2.3 If $f : \mathbb{T}^d \to \mathbb{C}$ is integrable, i.e., $f \in \mathrm{L}_1(\mathbb{T}^d)$, we call

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \tag{2.1}$$

the **Fourier series** of f with **Fourier coefficients**

$$\hat{f}_{\boldsymbol{k}} = \int_{\mathbb{T}^d} f(\boldsymbol{x}) \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \, \mathrm{d}\boldsymbol{x}.$$

◇

Furthermore, we will mainly focus on $\mathrm{L}_2(\mathbb{T}^d)$ where $\left(\mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}\right)_{\boldsymbol{k} \in \mathbb{Z}^d}$ is a complete orthonormal system, see [1, Theorem 3.5]. This property leads to the following results.

Lemma 2.4 *If $f \in \mathrm{L}_2(\mathbb{T}^d)$ then the series in* (2.1) *converges in the* L_2*-norm.*

Proof. see [1, Theorem 3.4] ■

Theorem 2.5 *Let $f, g \in \mathrm{L}_2(\mathbb{T}^d)$. Then f and g are equal if and only if*

$$\hat{f}_{\boldsymbol{k}} = \hat{g}_{\boldsymbol{k}} \ \forall \boldsymbol{k} \in \mathbb{Z}^d.$$

Proof. see [1, Corollary 2.2] ■

Theorem 2.6 (Parseval's identity) *For a function $f \in \mathrm{L}_2(\mathbb{T}^d)$ with Fourier coefficients $\hat{f}_{\boldsymbol{k}}, \boldsymbol{k} \in \mathbb{Z}^d$, we have* ***Parseval's identity***

$$\|f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \left| \hat{f}_{\boldsymbol{k}} \right|^2. \tag{2.2}$$

Proof. see [1, Theorem 3.4] ■

In the following we consider the smoothness of functions on the torus and classify them by the decay of their Fourier coefficients. To this end, we first define the p-Norm and subsequently the Sobolev spaces and weights as proposed in [11].

Definition 2.7 The p-**Norm** of a vector $\boldsymbol{x} \in \mathbb{R}^d$ is defined as

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}} & : p > 0 \\ \max_{i=1,2,\dots,d} |x_i| & : p = \infty \\ |\{i \colon x_i \neq 0\}| & : p = 0 \end{cases}.$$

While it is a norm for $1 \le p \le \infty$, it is only a quasi-norm for $0 < p < 1$. ◇

Definition 2.8 The **Sobolev type spaces** are defined as

$$\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d) = \{f \in \mathrm{L}_2(\mathbb{T}^d)\colon\ \|f\|_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)} < \infty\}$$

with the norm

$$\|f\|_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)} = \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d} w^2(\boldsymbol{k}) \left|\hat{f}_{\boldsymbol{k}}\right|^2\right)^{\frac{1}{2}}.$$

for the weight sequence $\boldsymbol{w} = (w(\boldsymbol{k}))_{\boldsymbol{k}\in\mathbb{Z}^d}$ with a weight function $w : \mathbb{Z}^d \to [1, \infty)$. ◇

Definition 2.9 The **classical isotropic Sobolev spaces** $\mathrm{H}^{s,p}(\mathbb{T}^d)$ are a special case of the Sobolev type spaces $\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$ from Definition 2.8 where we choose the weight functions

$$w_{s,p}(\boldsymbol{k}) = (1 + \|\boldsymbol{k}\|_p^p)^{\frac{s}{p}}, p \in (0, \infty)$$
$$w_{s,p}(\boldsymbol{k}) = \max(1, |k_1|, \dots, |k_d|), p = \infty$$

for $s > 0$. Furthermore, we define $\mathrm{H}^s(\mathbb{T}^d) \coloneqq \mathrm{H}^{s,2}(\mathbb{T}^d)$. ◇

Definition 2.10 The **Sobolev spaces of dominating mixed smoothness** $\mathrm{H}^s_{\mathrm{mix}}(\mathbb{T}^d)$ are a second special case of the Sobolev type spaces $\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$ from Definition 2.8 where we choose the weight function

$$w_s(\boldsymbol{k}) = \prod_{i=1}^{d}(1 + |k_i|^2)^s.$$

◇

Subsequently, we consider the weighted Wiener algebra of functions and its embeddings into other function spaces.

Definition 2.11 The weighted **Wiener algebra** is defined as

$$\mathcal{A}_w(\mathbb{T}^d) = \{f \in \mathrm{L}_1(\mathbb{T}^d)\colon\ \|f\|_{\mathcal{A}_w} < \infty\}$$

with the norm

$$\|f\|_{\mathcal{A}_w} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} w(\boldsymbol{k}) \left|\hat{f}_{\boldsymbol{k}}\right|.$$

Here, $w : \mathbb{Z}^d \to [1, \infty)$ is a weight function. Furthermore, we define

$$\mathcal{A}_1(\mathbb{T}^d) \eqqcolon \mathcal{A}(\mathbb{T}^d),$$

where 1 is the function w with $w(\boldsymbol{k}) = 1$ for all $\boldsymbol{k} \in \mathbb{Z}^d$. ◇

Theorem 2.12 *For the Wiener algebra and the continuous functions we have the embeddings*

$$\mathcal{A}_w(\mathbb{T}^d) \subset \mathcal{A}(\mathbb{T}^d) \subset \mathrm{C}(\mathbb{T}^d).$$

Proof. see [18, Lemma 8.2] ■

2.2 Rank-1 Lattice as Sampling Schemes

Now, we consider rank-1 lattices as sampling schemes to reconstruct high-dimensional functions. This section is based on [18, Chapter 8] and we restrict ourselves to a basic overview and mention results that are especially relevant to our later work.

In general, a function $f \in \mathrm{L}_1(\mathbb{T}^d)$ has an infinite number of Fourier coefficients $\hat{f}_{\boldsymbol{k}}$. In this section, we consider approximations using Fourier partial sums

$$(S_I f)(\boldsymbol{x}) \coloneqq \sum_{\boldsymbol{k}\in I} \hat{f}_{\boldsymbol{k}}\, \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \tag{2.3}$$

with a finite index set $I \subset \mathbb{Z}^d$. The Fourier partial sum $S_I f$ is a trigonometric polynomial. This gives rise to the following considerations.

Given a trigonometric polynomial $p(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in I} \hat{p}_{\boldsymbol{k}}\, \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}$ and a sampling set $X \subset \mathbb{T}^d$ with $|X| \in \mathbb{N}$ sampling nodes, we consider the two questions:

1. For given Fourier coefficients $\hat{p}_{\boldsymbol{k}}$ how to compute the values $p(\boldsymbol{x})$, $\boldsymbol{x} \in X$, efficiently?
2. How can we efficiently compute the Fourier coefficients $\hat{p}_{\boldsymbol{k}}$, given the evaluations $p(\boldsymbol{x})$, $\boldsymbol{x} \in X$?

Let us consider the Fourier matrix

$$\boldsymbol{F} = \left(\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\right)_{\boldsymbol{x}\in X, \boldsymbol{k}\in I} \in \mathbb{C}^{|X|,|I|}.$$

Then we formulate the first question as how to compute the matrix vector product

$$\boldsymbol{p} = \boldsymbol{F}\hat{\boldsymbol{p}} \tag{2.4}$$

efficiently and the second question as how to solve this system for $\hat{\boldsymbol{p}} = (\hat{p}_{\boldsymbol{k}})_{\boldsymbol{k}\in I}$. Here, $\boldsymbol{p} = (p(\boldsymbol{x}))_{\boldsymbol{x}\in X}$ is the vector of evaluations. We consider the questions for the special case that the sampling set X is a rank-1 lattice in the following.

Definition 2.13 For a vector $\boldsymbol{z} \in \mathbb{Z}^d$ and a positive integer $M \in \mathbb{N}$, we define the **rank-1 lattice**

$$\Lambda(\boldsymbol{z}, M) = \left\{\boldsymbol{x}_j = \frac{1}{M}(j\boldsymbol{z} \mod M\boldsymbol{1}) \in [0,1)^d \colon j = 0, 1, \dots, M-1\right\}$$

with $\boldsymbol{1} = (1)_{i=1}^d$ and $j\boldsymbol{z} \mod M\boldsymbol{1} = (jz_i \mod M)_{i=1}^d$. We call $\boldsymbol{z}$ the **generating vector** and M the **lattice size**. ◇

Rank-1 lattice have the following property with regard to dimension reduction. This is particularly important to the considerations in Section 4.

Lemma 2.14 *Let $\Lambda(\boldsymbol{z}, M)$ be a rank-1 lattice with generating vector $\boldsymbol{z} \in \mathbb{Z}^d$ and lattice size $M \in \mathbb{N}$. Furthermore, let*

$$\boldsymbol{x}_j = \frac{1}{M}(j\boldsymbol{z}_{\boldsymbol{u}} \mod M\boldsymbol{1}) \in \Lambda(\boldsymbol{z}_{\boldsymbol{u}}, M) \quad \textit{and} \quad \boldsymbol{y}_j = \frac{1}{M}(j\boldsymbol{z} \mod M\boldsymbol{1}) \in \Lambda(\boldsymbol{z}, M)$$

with $\emptyset \neq \boldsymbol{u} \subset \{1, 2, \ldots, d\}$ and $\boldsymbol{z}_{\boldsymbol{u}} = (z_i)_{i\in\boldsymbol{u}}$. Then

$$\boldsymbol{x}_j = (\boldsymbol{y}_j)_{\boldsymbol{u}} = ((\boldsymbol{y}_j)_i)_{i\in\boldsymbol{u}}.$$

Proof. The statement follows from the following simple consideration

$$\boldsymbol{x}_j = \frac{1}{M}(j\boldsymbol{z}_{\boldsymbol{u}} \mod M\boldsymbol{1}) = \frac{1}{M}(j\boldsymbol{z} \mod M\boldsymbol{1})_{\boldsymbol{u}} = \left(\boldsymbol{y}_j\right)_{\boldsymbol{u}}.$$

■

If $X = \Lambda(\boldsymbol{z}, M)$ we can use the LFFT or Lattice FFT, Algorithm 2.1, to compute the matrix-vector product $\boldsymbol{F}\hat{\boldsymbol{p}}$ efficiently and the adjoint LFFT, Algorithm 2.2, to do the same for the adjoint matrix-vector product $\boldsymbol{F}^H\boldsymbol{p}$. For a detailed explanation we refer to [18, Chapter 8.2.2].

Algorithm 2.1 LFFT

Input: $M \in \mathbb{N}$ lattice size of rank-1 lattice $\Lambda(\boldsymbol{z}, M)$
$\boldsymbol{z} \in \mathbb{Z}^d$ generating vector of $\Lambda(\boldsymbol{z}, M)$
$I \subset \mathbb{Z}^d$ finite frequency index set
$\hat{\boldsymbol{p}} = (\hat{p}_{\boldsymbol{k}})_{\boldsymbol{k}\in I}$ Fourier coefficients of p

1: $\hat{\boldsymbol{g}} \leftarrow (0)_{l=0}^{M-1}$
2: **for** $\boldsymbol{k} \in I$ **do**
3: $\quad \hat{g}_{\boldsymbol{k}\cdot\boldsymbol{z} \bmod M} \leftarrow \hat{g}_{\boldsymbol{k}\cdot\boldsymbol{z} \bmod M} + \hat{p}_{\boldsymbol{k}}$
4: **end for**
5: $\boldsymbol{p} \leftarrow \boldsymbol{F}_M^{-1}\hat{\boldsymbol{g}}$ ◁ one-dim. FFT of length M
6: $\boldsymbol{p} \leftarrow M\boldsymbol{p}$

Output: $\boldsymbol{p} = \boldsymbol{F}\hat{\boldsymbol{p}}$ values of trigonometric polynomial p
Arithmetic cost: $M \log M + d\,|I|$

Now, we consider the second question from above. The system belonging to (2.4) has only an unique solution if $|X| \geq |I|$ and the Fourier matrix $\boldsymbol{F}$ has full rank. From [18, Lemma 8.7] we know

$$\left(\boldsymbol{F}^H\boldsymbol{F}\right)_{\boldsymbol{k}\in I,\boldsymbol{h}\in I} = \begin{cases} M & : \boldsymbol{k}\cdot\boldsymbol{z} = \boldsymbol{h}\cdot\boldsymbol{z} \mod M \\ 0 & : \text{otherwise}, \end{cases} \tag{2.5}$$

which leads to the following definition and subsequent theorem.

Algorithm 2.2 adjoint LFFT

Input: $M \in \mathbb{N}$ lattice size of rank-1 lattice $\Lambda(\boldsymbol{z}, M)$
$\boldsymbol{z} \in \mathbb{Z}^d$ generating vector of $\Lambda(\boldsymbol{z}, M)$
$I \subset \mathbb{Z}^d$ finite frequency index set
$\boldsymbol{p} = (p(\frac{j}{M}\boldsymbol{z}))_{j=0}^{M-1}$ values of trigonometric polynomial p

1: $\hat{\boldsymbol{g}} \leftarrow \boldsymbol{F}_M \boldsymbol{p}$ ◁ one-dim. FFT of length M
2: **for** $\boldsymbol{k} \in I$ **do**
3: $\hat{a}_{\boldsymbol{k}} \leftarrow \hat{g}_{\boldsymbol{k}\cdot\boldsymbol{z} \mod M}$
4: **end for**

Output: $\hat{\boldsymbol{a}} = \boldsymbol{F}^H \boldsymbol{p}$
Arithmetic cost: $M \log M + d\,|I|$

Definition 2.15 We call a rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ **reconstructing rank-1 lattice** for the finite index set $I \subset \mathbb{Z}^d$ and denote it with $\Lambda(\boldsymbol{z}, M, I)$ if

$$\boldsymbol{k} \cdot \boldsymbol{z} \not\equiv \boldsymbol{h} \cdot \boldsymbol{z} \mod M \, \forall \boldsymbol{h} \neq \boldsymbol{k}$$

with $\boldsymbol{h}, \boldsymbol{k} \in I$. ◇

Theorem 2.16 *Let $\Lambda(\boldsymbol{z}, M, I)$ be a reconstructing rank-1 lattice with respect to the finite index set $I \subset \mathbb{Z}^d$. Then*

$$\boldsymbol{F}^H \boldsymbol{F} = M\,\boldsymbol{I}$$

for the Fourier matrix $\boldsymbol{F} = (\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}})_{\boldsymbol{x}\in\Lambda(\boldsymbol{z},M,I),\boldsymbol{k}\in I}$.

Proof. see [18, Lemma 8.7] ■

The previous considerations allow us to write the solution to the second question, i.e., solving (2.4) for $\hat{\boldsymbol{p}}$, as

$$\hat{\boldsymbol{p}} = (\boldsymbol{F}^H \boldsymbol{F})^{-1} \boldsymbol{F}^H \boldsymbol{p} = (M\mathbf{I})^{-1} \boldsymbol{F}^H \boldsymbol{p} = \frac{1}{M} \boldsymbol{F}^H \boldsymbol{p}$$

which can be computed efficiently using an adjoint LFFT. When approximating a function $f \in \mathrm{L}_1(\mathbb{T}^d)$ by a Fourier partial sum $(S_I f)(\boldsymbol{x})$ as described in [18, Chapter 8.3], we have the **aliasing formula**

$$\tilde{\hat{f}}_{\boldsymbol{k}} = \hat{f}_{\boldsymbol{k}} + \sum_{\boldsymbol{h}\in\Lambda^{\perp}(\boldsymbol{z},M)\setminus\{\boldsymbol{0}\}} \hat{f}_{\boldsymbol{k}+\boldsymbol{h}}. \tag{2.6}$$

Here, $\tilde{\hat{f}}_{\boldsymbol{k}}$ are the reconstructed Fourier coefficients and

$$\Lambda^{\perp}(\boldsymbol{z}, M) \coloneqq \{\boldsymbol{k} \in \mathbb{Z}^d \colon \boldsymbol{k} \cdot \boldsymbol{z} \equiv 0 \mod M\}$$

is the integer dual lattice.

3 The ANOVA Decomposition

The **analysis of variance (ANOVA)** decomposition [13, 3, 4, 17] is an important model in the analysis of dimension interactions of multivariate functions. Besides being a tool in understanding certain quadrature algorithms such as the quasi-Monte Carlo method [16, 20] or the Multivariate Decomposition Method to approximate infinite-variate integrals [12, 2], the ANOVA model was recently used as a basis for learning highdimensional sparse additive models from point queries [21]. Furthermore, Analysis of Variance is well-known in statistics as a tool to study dimension relations.

We start by introducing the classical ANOVA decomposition in Section 3.1 where we also translate the established terms to a Fourier context. We then go on to the properties of the model in Section 3.2 where we focus on sensitivity analysis to determine dimension interactions. Those considerations lead to an approximate ANOVA model in Section 3.3 which will be especially relevant for approximation.

3.1 The Classical ANOVA Decomposition

We consider multivariate periodic functions

$$f : \mathbb{T}^d \to \mathbb{R}, \boldsymbol{x} \mapsto f(\boldsymbol{x}) \tag{3.1}$$

with dimension $d \in \mathbb{N}$ and the smooth manifold of the Torus $\mathbb{T}$ which we identify with $(-1/2, 1/2]$. Here, the right and left interval limits are identified with each other. We want to restrict ourselves to the square-integrable functions in $f \in \mathrm{L}_2(\mathbb{T}^d)$ which can be represented as Fourier series

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}, \tag{3.2}$$

see Section 2.

Now, let $\mathcal{D} = \{1, 2, \ldots, d\}$ denote the set of coordinate indices of $\boldsymbol{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{T}^d$. Note that we will often say that variables are in a subset $\boldsymbol{u} \subset \mathcal{D}$ although technical this refers to the coordinate indices of the variables. The first important tool for the ANOVA model are projections on subspaces. We use the notation $\boldsymbol{x}_{\boldsymbol{u}} = (x_i)_{i \in \boldsymbol{u}}$ for $\boldsymbol{u} \subset \mathcal{D}$.

Definition 3.1 Let $\boldsymbol{u} \subset \mathcal{D}$ and $f \in \mathrm{L}_2(\mathbb{T}^d)$ as in (3.1). We define a **projection operator**

$$P_{\boldsymbol{u}} : \mathrm{L}_2(\mathbb{T}^d) \to \mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|}), f \mapsto \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}_{\mathcal{D} \setminus \boldsymbol{u}}. \tag{3.3}$$

$\diamond$

Note that $P_{\boldsymbol{u}} f$ only depends on the variables $\boldsymbol{x}_{\boldsymbol{u}}$, i.e., $(P_{\boldsymbol{u}} f)(\boldsymbol{x}) = (P_{\boldsymbol{u}} f)(\boldsymbol{x}_{\boldsymbol{u}})$. For the special case $\boldsymbol{u} = \emptyset$ we have

$$P_{\emptyset} = \int_{\mathbb{T}^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} =: If$$

which does not depend on any variable. Clearly, $P_{\boldsymbol{u}}$ is well-defined and maps into $\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})$ since

$$\int_{\mathbb{T}^{|\boldsymbol{u}|}} \left| \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \right|^2 \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \le \int_{\mathbb{T}^{|\boldsymbol{u}|}} \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} |f(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} = \|f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} \,. \tag{3.4}$$

In order to translate the projections to a Fourier context, we need the index set

$$\mathbb{P}^{(d)}_{\boldsymbol{u}} := \{\boldsymbol{k} \in \mathbb{Z}^d : \boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}\} \tag{3.5}$$

which is isomorphic to $\mathbb{Z}^{|\boldsymbol{u}|}$ with the identification $\boldsymbol{k} \leftrightarrow \boldsymbol{k}_{\boldsymbol{u}}$ since always $\boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}$. We use the convention $\mathbb{Z}^0 = \{0\}$. Now, we prove some properties of the projections $P_{\boldsymbol{u}}$. Note that we understand $\boldsymbol{k}_{\boldsymbol{u}} = (k_i)_{i\in\boldsymbol{u}}$ in the same manner as $\boldsymbol{x}_{\boldsymbol{u}}$.

Lemma 3.2 *Let $P_{\boldsymbol{u}}$ be the projection operator* (3.3) *and $\hat{f}_{\boldsymbol{k}}, \boldsymbol{k} \in \mathbb{Z}^d$, the Fourier coefficients of f in* (3.2). *$P_{\boldsymbol{u}}$ has the following properties:*

(i) *$P_{\boldsymbol{u}}$ is idempotent, i.e., $P^2_{\boldsymbol{u}} = P_{\boldsymbol{u}}$,*

(ii) *if $\boldsymbol{u} = \mathcal{D}$ then $P_{\boldsymbol{u}} = \mathrm{id}_{\mathrm{L}_2(\mathbb{T}^d)}$,*

(iii) $(P_{\boldsymbol{u}} f)(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{P}^{(d)}_{\boldsymbol{u}}} \hat{f}_{\boldsymbol{k}} \,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}$ *with $\mathbb{P}^{(d)}_{\boldsymbol{u}}$ from* (3.5),

(iv) $(P_{\boldsymbol{u}} f)(\boldsymbol{x}) = \sum_{\boldsymbol{\ell}\in\mathbb{Z}^{|\boldsymbol{u}|}} \hat{p}_{\boldsymbol{\ell},\boldsymbol{u}} \,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{x}_{\boldsymbol{u}}}$ *with $\hat{p}_{\boldsymbol{\ell},\boldsymbol{u}} = \hat{f}_{\boldsymbol{k}}$ for $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}$.*

Proof. We prove property (i) by

$$P_{\boldsymbol{u}}(P_{\boldsymbol{u}} f) = \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} = P_{\boldsymbol{u}} f \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} \mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} = P_{\boldsymbol{u}} f$$

since $P_{\boldsymbol{u}} f$ does not depend on variables in $\mathcal{D} \setminus \boldsymbol{u}$. One can also directly see property (ii) since we are not integrating over any variables in this case which just leaves us with the function f itself.

In order to prove property (iii) we start with the simple case of $P_{\emptyset} f = If$ and get

$$If = \int_{\mathbb{T}^d} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{T}^d} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \,\mathrm{d}\boldsymbol{x} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \int_{\mathbb{T}^d} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \,\mathrm{d}\boldsymbol{x} = \hat{f}_{\boldsymbol{0}}.$$

by Fubini's Theorem. Since $\mathbb{P}^{(d)}_{\emptyset} = \{\boldsymbol{0}\}$, property (iii) holds for $\boldsymbol{u} = \emptyset$. For the case $\boldsymbol{u} = \mathcal{D}$, we have $P_{\boldsymbol{u}} = \mathrm{id}_{\mathrm{L}_2(\mathbb{T}^d)}$ and $\mathbb{P}^{(d)}_{\boldsymbol{u}} = \mathbb{Z}^d$ and therefore arrive again at the Fourier series representation (3.2). For a general $\boldsymbol{u} \subset \mathcal{D}$, we can again interchange sum and integral

$$\begin{aligned} P_{\boldsymbol{u}} f(\boldsymbol{x}) &= \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \\ &= \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}_{\boldsymbol{u}}\cdot\boldsymbol{x}_{\boldsymbol{u}}} \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}}\cdot\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}}} \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \delta_{\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}}, \boldsymbol{0}} \\
&= \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} \hat{f}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}.
\end{aligned}
$$

Here, $\delta_{\boldsymbol{a},\boldsymbol{b}}$, the Kronecker delta, is one if $\boldsymbol{a} = \boldsymbol{b}$ and zero otherwise. So, property (iii) holds.

For property (iv), we just need to identify $\mathbb{P}_{\boldsymbol{u}}^{(d)}$ with $\mathbb{Z}^{|\boldsymbol{u}|}$ as done above. We omit the zeros from $\boldsymbol{k}$, i.e., take $\boldsymbol{k}_{\boldsymbol{u}}$, and do the same for the entries of $\boldsymbol{x}$ that are not represented in $\boldsymbol{u}$, i.e., use $\boldsymbol{x}_{\boldsymbol{u}}$. ■

We especially emphasize that the difference of (iii) and (iv) lies only in the dimension of the indices within the different index sets. We will use both variants to represent $P_{\boldsymbol{u}} f$ based on the context.

Given the projections $P_{\boldsymbol{u}} f$, we define the **ANOVA term**

$$
f_{\boldsymbol{u}}(\boldsymbol{x}) \coloneqq (P_{\boldsymbol{u}} f)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} f_{\boldsymbol{v}}(\boldsymbol{x}). \tag{3.6}
$$

Similar to the projections, the ANOVA term $f_{\boldsymbol{u}}$ only depends on the variables $\boldsymbol{x}_{\boldsymbol{u}}$, i.e., $f_{\boldsymbol{u}}(\boldsymbol{x}) = f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$.

Lemma 3.3 *Let $f \in \mathrm{L}_2(\mathbb{T}^d)$. Then*

$$
f_{\boldsymbol{u}} \in \mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})
$$

for each $\boldsymbol{u} \subset \mathcal{D}$.

Proof. The projection $P_{\boldsymbol{u}} f$ as well as the ANOVA terms of lower order $f_{\boldsymbol{v}}$ are (by adding independent variables) in $\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|}) = \mathrm{Im}(\mathrm{P}_{\boldsymbol{u}})$. Therefore, $f_{\boldsymbol{u}}$ is a sum of $\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})$ functions and as a consequence in $\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})$ itself. ■

Now, we want to find a Fourier series representation for the ANOVA terms $f_{\boldsymbol{u}}$. As for the projections, we need a special index set

$$
\mathbb{F}_{\boldsymbol{u}}^{(d)} \coloneqq \{\boldsymbol{k} \in \mathbb{Z}^d : \boldsymbol{k}_{\boldsymbol{u}} \neq \boldsymbol{0}, \boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}} = \boldsymbol{0}\}. \tag{3.7}
$$

This can be identified with $(\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}$ through $\boldsymbol{k} \leftrightarrow \boldsymbol{\ell}$ with $\boldsymbol{\ell} = \boldsymbol{k}_{\boldsymbol{u}}$ and $\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}} = \boldsymbol{0}$. We use the convention that $(\mathbb{Z} \setminus \{0\})^0 = \{0\}$. In order to prove the alternate representation, we need the following two lemmas.

Lemma 3.4 *Let $d \in \mathbb{N}$ and $\boldsymbol{u} \subset \mathcal{D}$. Then*

$$
\bigcup_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \mathbb{F}_{\boldsymbol{v}}^{(d)} = \{\boldsymbol{k} \in \mathbb{Z}^d : \exists i \in \boldsymbol{u} : k_i = 0, \boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}} = \boldsymbol{0}\}
$$

and as a direct consequence

$$
\mathbb{F}_{\boldsymbol{u}}^{(d)} = \mathbb{P}_{\boldsymbol{u}}^{(d)} \setminus \bigcup_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \mathbb{F}_{\boldsymbol{v}}^{(d)}.
$$

Proof. '$\subset$': If $\boldsymbol{k}$ is in the set on the left-hand side then we can find $\boldsymbol{v} \subsetneq \boldsymbol{u}$ such that $\boldsymbol{k}_{\boldsymbol{v}} \neq \boldsymbol{0}$ and $\boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{v}} = \boldsymbol{0}$. If we choose $\boldsymbol{w} = \boldsymbol{u} \setminus \boldsymbol{v}$ then $\boldsymbol{k}_{\boldsymbol{w}} = \boldsymbol{0}$ and $\boldsymbol{k}$ is in the set on the right-hand side.

'$\supset$': If $\boldsymbol{k}$ is in the set on the right-hand side then we can find $\boldsymbol{w} = \{i\} \subset \boldsymbol{u}$ such that $\boldsymbol{k}_{\boldsymbol{w}} = \boldsymbol{0}$. We can now either find a maximal set $\boldsymbol{v} \subset \boldsymbol{u}$ such that $\boldsymbol{k}_{\boldsymbol{v}} \neq \boldsymbol{0}$ or $\boldsymbol{k} = \boldsymbol{0}$. In both cases, $\boldsymbol{k}$ is in the set on the left-hand side. ■

Lemma 3.5 *Let $d \in \mathbb{N}$ and $\boldsymbol{u}, \boldsymbol{v} \subset \mathcal{D}$. Then*

$$\mathbb{F}_{\boldsymbol{u}}^{(d)} \cap \mathbb{F}_{\boldsymbol{v}}^{(d)} = \emptyset$$

for $\boldsymbol{u} \neq \boldsymbol{v}$ and

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

is a disjoint union.

Proof. We prove the first statement by contradiction. 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 $\tilde{\boldsymbol{k}}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}$ 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 $j \in (\mathcal{D} \setminus \boldsymbol{v}) \cap \boldsymbol{u}$. Then $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{v}}^{(d)}$ implies that $\tilde{k}_j = 0$ which contradicts $\tilde{\boldsymbol{k}} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}$.

For the second statement we consider the two inclusions. The inclusion $\bigcup_{\boldsymbol{u} \subset \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)} \subset \mathbb{Z}^d$ is clear since $\mathbb{F}_{\boldsymbol{u}}^{(d)} \subset \mathbb{Z}^d \, \forall \boldsymbol{u} \subset \mathcal{D}$. For the other inclusion we take a $\boldsymbol{k} \in \mathbb{Z}^d$. Then we can find the maximal set $\boldsymbol{v} \subset \mathcal{D}$ such that $\boldsymbol{k}_{\boldsymbol{v}} \neq \boldsymbol{0}$. Note that $\boldsymbol{v} = \emptyset$ is possible. In this case it holds that $\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{v}}^{(d)}$. ■

Corollary 3.6 *Let $\boldsymbol{u} \subset \mathcal{D}$ and $f \in \mathrm{L}_2(\mathbb{T}^d)$. The ANOVA term $f_{\boldsymbol{u}}$ has the Fourier series representations*

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} \hat{f}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} \tag{3.8}$$

and

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \sum_{\boldsymbol{\ell} \in (\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} \hat{f}_{\boldsymbol{\ell},\boldsymbol{u}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}}$$

with $\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}} = \hat{f}_{\boldsymbol{k}}$ for $\boldsymbol{\ell} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}$ with $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}$.

Proof. We use structural induction over the cardinality of $|\boldsymbol{u}|$. The statement (3.8) is trivial for $\boldsymbol{u} = \emptyset$ since

$$f_{\emptyset} = If = \hat{f}_{\boldsymbol{0}}$$

and $\mathbb{F}_{\emptyset}^{(d)} = \{\boldsymbol{0}\}$. Now, let (3.8) be true for $|\boldsymbol{v}| = 0, 1, \ldots, n-1$ and choose $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| = n$. Using the recursive formula (3.6) for the ANOVA terms and the Fourier representation of $P_{\boldsymbol{u}} f$ from Lemma 3.2, we have

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} \hat{f}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{v}}^{(d)}} \hat{f}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}$$

$$\overset{\text{Lemma 3.5}}{=} \sum_{\boldsymbol{k}\in\mathbb{P}_{\boldsymbol{u}}^{(d)}\setminus\bigcup_{\boldsymbol{v}\subsetneq\boldsymbol{u}}\mathbb{F}_{\boldsymbol{v}}^{(d)}} \hat{f}_{\boldsymbol{k}}\,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}$$
$$\overset{\text{Lemma 3.4}}{=} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \hat{f}_{\boldsymbol{k}}\,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}.$$

■

We can use the Fourier series representation (3.8) in order to prove two important properties of the ANOVA terms.

Lemma 3.7 *Let $\boldsymbol{u},\boldsymbol{v}\subset D$ with $\boldsymbol{u}\neq\boldsymbol{v}$. Then*

$$\langle f_{\boldsymbol{u}}, f_{\boldsymbol{v}}\rangle_{\mathrm{L}_2(\mathbb{T}^d)} = 0.$$

Proof. We take the scalar product and replace the ANOVA terms by their Fourier series (3.8)

$$\begin{aligned}
\langle f_{\boldsymbol{u}}, f_{\boldsymbol{v}}\rangle &= \Big\langle \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \hat{f}_{\boldsymbol{k}}\,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \sum_{\boldsymbol{\ell}\in\mathbb{F}_{\boldsymbol{v}}^{(d)}} \hat{f}_{\boldsymbol{\ell}}\,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{x}}\Big\rangle \\
&= \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \sum_{\boldsymbol{\ell}\in\mathbb{F}_{\boldsymbol{v}}^{(d)}} \hat{f}_{\boldsymbol{k}}\hat{f}_{\boldsymbol{\ell}}\,\underbrace{\langle \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{x}}\rangle}_{=\delta_{\boldsymbol{k},\boldsymbol{\ell}}} \\
&= \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \sum_{\boldsymbol{\ell}\in\mathbb{F}_{\boldsymbol{v}}^{(d)}} \hat{f}_{\boldsymbol{k}}\hat{f}_{\boldsymbol{\ell}}\delta_{\boldsymbol{k},\boldsymbol{\ell}} \\
&= 0
\end{aligned}$$

The last step follows by Lemma 3.5, i.e., the fact that the two index sets $\mathbb{F}_{\boldsymbol{u}}^{(d)}$ and $\mathbb{F}_{\boldsymbol{v}}^{(d)}$ are disjoint. ■

Lemma 3.8 *Let $\boldsymbol{u}\subset D$ with $\boldsymbol{u}\neq\emptyset$. Then*

$$I f_{\boldsymbol{u}} = \int_{\mathbb{T}^{|\boldsymbol{u}|}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})\,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} = 0.$$

Proof. This follows directly from Corollary 3.6 since $\mathbf{0}\notin\mathbb{F}_{\boldsymbol{u}}^{(d)}$. ■

We now use the ANOVA terms to decompose a function f in a unique way through the ANOVA decomposition.

Theorem 3.9 (ANOVA decomposition) *Let $f\in\mathrm{L}_2(\mathbb{T}^d)$ and $f_{\boldsymbol{u}}$ as in* (3.6)*. Then we can decompose f uniquely as*

$$f(\boldsymbol{x}) = f_\emptyset + \sum_{i=1}^{d} f_{\{i\}}(\boldsymbol{x}) + \sum_{\substack{i,j=1\\ i<j}}^{d} f_{\{i,j\}}(\boldsymbol{x}) + \cdots + f_{\{1,2,\ldots,d\}}(\boldsymbol{x})$$

and call this ***analysis of variance (ANOVA) decomposition****. We will also use the more compact notation*

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{u} \subset \mathcal{D}} f_{\boldsymbol{u}}(\boldsymbol{x}).$$

with at most 2^d terms.

Proof. First, we write the ANOVA decomposition with Fourier series of the ANOVA terms

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

Since we know that $\bigcup_{\boldsymbol{u} \subset \mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)} = \mathbb{Z}^d$ is a disjoint union by Lemma 3.5, we rewrite this as

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

and arrive at our function f.

This decomposition is unique because we split the index set $\mathbb{Z}^d$ in disjoint subsets. Since we sum over all elements within the power set, we have at most $|\mathcal{P}(\mathcal{D})| = 2^d$ terms as some of them might be 0. ■

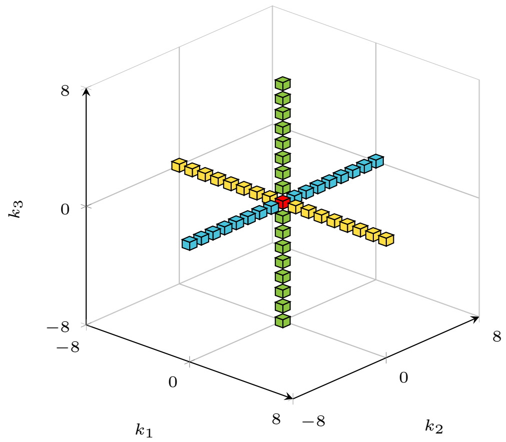
Remark 3.10 *The ANOVA decomposition as proposed in Theorem 3.9 depends strongly on the projection operator $P_{\boldsymbol{u}}$ from Definition 3.3. The ANOVA decomposition with this projection operator is sometimes referred to as the* ***classical ANOVA decomposition****. There are certain conditions as stated in [13] to a general projection operator under which we still have a generalization of the ANOVA decomposition or general decomposition formula. One special example is the* ***anchored ANOVA decomposition*** *with the projection operator*

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

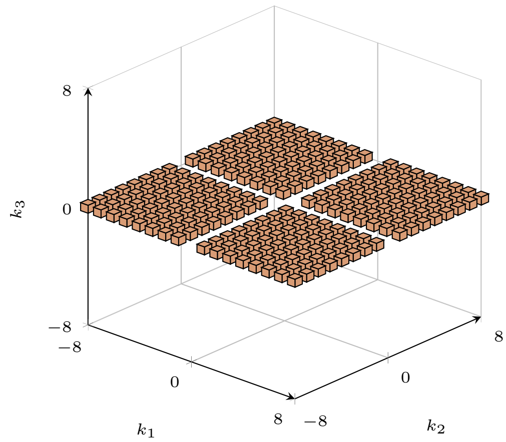
with an anchor point $\boldsymbol{c} \in \mathbb{T}^d$. Here, $(\boldsymbol{x}, \boldsymbol{c})_{\boldsymbol{u}} \in \mathbb{T}^d$ such that

$$((\boldsymbol{x}, \boldsymbol{c})_{\boldsymbol{u}})_i = \begin{cases} x_i & : i \in \boldsymbol{u} \\ c_i & : i \notin \boldsymbol{u} \end{cases}.$$

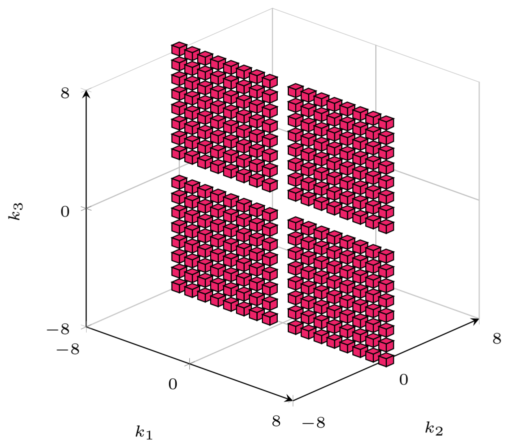
Figure 3.1 shows how the ANOVA decomposition is working on the subset $[-8, 8]^d$ of the index set $\mathbb{Z}^d$ for $d = 3$. It breaks the cube into the disjoint subsets $\mathbb{F}_{\boldsymbol{u}}^{(3)}$ for $\boldsymbol{u} \subset \{1, 2, 3\}$. Figure 3.2 shows the projection index sets $\mathbb{P}_{\boldsymbol{u}}$ in $[-8, 8]^d$. We can clearly see that the sets are not disjoint.



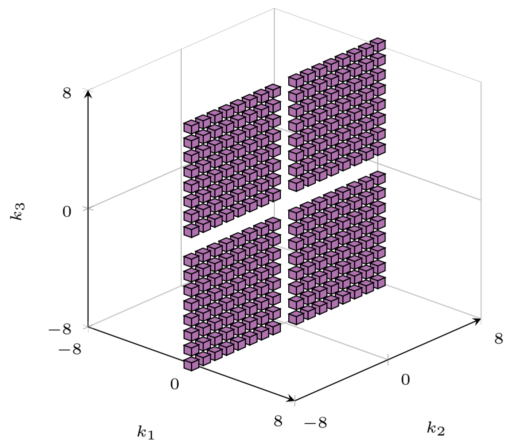
(a) $\mathbb{F}_{\emptyset}^{(3)}$ in red, $\mathbb{F}_{\{1\}}^{(3)} \cap [-8, 8]^3$ in yellow, $\mathbb{F}_{\{2\}}^{(3)} \cap [-8, 8]^3$ in blue, $\mathbb{F}_{\{3\}}^{(3)} \cap [-8, 8]^3$ in green



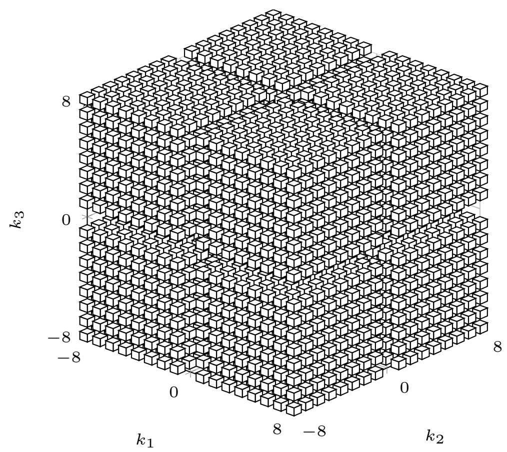
(b) $\mathbb{F}_{\{1,2\}}^{(3)} \cap [-8, 8]^3$



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

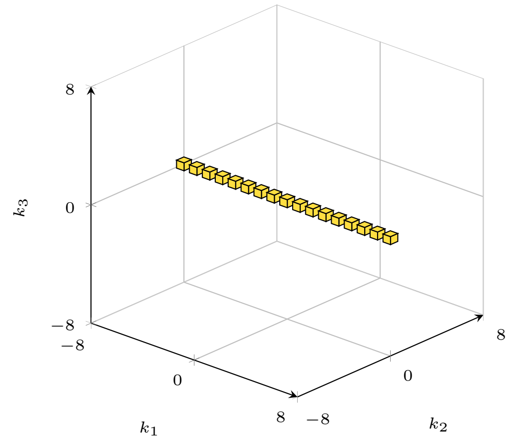


(d) $\mathbb{F}_{\{2,3\}}^{(3)} \cap [-8, 8]^3$

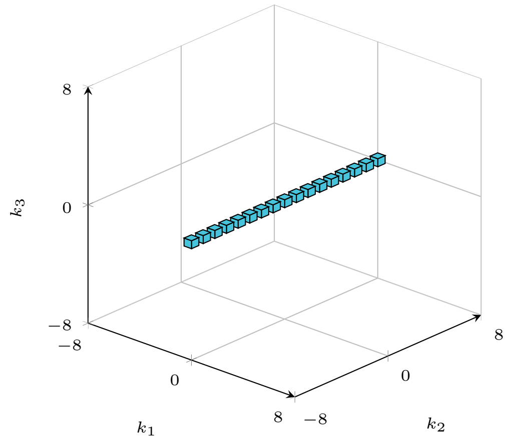


(e) $\mathbb{F}_{\{1,2,3\}}^{(3)} \cap [-8, 8]^3$

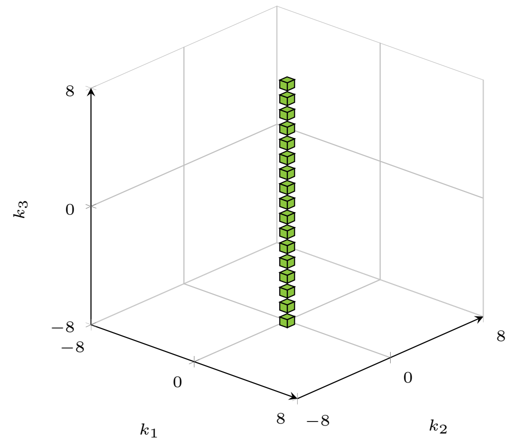
Figure 3.1: The ANOVA decomposition working on the hypercube $[-8, 8]^3$ as a part of the 3-dimensional index set $\mathbb{Z}^3$.



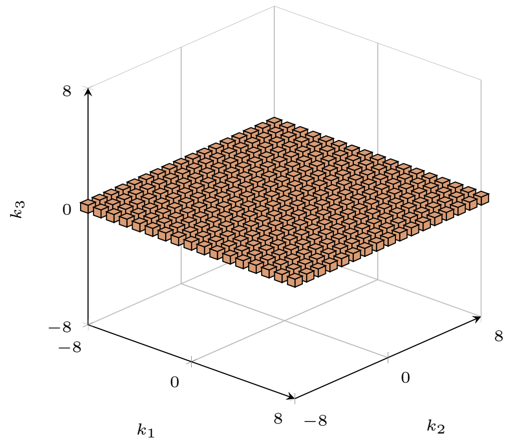
(a) $\mathbb{P}^{(3)}_{\{1\}} \cap [-8,8]^3$



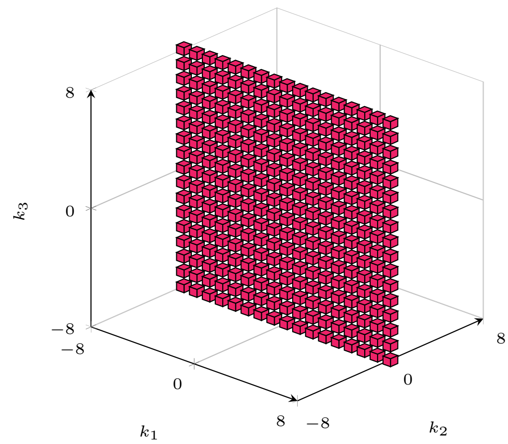
(b) $\mathbb{P}^{(3)}_{\{2\}} \cap [-8,8]^3$



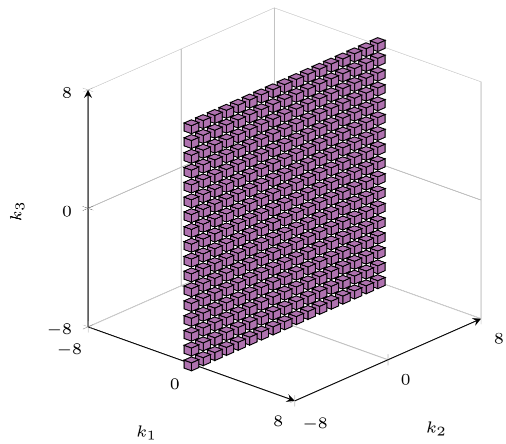
(c) $\mathbb{P}^{(3)}_{\{3\}} \cap [-8,8]^3$



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



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



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

Figure 3.2: The projection index sets $\mathbb{P}^{(d)}_{\boldsymbol{u}}$ with $1 \leq |\boldsymbol{u}| \leq 2$ in the hypercube $[-8,8]^3$ as a part of the 3-dimensional index set $\mathbb{Z}^3$.

Furthermore, it is possible to find a direct representation for the ANOVA terms using only projections which will be useful later on. A proof of this alternate expression in Theorem 3.12 is given in [13] using a property of projection operators. Here, we give a proof through counting arguments using the following lemma.

Lemma 3.11 *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

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

■

Theorem 3.12 *Let $\boldsymbol{u} \subset \mathcal{D}$. Then*

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \sum_{\boldsymbol{v} \subset \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \left(P_{\boldsymbol{v}} f\right)(\boldsymbol{x}) \tag{3.9}$$

is an equivalent expression for the ANOVA terms (3.6)*.*

Proof. 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 (3.9) be true for $\boldsymbol{v} \subset \mathcal{D}$, $|\boldsymbol{v}| = 0, 1, \dots, n-1$ and take a subset $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| = n$. We start from the recursive expression in (3.6)

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \left(P_{\boldsymbol{u}} f\right)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} f_{\boldsymbol{v}}(\boldsymbol{x})$$

$$
\begin{aligned}
&= (P_{\boldsymbol{u}} f)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \sum_{\boldsymbol{w} \subset \boldsymbol{v}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} (P_{\boldsymbol{w}} f)(\boldsymbol{x}) \\
&= (P_{\boldsymbol{u}} f)(\boldsymbol{x}) - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} (P_{\boldsymbol{w}} f)(\boldsymbol{x}) \delta_{\boldsymbol{w} \subset \boldsymbol{v}} \\
&= (P_{\boldsymbol{u}} f)(\boldsymbol{x}) - \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (P_{\boldsymbol{w}} f)(\boldsymbol{x}) \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \delta_{\boldsymbol{w} \subset \boldsymbol{v}} \\
&= (P_{\boldsymbol{u}} f)(\boldsymbol{x}) - \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (P_{\boldsymbol{w}} f)(\boldsymbol{x}) \sum_{n=|\boldsymbol{w}|}^{|\boldsymbol{u}|-1} \sum_{\substack{\boldsymbol{v} \subset \boldsymbol{u} \\ |\boldsymbol{v}|=n}} (-1)^{|\boldsymbol{v}|-|\boldsymbol{w}|} \delta_{\boldsymbol{w} \subset \boldsymbol{v}} \\
&= (P_{\boldsymbol{u}} f)(\boldsymbol{x}) - \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (P_{\boldsymbol{w}} f)(\boldsymbol{x}) \sum_{n=|\boldsymbol{w}|}^{|\boldsymbol{u}|-1} (-1)^{n-|\boldsymbol{w}|} \sum_{\substack{\boldsymbol{v} \subset \boldsymbol{u} \\ |\boldsymbol{v}|=n}} \delta_{\boldsymbol{w} \subset \boldsymbol{v}} \\
&= (P_{\boldsymbol{u}} f)(\boldsymbol{x}) + \sum_{\boldsymbol{w} \subsetneq \boldsymbol{u}} (P_{\boldsymbol{w}} f)(\boldsymbol{x}) \underbrace{\sum_{n=|\boldsymbol{w}|}^{|\boldsymbol{u}|-1} (-1)^{n-|\boldsymbol{w}|+1} \binom{|\boldsymbol{u}|-|\boldsymbol{w}|}{n-|\boldsymbol{w}|}}_{\overset{\text{Lemma 3.11}}{=} (-1)^{|\boldsymbol{u}|-|\boldsymbol{w}|}} \\
&= \sum_{\boldsymbol{w} \subset \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{w}|} (P_{\boldsymbol{w}} f)(\boldsymbol{x})
\end{aligned}
$$

with

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

■

To conclude this section, we discuss some examples for ANOVA decompositions of functions.

Example 3.13 We consider the function

$$
f(x_1, x_2) = \sin\left(\frac{x_1 \pi}{3}\right) + \cos\left(\frac{x_2 \pi}{3}\right) + \sin\left(\frac{x_1 \pi}{3}\right) \cos\left(\frac{x_2 \pi}{3}\right)
$$

which is clearly an element of $\mathrm{L}_2(\mathbb{T}^2)$. First, we calculate the projections

$$
\begin{aligned}
P_{\emptyset} f(x_1, x_2) &= \int_{\mathbb{T}^2} f(x_1, x_2) \, \mathrm{d}(x_1, x_2) = \frac{3}{\pi} \\
P_{\{1\}} f(x_1, x_2) &= \int_{\mathbb{T}} f(x_1, x_2) \, \mathrm{d}x_2 = \sin\left(\frac{x_1 \pi}{3}\right) \left[1 + \frac{3}{\pi}\right] + \frac{3}{\pi} \\
P_{\{2\}} f(x_1, x_2) &= \int_{\mathbb{T}} f(x_1, x_2) \, \mathrm{d}x_1 = \cos\left(\frac{x_2 \pi}{3}\right) \\
P_{\{1,2\}} f(x_1, x_2) &= f(x_1, x_2).
\end{aligned}
$$

Now, we use the direct formula (3.9) in order to compute the ANOVA terms

$$
\begin{aligned}
f_\emptyset(x_1,x_2) &= (-1)^{0-0} P_\emptyset f(x_1,x_2) = \frac{3}{\pi} \\
f_{\{1\}}(x_1,x_2) &= (-1)^{1-0} P_\emptyset f(x_1,x_2) + (-1)^{1-1} P_{\{1\}} f(x_1,x_2) \\
&= \sin\left(\frac{x_1\pi}{3}\right)\left[1+\frac{3}{\pi}\right] \\
f_{\{2\}}(x_1,x_2) &= (-1)^{1-0} P_\emptyset f(x_1,x_2) + (-1)^{1-1} P_{\{2\}} f(x_1,x_2) \\
&= \cos\left(\frac{x_2\pi}{3}\right) - \frac{3}{\pi} \\
f_{\{1,2\}}(x_1,x_2) &= (-1)^{2-0} P_\emptyset f(x_1,x_2) + (-1)^{2-1} P_{\{1\}} f(x_1,x_2) \\
&\quad + (-1)^{2-1} P_{\{2\}} f(x_1,x_2) + (-1)^{2-2} P_{\{1,2\}} f(x_1,x_2) \\
&= \sin\left(\frac{x_1\pi}{3}\right)\cos\left(\frac{x_2\pi}{3}\right) - \frac{3}{\pi}\sin\left(\frac{x_1\pi}{3}\right).
\end{aligned}
$$

We can see that the ANOVA terms add up to f, i.e., $f = \sum_{\boldsymbol{u}\subset\mathcal{D}} f_{\boldsymbol{u}}$. It is also important to observe that the ANOVA term $f_{\boldsymbol{u}}$ does not necessarily coincide with the summand in f that depends on $\boldsymbol{x}_{\boldsymbol{u}}$ if f is already given in an additive form. For our particular f, we have

$$
f_{\{1\}}(x_1,x_2) \neq \sin\left(\frac{x_1\pi}{3}\right) \quad \text{and} \quad f_{\{2\}}(x_1,x_2) \neq \cos\left(\frac{x_2\pi}{3}\right).
$$

◇

Example 3.14 As a second example, we consider

$$
f(x_1,x_2,x_3) = x_1 + x_2 + x_3 + x_1x_2x_3 \in \mathrm{L}_2(\mathbb{T}^2).
$$

Calculating the projections yields

$$
\begin{aligned}
P_\emptyset f(x_1,x_2,x_3) &= \int_{\mathbb{T}^3} f(x_1,x_2,x_3)\,\mathrm{d}(x_1,x_2,x_3) = 0, \\
P_{\{i\}} f(x_1,x_2,x_3) &= \int_{\mathbb{T}^2} f(x_1,x_2,x_3)\,\mathrm{d}\boldsymbol{x}_{\{1,2,3\}\setminus\{i\}} = x_i, i = 1,2,3, \\
P_{\{i,j\}} f(x_1,x_2,x_3) &= \int_{\mathbb{T}} f(x_1,x_2,x_3)\,\mathrm{d}\boldsymbol{x}_{\{1,2,3\}\setminus\{i,j\}} = x_i + x_j, i = 2,3, j = 1,2, j < i, \\
P_{\{1,2,3\}} f(x_1,x_2,x_3) &= f(x_1,x_2,x_3).
\end{aligned}
$$

Applying the direct formula (3.9) again, we obtain the ANOVA terms

$$
\begin{aligned}
f_\emptyset(x_1,x_2,x_3) &= (-1)^{0-0} P_\emptyset f(x_1,x_2,x_3) = 0 \\
f_{\{i\}}(x_1,x_2,x_3) &= (-1)^{1-1} P_{\{i\}} f(x_1,x_2,x_3) = x_i, i = 1,2,3, \\
f_{\{i,j\}}(x_1,x_2,x_3) &= (-1)^{2-0} P_\emptyset f(x_1,x_2,x_3) - \sum_{i=1}^{3} P_{\{i\}} f(x_1,x_2,x_3) = 0
\end{aligned}
$$

$$
\begin{aligned}
f_{\{1,2,3\}}(x_1,x_2,x_3) &= f(x_1,x_2,x_3) - \sum_{\substack{i=2,3\\ j=1,2\\ j<i}} P_{\{i,j\}} f(x_1,x_2,x_3) + \sum_{i=1}^{3} P_{\{i\}} f(x_1,x_2,x_3) \\
&= x_1 x_2 x_3
\end{aligned}
$$

We observe that contrary to Example 3.13, the summands in f coincide with the ANOVA terms, i.e., f was already given in ANOVA decomposition. ◇

Example 3.15 In order to visualize the effect of the ANOVA decomposition, we take

$$f(x_1,x_2) = \mathrm{e}^{x_1+x_2}.$$

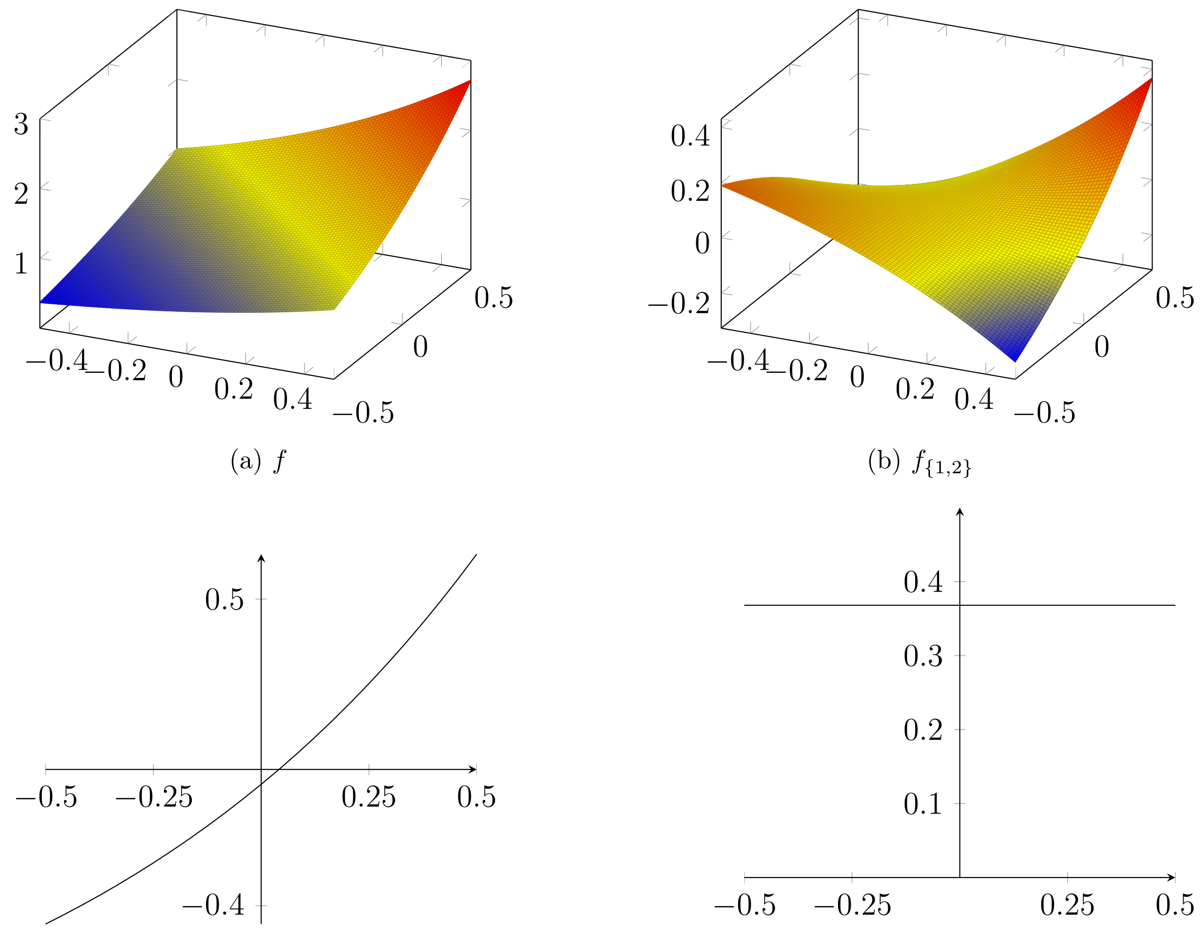
If we define $c := \int_{\mathbb{T}} e^x \,\mathrm{d}x = \mathrm{e}^{1/2} - \mathrm{e}^{-1/2}$ then

$$P_\emptyset f = c^2, P_{\{i\}} f(x_i) = c\mathrm{e}^{x_i} \text{ and } P_{\{1,2\}} f(x_1,x_2) = f(x_1,x_2)$$

and therefore the ANOVA terms are

$$f_\emptyset = c^2, f_{\{i\}}(x_i) = c\mathrm{e}^{x_i} - c^2 \text{ and } f_{\{1,2\}}(x_1,x_2) = \mathrm{e}^{x_1+x_2} - c(\mathrm{e}^{x_1} + \mathrm{e}^{x_2}) + c^2$$

with $i = 1, 2$. Figure 3.3 shows our function f with all its ANOVA terms.



(a) f

(b) $f_{\{1,2\}}$

(c) $f_{\{1\}}$ and $f_{\{2\}}$

(d) $f_{\emptyset}$

Figure 3.3: The function $f(x_1, x_2) = \mathrm{e}^{x_1+x_2}$ and its ANOVA terms.

3.2 Smoothness and Variance

We start this section by discussing the inheritance of affiliation to a function space from f to its ANOVA terms $f_{\boldsymbol{u}}$. This translates to the inheritance of smoothness from the function f to the ANOVA terms $f_{\boldsymbol{u}}$ if one considers subspaces that are defined through the smoothness of the function.

At first we consider the classical Sobolev spaces

$$\mathrm{H}^s(\mathbb{T}^d) = \left\{ f \in \mathrm{L}_2(\mathbb{T}^d) \colon D^{\boldsymbol{\alpha}} f \in \mathrm{L}_2(\mathbb{T}^d) \text{ for } \boldsymbol{\alpha} \in \mathbb{N}_0^d, \|\boldsymbol{\alpha}\|_1 \leq s \right\}. \tag{3.10}$$

They are a special case of the Sobolev classes considered in [3] and we start by following the general path established therein. Since our context is $\mathrm{L}_2(\mathbb{T}^d)$ we use a different proof based on the equivalence of Fourier coefficients. We then generalize results of inheritance of smoothness to Sobolev type spaces, see [11], and the weighted Wiener algebra.

Corollary 3.16 *Let $\boldsymbol{u} \subset \mathcal{D}$ and $f \in \mathrm{H}^s(\mathbb{T}^d)$ with $s > 0$. Then*

$$\left\| D^{\boldsymbol{\beta}} P_{\boldsymbol{u}} f \right\|_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})} < \infty \quad \forall \boldsymbol{\beta} \in \mathbb{N}_0^d, \boldsymbol{\beta}_{\mathcal{D}\setminus\boldsymbol{u}} = 0, \|\boldsymbol{\beta}\|_1 \leq s$$

and therefore $P_{\boldsymbol{u}} f \in \mathrm{H}^s(\mathbb{T}^{|\boldsymbol{u}|})$.

Proof. This is a consequence of the well-known Leibniz Theorem or Leibniz rule as

$$\begin{aligned}
\int_{\mathbb{T}^{|\boldsymbol{u}|}} \left| D^{\boldsymbol{\beta}} \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \right|^2 \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} &= \int_{\mathbb{T}^{|\boldsymbol{u}|}} \left| \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} D^{\boldsymbol{\beta}} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \right|^2 \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\
&\leq \int_{\mathbb{T}^{|\boldsymbol{u}|}} \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} \left| D^{\boldsymbol{\beta}} f(\boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x}_{\mathcal{D}\setminus\boldsymbol{u}} \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\
&= \left\| D^{\boldsymbol{\beta}} f \right\|_{\mathrm{L}_2(\mathbb{T}^d)}^2 < \infty.
\end{aligned}$$

■

Theorem 3.17 (Extended Leibniz Theorem) *Let $\boldsymbol{u} \subset \mathcal{D}$ and $f \in \mathrm{H}^s(\mathbb{T}^d)$ with order $s > 0$. Then*

$$D^{\boldsymbol{\beta}} \left(P_{\boldsymbol{u}} f \right)(\boldsymbol{x}) = \left(P_{\boldsymbol{u}} \left(D^{\boldsymbol{\beta}} f \right) \right)(\boldsymbol{x})$$

with $\boldsymbol{\beta} \in \mathbb{N}_0^d, \|\boldsymbol{\beta}\|_1 \leq s$ and $\boldsymbol{\beta}_{\mathcal{D}\setminus\boldsymbol{u}} = 0$.

Proof. It is sufficient to prove that both $D^{\boldsymbol{\beta}} P_{\boldsymbol{u}} f$ and $P_{\boldsymbol{u}} D^{\boldsymbol{\beta}} f$ have the same Fourier coefficients by Theorem 2.5. Using the differentiation property of the Fourier coefficients and Lemma 3.2, we derive

$$D^{\boldsymbol{\beta}} \left(P_{\boldsymbol{u}} f \right)(\boldsymbol{x}) = D^{\boldsymbol{\beta}} \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} \hat{f}_{\boldsymbol{k}} \,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} (2\pi\mathrm{i}\boldsymbol{k})^{\boldsymbol{\beta}} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}.$$

If we set $g(\boldsymbol{x}) = D^{\boldsymbol{\beta}} f(\boldsymbol{x})$ then

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{g}_{\boldsymbol{k}} \,\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \text{ with } \hat{g}_{\boldsymbol{k}} = (2\pi\mathrm{i}\boldsymbol{k})^{\boldsymbol{\beta}} \hat{f}_{\boldsymbol{k}}.$$

Applying the projection operator $P_{\boldsymbol{u}}$ yields

$$(P_{\boldsymbol{u}} g)(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} \hat{g}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} = \sum_{\boldsymbol{k} \in \mathbb{P}_{\boldsymbol{u}}^{(d)}} (2\pi \mathrm{i} \boldsymbol{k})^{\boldsymbol{\beta}} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}},$$

see Lemma 3.2, and therefore the Fourier coefficients are identical. ■

Using Theorem 3.17 and Corollary 3.16 we show that f indeed inherits its smoothness to the ANOVA terms.

Theorem 3.18 (Inheritance of Smoothness) *Let $\boldsymbol{u} \subset \mathcal{D}$ and $f \in \mathrm{H}^s(\mathbb{T}^d)$ with $s > 0$. Then*

$$\left\| D^{\boldsymbol{\beta}} f_{\boldsymbol{u}} \right\|_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})} < \infty \quad \forall \boldsymbol{\beta} \in \mathbb{N}_0^d, \boldsymbol{\beta}_{\mathcal{D} \setminus \boldsymbol{u}} = 0, \|\boldsymbol{\beta}\|_1 \leq s$$

and therefore $f_{\boldsymbol{u}} \in \mathrm{H}^s(\mathbb{T}^{|\boldsymbol{u}|})$.

Proof. We prove this theorem by estimating the norm

$$\begin{aligned}
\left\| D^{\boldsymbol{\beta}} f_{\boldsymbol{u}} \right\|^2_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})} &= \int_{\mathbb{T}^{|\boldsymbol{u}|}} \left| D^{\boldsymbol{\beta}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) \right|^2 \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \overset{(3.9)}{=} \int_{\mathbb{T}^{|\boldsymbol{u}|}} \left| D^{\boldsymbol{\beta}} \sum_{\boldsymbol{v} \subset \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} (P_{\boldsymbol{v}} f)(\boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\
&\leq \sum_{\boldsymbol{v} \subset \boldsymbol{u}} \int_{\mathbb{T}^{|\boldsymbol{u}|}} \left| D^{\boldsymbol{\beta}} (P_{\boldsymbol{v}} f)(\boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \leq \sum_{\boldsymbol{v} \subset \boldsymbol{u}} \left\| D^{\boldsymbol{\gamma}^{(\boldsymbol{v})}} P_{\boldsymbol{v}} f \right\|_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{v}|})} < \infty
\end{aligned}$$

with $\boldsymbol{\gamma}^{(\boldsymbol{v})} \in \mathbb{N}^d$, $\boldsymbol{\gamma}^{(\boldsymbol{v})}_{\boldsymbol{v}} = \boldsymbol{\beta}_{\boldsymbol{v}}$ and $\boldsymbol{\gamma}^{(\boldsymbol{v})}_{\mathcal{D} \setminus \boldsymbol{v}} = \boldsymbol{0}$. ■

The previous results can be generalized to a much broader class of function spaces over the torus characterized by the decay of the Fourier coefficients. The first class we consider are the Sobolev type spaces $\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$ from Definition 2.8.

Theorem 3.19 (Inheritance of Smoothness for Sobolev Type Spaces) *Let $f \in \mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$ and $w : \mathbb{Z}^d \to [1, \infty)$ the function that generates the weight sequence $\boldsymbol{w}$. Then*

$$f_{\boldsymbol{u}} \in \mathrm{H}^{\overline{\boldsymbol{w}}}(\mathbb{T}^{|\boldsymbol{u}|}) \quad \forall \boldsymbol{u} \subset \mathcal{D}$$

for $\overline{\boldsymbol{w}} = (\overline{w}(\boldsymbol{\ell}))_{\boldsymbol{\ell} \in (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}}$ and $\overline{w}(\boldsymbol{\ell}) = w(\boldsymbol{k})$ if $\boldsymbol{k} \in \mathbb{Z}^d$ is chosen such that $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}} = \boldsymbol{0}$.

Proof. We use the identification from Corollary 3.6 and estimate the sum

$$\begin{aligned}
\|f_{\boldsymbol{u}}\|^2_{\mathrm{H}^{\overline{\boldsymbol{w}}}(\mathbb{T}^{|\boldsymbol{u}|})} = \sum_{\boldsymbol{\ell} \in (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}} \overline{w}^2(\boldsymbol{\ell}) \left| \hat{f}_{\boldsymbol{\ell}, \boldsymbol{u}} \right|^2 &= \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} \overline{w}^2(\boldsymbol{k}_{\boldsymbol{u}}) \left| \hat{f}_{\boldsymbol{k}} \right|^2 \\
&= \sum_{\boldsymbol{k} \in \mathbb{F}_{\boldsymbol{u}}^{(d)}} w^2(\boldsymbol{k}) \left| \hat{f}_{\boldsymbol{k}} \right|^2 \\
&\leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w^2(\boldsymbol{k}) \left| \hat{f}_{\boldsymbol{k}} \right|^2 = \|f\|^2_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)} < \infty.
\end{aligned}$$

■

With Theorem 3.19, the inheritance of smoothness follows immediately for the well-known special cases of Sobolev type spaces.

Corollary 3.20 *Let $f \in \mathrm{H}^{s,p}(\mathbb{T}^d)$ for $s > 0$ and $p > 0$. Then*

$$f_{\boldsymbol{u}} \in \mathrm{H}^{s,p}(\mathbb{T}^{|\boldsymbol{u}|}).$$

Proof. The space $\mathrm{H}^{s,p}(\mathbb{T}^d)$ with the weight function

$$w_{s,p}(\boldsymbol{k}) = \begin{cases} (1 + \|\boldsymbol{k}\|_p^p)^{\frac{s}{p}} & : p < \infty \\ \max(1, |k_1|, \ldots, |k_d|) & : p = \infty \end{cases}$$

from Definition 2.9 is a Sobolev type space. Therefore, the statement follows directly from Theorem 3.19. ■

Corollary 3.21 *Let $f \in \mathrm{H}^s_{\mathrm{mix}}(\mathbb{T}^d)$ for $s > 0$. Then*

$$f_{\boldsymbol{u}} \in \mathrm{H}^s_{\mathrm{mix}}(\mathbb{T}^{|\boldsymbol{u}|}).$$

Proof. The space $\mathrm{H}^s_{\mathrm{mix}}(\mathbb{T}^d)$ is equipped with the weight function

$$w_s(\boldsymbol{k}) = \prod_{i=1}^{d} (1 + |k_i|^2)^s$$

from Definition 2.10. Therefore, it is a Sobolev type space and we use Theorem 3.19. ■

Finally, we consider the Wiener algebra $\mathcal{A}_w(\mathbb{T}^d)$, see Definition 2.11, with weight function $w : \mathbb{Z}^d \to [1, \infty)$. Here, the inheritance of smoothness also holds.

Theorem 3.22 (Inheritance of Smoothness for the Wiener Algebra) *Let $f \in \mathcal{A}_w(\mathbb{T}^d)$ with weight function $w : \mathbb{Z}^d \to [1, \infty)$. Then*

$$f_{\boldsymbol{u}} \in \mathcal{A}_{\overline{w}}(\mathbb{T}^{|\boldsymbol{u}|}) \quad \forall \boldsymbol{u} \subset \mathcal{D}$$

with $\overline{w} : \mathbb{Z}^{|\boldsymbol{u}|} \to [1, \infty)$ given by $\overline{w}(\boldsymbol{\ell}) = w(\boldsymbol{k})$ for $\boldsymbol{k} \in \mathbb{Z}^d$, $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}} = \boldsymbol{0}$.

Proof. We have to show that the norm

$$\|f_{\boldsymbol{u}}\|_{\mathcal{A}_{\overline{w}}(\mathbb{T}^{|\boldsymbol{u}|})} = \sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} \overline{w}(\boldsymbol{\ell}) \left| \hat{f}_{\boldsymbol{\ell}, \boldsymbol{u}} \right|$$

is finite. Using the same trick as before we estimate

$$\sum_{\boldsymbol{\ell} \in \mathbb{Z}^{|\boldsymbol{u}|}} \overline{w}(\boldsymbol{\ell}) \left| \hat{f}_{\boldsymbol{\ell}, \boldsymbol{u}} \right| = \sum_{\boldsymbol{k} \in \mathbb{F}^d_{\boldsymbol{u}}} w(\boldsymbol{k}) \left| \hat{f}_{\boldsymbol{k}} \right| \leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} w(\boldsymbol{k}) \left| \hat{f}_{\boldsymbol{k}} \right| = \|f\|_{\mathcal{A}_w(\mathbb{T}^d)} < \infty.$$

■

The second part of this section is about the importance of a specific ANOVA term $f_{\boldsymbol{u}}$ in relation to the entire function. This is especially important in our overarching context of learning sparse additive models. Identifying terms of importance is one of the main ideas of the approximation approach in Sections 4 and 5. We start by defining the variance of the function $f \in \mathrm{L}_2(\mathbb{T}^d)$

$$\sigma^2(f) \coloneqq \int_{\mathbb{T}^d} (f(\boldsymbol{x}) - If)^2 \,\mathrm{d}\boldsymbol{x}. \tag{3.11}$$

We can relate this to the $\mathrm{L}_2(\mathbb{T}^d)$-norm through the following lemma.

Lemma 3.23 *Let $f : \mathbb{T}^d \to \mathbb{R}$ as before and $f \in \mathrm{L}_2(\mathbb{T}^d)$. Then we can express* (3.11) *as*

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

In a Fourier context, this translates to

$$\sigma^2(f) = \|f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} - \hat{f}_{\boldsymbol{0}}^2 = \sum_{\boldsymbol{k} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}} \left| \hat{f}_{\boldsymbol{k}} \right|^2. \tag{3.12}$$

Proof. Starting from (3.11), we derive

$$\begin{aligned}
\int_{\mathbb{T}^d} (f(\boldsymbol{x}) - If)^2 \,\mathrm{d}\boldsymbol{x} &= \int_{\mathbb{T}^d} \left[f^2(\boldsymbol{x}) - 2f(If) + (If)^2 \right] \,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{T}^d} f^2(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} - 2(If) \underbrace{\int_{\mathbb{T}^d} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}_{=If} + (If)^2 \\
&= \|f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} - (If)^2.
\end{aligned}$$

The second equation (3.12) follows directly from Parseval's identity (2.2). ■

For the variance of the ANOVA terms $f_{\boldsymbol{u}}$, we can apply Lemma 3.8 and get

$$\sigma^2(f_{\boldsymbol{u}}) = \|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})} - \underbrace{(If_{\boldsymbol{u}})^2}_{=0} = \|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})}$$

for $\boldsymbol{u} \neq \emptyset$. Together with Parseval's identity and Corollary 3.6 we can rewrite this as the sum

$$\sigma^2(f_{\boldsymbol{u}}) = \sum_{\boldsymbol{k} \in \mathbb{F}^{(d)}_{\boldsymbol{u}}} \left| \hat{f}_{\boldsymbol{k}} \right|^2 = \sum_{\boldsymbol{\ell} \in (\mathbb{Z} \setminus \{0\})^{|\boldsymbol{u}|}} \left| \hat{f}_{\boldsymbol{\ell}, \boldsymbol{u}} \right|^2. \tag{3.13}$$

With the following Theorem, we are able to relate the variance of ANOVA terms to the variance of the corresponding function.

Theorem 3.24 *Let $f \in \mathrm{L}_2(\mathbb{T}^d)$ with $Im(f) \subset \mathbb{R}$. Then we can express the variance of f as the sum of the variances of its ANOVA terms*

$$\sigma^2(f) = \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ \boldsymbol{u} \neq \emptyset}} \sigma^2(f_{\boldsymbol{u}}). \tag{3.14}$$

Proof. We start from the right-hand side in (3.14) and use (3.13)

$$\sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ \boldsymbol{u}\neq\emptyset}} \sigma^2(f_{\boldsymbol{u}}) = \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ \boldsymbol{u}\neq\emptyset}} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \left|\hat{f}_{\boldsymbol{k}}\right|^2.$$

Using the fact that $\mathbb{F}_{\boldsymbol{u}}^{(d)} \cap \mathbb{F}_{\boldsymbol{v}}^{(d)} = \emptyset$ for $\boldsymbol{u} \neq \boldsymbol{v}$ from Lemma 3.5 and $\bigcup_{\boldsymbol{u}\subset\mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)} = \mathbb{Z}^d$ as a consequence of Corollary 3.6 and Theorem 3.9, we arrive at

$$\sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ \boldsymbol{u}\neq\emptyset}} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \left|\hat{f}_{\boldsymbol{k}}\right|^2 = \sum_{\boldsymbol{k}\in\bigcup_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ \boldsymbol{u}\neq\emptyset}} \mathbb{F}_{\boldsymbol{u}}^{(d)}} \left|\hat{f}_{\boldsymbol{k}}\right|^2 = \sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} \left|\hat{f}_{\boldsymbol{k}}\right|^2.$$

Since this is equal to (3.12), we have proven the statement. ■

In order to measure the importance of a particular ANOVA term $f_{\boldsymbol{u}}$ we consider the global sensitivity indices as proposed in [15, 14]

$$\varrho(\boldsymbol{u}; f) \coloneqq \frac{\sigma^2(f_{\boldsymbol{u}})}{\sigma^2(f)} \in [0,1]. \tag{3.15}$$

Using (3.14) it directly follows that

$$\sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ \boldsymbol{u}\neq\emptyset}} \varrho(\boldsymbol{u}; f) = 1$$

and

$$\varrho(\boldsymbol{u}; f) = 1 - \sum_{\substack{\boldsymbol{v}\subset\mathcal{D}\\ \boldsymbol{v}\neq\boldsymbol{u}\\ \boldsymbol{v}\neq\emptyset}} \varrho(\boldsymbol{v}; f).$$

Now, we take another look at the Example 3.15 from before.

Example 3.25 We again consider the function $f(x_1, x_2) = \mathrm{e}^{x_1+x_2}$. In order to calculate the global sensitivity indices, we start with the variances

$$\begin{aligned}
\sigma^2(f) &= \int_{\mathbb{T}^2} (\mathrm{e}^{x_1+x_2} - c^2)^2 \,\mathrm{d}\boldsymbol{x} = \left(\frac{\mathrm{e}}{2} - \frac{1}{2\mathrm{e}}\right)^2 - c^4 \approx 0.201352\\
\sigma^2(f_{\{i\}}) &= \int_{\mathbb{T}} \left(c\mathrm{e}^{x_i} - c^2\right)^2 \,\mathrm{d}x_i = c^2\left[\frac{\mathrm{e}}{2} - \frac{1}{2\mathrm{e}}\right] - c^4 \approx 0.0967117, i = 1, 2,\\
\sigma^2(f_{\{1,2\}}) &= \int_{\mathbb{T}^2} \left(\mathrm{e}^{x_1+x_2} - c\left(\mathrm{e}^{x_1} + \mathrm{e}^{x_2}\right) + c^2\right)^2 \,\mathrm{d}\boldsymbol{x} \approx 0.00792811.
\end{aligned}$$

where c is given in Example 3.15. Clearly, $\sigma^2(f_1) + \sigma^2(f_2) + \sigma^2(f_{1,2}) = \sigma^2(f)$ and for the global sensitivity indices we have

$$\varrho(\{1\}; f) = \varrho(\{2\}; f) \approx 0.480313 \text{ and } \varrho(\{1,2\}; f) = 0.0393745.$$

This means that the two one-dimensional terms explain together about 96% of $\sigma^2(f)$ while the two-dimensional term contributes only about 4%. ◇

Remark 3.26 *One could get the impression from Example 3.25 that lower order terms do contribute more to the function in general. However, this does not need to be the case as we can create a counterexample by defining a function where low order ANOVA terms are zero.*

Nevertheless, we make the observation that e.g. in the classical isotropic Sobolev spaces $\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$*, the weights* $w_{s,p}(\boldsymbol{k})$ *with* $s > 0$ *and* $1 \leq p < \infty$ *are larger if more entries of* $\boldsymbol{k}$ *are nonzero which forces the Fourier coefficients to decay faster in those directions.* $\diamond$

3.3 Effective Dimensions

Now, we want to introduce the so called **effective dimensions**, see [4, Chapter 2.1.1]. Given a proportion $\alpha \in (0,1]$, we have two notions of an effective dimension. The first one is called **truncation dimension** $d_t \in \{1,2,\dots,d\}$ with respect to the function f. It is defined as the smallest value d_t such that

$$\sum_{\substack{\boldsymbol{u}\subset\{1,2,\dots,d_t\}\\ \boldsymbol{u}\neq\emptyset}} \sigma^2(f_{\boldsymbol{u}}) \geq \alpha\sigma^2(f).$$

Practically speaking, this reduces the dimension of the function f in a way that still allows us to explain a determined portion α of it. Note that arbitrary interactions of the remaining dimensions are still allowed. This provides the means to identify the number of *important* variables of the function f.

The second notion is the **superposition dimension** $d_s \in \{0,1,2,\dots,d-1\}$. This is the smallest value d_s such that

$$\sum_{\substack{\boldsymbol{u}\subset\{1,2,\dots,d\}\\ 0<|\boldsymbol{u}|\leq d_s}} \sigma^2(f_{\boldsymbol{u}}) \geq \alpha\sigma^2(f).$$

Here, we do not limit the variables itself but the number of variables that can interact with each other. The superposition dimension d_s is therefore also the highest order of allowed interactions. In some further considerations, we omit the specification of an α and instead directly choose d_s. In this case, we have

$$\alpha := \frac{\sum_{\substack{\boldsymbol{u}\subset\{1,2,\dots,d\}\\ 0<|\boldsymbol{u}|\leq d_s}} \sigma^2(f_{\boldsymbol{u}})}{\sigma^2(f)}.$$

While both effective dimensions can be useful in different context, in this thesis we will only consider the superposition dimension and build a corresponding model. The reason for using this as a basis for our approximation approach will become clear in Sections 4 and 5. A certain variant of this idea (in a different context) was proposed in [4, Chapter 3.2.4].

Definition 3.27 For a given superposition dimension d_s, we define the truncation operator

$$\mathcal{T}_{d_s} : \mathrm{L}_2(\mathbb{T}^d) \to \mathrm{L}_2(\mathbb{T}^d), f \mapsto \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ |\boldsymbol{u}|\leq d_s}} f_{\boldsymbol{u}}.$$

For a fixed function $f : \mathbb{T}^d \to \mathbb{R}$, we call the image of the truncation operator

$$\mathcal{T}_{d_s} f := \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\ |\boldsymbol{u}|\leq d_s}} f_{\boldsymbol{u}} \tag{3.16}$$

the **approximate ANOVA model** with respect to d_s. ◇

Clearly, the operator $\mathcal{T}_{d_s}$ is well-defined since each $f \in \mathrm{L}_2(\mathbb{T}^d)$ has a unique ANOVA decomposition by Theorem 3.9 and $f_{\boldsymbol{u}} \in \mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})$ for each $\boldsymbol{u} \subset \mathcal{D}$ by Lemma 3.3.

Remark 3.28 *For a fixed superposition dimension d_s the number of terms has polynomial growth in d as we can estimate*

$$\sum_{i=0}^{d_s} \binom{d}{i} \leq (1+d)^{d_s}$$

by applying the Binomial Theorem. ◇

First, we want to propose a direct formula with the projections $P_{\boldsymbol{u}} f$ using the following theorem.

Theorem 3.29 *Let $d_s \in \{1, 2, \ldots, d-1\}$. Then we can express the approximate ANOVA model directly by*

$$\mathcal{T}_{d_s} f = \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} c(|\boldsymbol{u}|, d, d_s) P_{\boldsymbol{u}} f \tag{3.17}$$

with coefficients

$$c(|\boldsymbol{u}|, d, d_s) = \sum_{n=|\boldsymbol{u}|}^{d_s} \binom{d-|\boldsymbol{u}|}{n-|\boldsymbol{u}|} (-1)^{n-|\boldsymbol{u}|}. \tag{3.18}$$

Proof. Here, a similar technique as in Theorem 3.12 can be applied. We start with (3.16) and use (3.9)

$$\begin{aligned}
\sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} f_{\boldsymbol{u}} &= \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} \sum_{\boldsymbol{v} \subset \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} P_{\boldsymbol{v}} f \\
&= \sum_{n=0}^{d_s} \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| = n}} \sum_{\substack{\boldsymbol{v} \subset \mathcal{D} \\ |\boldsymbol{v}| \leq d_s}} (-1)^{n-|\boldsymbol{v}|} P_{\boldsymbol{v}} f \delta_{\boldsymbol{v} \subset \boldsymbol{u}} \\
&= \sum_{n=0}^{d_s} \sum_{\substack{\boldsymbol{v} \subset \mathcal{D} \\ |\boldsymbol{v}| \leq d_s}} (-1)^{n-|\boldsymbol{v}|} P_{\boldsymbol{v}} f \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| = n}} \delta_{\boldsymbol{v} \subset \boldsymbol{u}} \\
&= \sum_{n=0}^{d_s} \sum_{\substack{\boldsymbol{v} \subset \mathcal{D} \\ |\boldsymbol{v}| \leq d_s}} (-1)^{n-|\boldsymbol{v}|} P_{\boldsymbol{v}} f \binom{d-|\boldsymbol{v}|}{n-|\boldsymbol{v}|} \\
&= \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} \left[\sum_{n=|\boldsymbol{u}|}^{d_s} \binom{d-|\boldsymbol{u}|}{n-|\boldsymbol{u}|} (-1)^{n-|\boldsymbol{u}|} \right] P_{\boldsymbol{u}} f.
\end{aligned}$$

■

Now, we consider the Fourier coefficients of the approximate model $\mathcal{T}_{d_s} f$.

Corollary 3.30 *For a given $\boldsymbol{k} \in \mathbb{Z}^d$, we have*

$$c_{\boldsymbol{k}}(\mathcal{T}_{d_s} f) = \begin{cases} \hat{f}_{\boldsymbol{k}} & : \|\boldsymbol{k}\|_0 \leq d_s \\ 0 & : \text{otherwise} \end{cases} \tag{3.19}$$

with $c_{\boldsymbol{k}}(\mathcal{T}_{d_s} f)$ being the Fourier coefficient of $\mathcal{T}_{d_s} f$ for the frequency $\boldsymbol{k}$ and $\|\boldsymbol{k}\|_0$ the number of entries in $\boldsymbol{k}$ that are non-zero. Furthermore, the coefficients $c_{\boldsymbol{k}}(\mathcal{T}_{d_s} f)$ can be expressed via the Fourier coefficients of the projections as follows

$$c_{\boldsymbol{k}}(\mathcal{T}_{d_s} f) = \begin{cases} \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} c(|\boldsymbol{u}|, d, d_s) \hat{p}_{\boldsymbol{k}_{\boldsymbol{u}}, \boldsymbol{u}} \, \delta_{\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}}, \boldsymbol{0}} & : \|\boldsymbol{k}\|_0 \leq d_s \\ 0 & : \text{otherwise} \end{cases}. \tag{3.20}$$

Here, $c(|\boldsymbol{u}|, d, d_s)$ is from (3.18) *and $\hat{p}_{\boldsymbol{k}_{\boldsymbol{u}}, \boldsymbol{u}}$ is the $\boldsymbol{k}_{\boldsymbol{u}}$-th Fourier coefficient of $P_{\boldsymbol{u}} f$.*

Proof. Clearly, the formula (3.19) holds since the operator $\mathcal{T}_{d_s}$ simply cuts all ANOVA terms of order larger than d_s and therefore the corresponding Fourier coefficients.

In order to prove (3.20), we use Theorem 3.29 and calculate the integral

$$\begin{aligned} c_{\boldsymbol{k}}(\mathcal{T}_{d_s} f) &= \int_{\mathbb{T}^d} \left(\sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} f_{\boldsymbol{u}}(\boldsymbol{x}) \right) \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k}\boldsymbol{x}} \, \mathrm{d}\boldsymbol{x} \\ &\overset{\text{Theorem 3.29}}{=} \int_{\mathbb{T}^d} \left(\sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} c(|\boldsymbol{u}|, d, d_s)(P_{\boldsymbol{u}} f)(\boldsymbol{x}) \right) \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k}\boldsymbol{x}} \, \mathrm{d}\boldsymbol{x} \\ &= \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} c(|\boldsymbol{u}|, d, d_s) \int_{\mathbb{T}^{|\boldsymbol{u}|}} (P_{\boldsymbol{u}} f)(\boldsymbol{x}_{\boldsymbol{u}}) \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k}_{\boldsymbol{u}} \boldsymbol{x}_{\boldsymbol{u}}} \, \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \, \delta_{\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}}, \boldsymbol{0}} \\ &= \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} c(|\boldsymbol{u}|, d, d_s) \delta_{\boldsymbol{k}_{\mathcal{D} \setminus \boldsymbol{u}}, \boldsymbol{0}} \, \hat{p}_{\boldsymbol{k}_{\boldsymbol{u}}, \boldsymbol{u}}. \end{aligned}$$

■

With regard to the approximation of functions, we are interested in the error if we approximate f by $\mathcal{T}_{d_s} f$. Here, we can look at the same function spaces as in Section 3.2 and derive the following results.

Lemma 3.31 *Let $f \in \mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$ with generating weight function $w : \mathbb{Z}^d \to [1, \infty)$. Then*

$$\sum_{\boldsymbol{u} \subset \mathcal{D}} \|f_{\boldsymbol{u}}\|^2_{\mathrm{H}^{\boldsymbol{w}_{\boldsymbol{u}}}(\mathbb{T}^{|\boldsymbol{u}|})} = \|f\|^2_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)}$$

with $\boldsymbol{w_u} = (w_{\boldsymbol{u}}(\boldsymbol{l}))_{\boldsymbol{l}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}}$ *and the weight functions* $w_{\boldsymbol{u}} : \mathbb{Z}^{|\boldsymbol{u}|} \to [1,\infty)$ *under the condition that*

$$w_{\boldsymbol{u}}(\boldsymbol{k}_{\boldsymbol{u}}) = w(\boldsymbol{k})\, \forall \boldsymbol{u} \subset \mathcal{D} \colon \forall \boldsymbol{k} \in \mathbb{Z}^d : \boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}. \tag{3.21}$$

Proof. Using the definition of the norm and Lemma 3.5 we have

$$\begin{aligned}\sum_{\boldsymbol{u}\subset\mathcal{D}} \|f_{\boldsymbol{u}}\|^2_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)} &= \sum_{\boldsymbol{u}\subset\mathcal{D}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} w^2_{\boldsymbol{u}}(\boldsymbol{\ell}) \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|^2 = \sum_{\boldsymbol{u}\subset\mathcal{D}} \sum_{\boldsymbol{k}\in\mathbb{F}^{(d)}_{\boldsymbol{u}}} w^2(\boldsymbol{k}) \left|\hat{f}_{\boldsymbol{k}}\right|^2 \\ &= \sum_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\subset\mathcal{D}}\mathbb{F}^{(d)}_{\boldsymbol{u}}} w^2(\boldsymbol{k}) \left|\hat{f}_{\boldsymbol{k}}\right|^2 = \|f\|^2_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)}.\end{aligned}$$

■

Theorem 3.32 *Let* $d_s \in \{1,2,\ldots,d-1\}$ *be the superposition dimension of an approximate ANOVA model* $\mathcal{T}_{d_s} f$ *and* $f \in \mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$ *with weight function* $w : \mathbb{Z}^d \to [1,\infty)$. *Furthermore, let* $\mathrm{H}^{\boldsymbol{w_u}}$ *for* $\boldsymbol{u} \subset \mathcal{D}$ *be generated by a weight function such that* (3.21) *holds. Then we estimate the error as*

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_2(\mathbb{T}^d)} \leq \frac{1}{\inf_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\subset\mathcal{D},|\boldsymbol{u}|>d_s}\mathbb{F}^d_{\boldsymbol{u}}} w(\boldsymbol{k})} \|f\|_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)}.$$

Proof. We start by using the orthogonality of the ANOVA terms (Lemma 3.7) and get

$$\|f - \mathcal{T}_{d_s} f\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \left\| \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} f_{\boldsymbol{u}} \right\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^d)} = \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})}.$$

We apply Parseval's identity (2.2) and Lemma 3.31 to estimate

$$\begin{aligned}\sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2(\mathbb{T}^{|\boldsymbol{u}|})} &= \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|^2 \\ &= \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} \frac{w^2_{\boldsymbol{u}}(\boldsymbol{\ell})}{w^2_{\boldsymbol{u}}(\boldsymbol{\ell})} \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|^2 \\ &\leq \frac{1}{\min_{\boldsymbol{u}\subset\mathcal{D},|\boldsymbol{u}|>d_s} \inf_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} w^2_{\boldsymbol{u}}(\boldsymbol{\ell})} \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} w^2_{\boldsymbol{u}}(\boldsymbol{\ell}) \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|^2 \\ &\leq \frac{1}{\min_{\boldsymbol{u}\subset\mathcal{D},|\boldsymbol{u}|>d_s} \inf_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} w^2_{\boldsymbol{u}}(\boldsymbol{\ell})} \|f\|^2_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)} \\ &= \frac{1}{\inf_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\subset\mathcal{D},|\boldsymbol{u}|>d_s}\mathbb{F}^d_{\boldsymbol{u}}} w^2(\boldsymbol{k})} \|f\|^2_{\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)}.\end{aligned}$$

■

This theorem can be directly applied to the special Sobolev type spaces we considered earlier, see Definitions 2.9 and 2.10.

Corollary 3.33 *Let $f \in \mathrm{H}^{s,p}(\mathbb{T}^d)$ for $s > 0$ and $0 < p < \infty$. Then*

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_2(\mathbb{T}^d)} \leq (2 + d_s)^{-\frac{s}{p}} \|f\|_{\mathrm{H}^{s,p}(\mathbb{T}^d)} .$$

Proof. For the weight functions we know that

$$\inf_{\boldsymbol{\ell} \in (\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} (1 + \|\boldsymbol{\ell}\|_p^p)^{\frac{s}{p}} = (1 + |\boldsymbol{u}|)^{\frac{s}{p}}$$

and moreover

$$\min_{\boldsymbol{u} \subset \mathcal{D}, |\boldsymbol{u}| > d_s} (1 + |\boldsymbol{u}|)^{\frac{s}{p}} = (2 + d_s)^{\frac{s}{p}}.$$

Now the estimate follows directly from Theorem 3.32. ∎

Corollary 3.34 *Let $f \in \mathrm{H}^s_{\mathrm{mix}}(\mathbb{T}^d)$ for $s > 0$. Then*

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_2(\mathbb{T}^d)} \leq 2^{-s(d_s+1)} \|f\|_{\mathrm{H}^s_{\mathrm{mix}}(\mathbb{T}^d)} .$$

Proof. We start again by the minimization problem

$$\inf_{\boldsymbol{\ell} \in (\mathbb{Z}\setminus\{0\})^{|\boldsymbol{u}|}} \prod_{i=1}^{d} (1 + |l_i|^2)^s = 2^{s|\boldsymbol{u}|}$$

and then

$$\min_{\boldsymbol{u} \subset \mathcal{D}, |\boldsymbol{u}| > d_s} 2^{s|\boldsymbol{u}|} = 2^{s(d_s+1)}.$$

The estimate follows again directly from Theorem 3.32. ∎

As before we also consider subsets of the weighted Wiener algebra $\mathcal{A}_w(\mathbb{T}^d)$ where we find that a similar estimate holds.

Theorem 3.35 *Let $d_s \in \{1, 2, \ldots, d-1\}$ be the superposition dimension of an approximate ANOVA model $\mathcal{T}_{d_s} f$ and $f \in \mathcal{A}_w(\mathbb{T}^d)$ with $\mathcal{A}_w(\mathbb{T}^d)$ as in Theorem 3.22. Then we estimate the error as*

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq \frac{1}{\inf_{\boldsymbol{k} \in \bigcup_{\boldsymbol{u} \subset \mathcal{D}, |\boldsymbol{u}| > d_s} \mathbb{F}^d_{\boldsymbol{u}}} w(\boldsymbol{k})} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)} .$$

Proof. Using that $\mathcal{A}_w(\mathbb{T}^d) \subset \mathrm{C}(\mathbb{T}^d)$ by Theorem 2.12 and the Fourier series representation of $f_{\boldsymbol{u}}$ from Corollary 3.6, we deduce

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} = \sup_{\boldsymbol{x} \in \mathbb{T}^d} |f(\boldsymbol{x}) - \mathcal{T}_{d_s} f(\boldsymbol{x})| = \sup_{\boldsymbol{x} \in \mathbb{T}^d} \left| \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| > d_s}} f_{\boldsymbol{u}}(\boldsymbol{x}) \right| \leq \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| > d_s}} \sum_{\boldsymbol{k} \in \mathbb{F}^{(d)}_{\boldsymbol{u}}} \left| \hat{f}_{\boldsymbol{k}} \right| .$$

With the same arguments as in the proof of Theorem 3.32, we estimate

$$\sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|>d_s}} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \left|\hat{f}_{\boldsymbol{k}}\right| \leq \frac{1}{\inf_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\subset\mathcal{D},|\boldsymbol{u}|>d_s}\mathbb{F}_{\boldsymbol{u}}^d} w(\boldsymbol{k})} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.$$

■

Now, we directly apply the result for mixed and isotropic smoothness spaces.

Corollary 3.36 *Let $f \in \mathcal{A}_w(\mathbb{T}^d)$ with*

$$w(\boldsymbol{k}) = w_{s,p}(\boldsymbol{k}) \coloneqq (1 + \|\boldsymbol{k}\|_p^p)^{s/p}.$$

Then

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq (2 + d_s)^{-\frac{s}{p}} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.$$

Proof. The result follows immediately from the arguments in the proof of Corollary 3.33 and Theorem 3.35. ■

Corollary 3.37 *Let $f \in \mathcal{A}_w(\mathbb{T}^d)$ with*

$$w(\boldsymbol{k}) = \prod_{i=1}^{d} (1 + |k_i|^2)^s.$$

Then

$$\|f - \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq 2^{-s(d_s+1)} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.$$

Proof. We discussed the solution to the minimization problem in the statement of Theorem 3.35 in the proof of Corollary 3.34 and can therefore apply the theorem. ■

4 Approximation with Black-Box-Access

In this section we consider the following fundamental problem.

Problem 4.1 *Let $f : \mathbb{T}^d \to \mathbb{R}$ be a function in $C(\mathbb{T}^d) \subset \mathrm{L}_2(\mathbb{T}^d)$ with $d \in \mathbb{N}$. We assume to have **black-box-access**, i.e., the ability to evaluate f at any point $\boldsymbol{x} \in \mathbb{T}^d$. Furthermore, a superposition dimension $d_s \in \mathbb{N}$ with $d_s \leq d$ is given.*
We want to find an approximation for f based on the approximate ANOVA model $\mathcal{T}_{d_s} f$. Furthermore, we are looking for important dimension interactions, i.e., the ANOVA terms that contribute significantly to f, in other words, sets $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$ whose global sensitivity index $\varrho(\boldsymbol{u}; f)$ is large.

The general problem of finding an approximation for f with black-box-access has already been considered a number of times with a multitude of approaches, e.g. in [21, 8, 9] to name a select few. We specifically emphasize that we choose our approach with regard to the secondary problem of finding the important dimension interactions. Using the approximate ANOVA model as a foundation will allow us to do that.

Furthermore, we choose rank-1 lattice, see Section 2.2, as sampling schemes and make use of the already existing theory surrounding function approximation using those lattice, see [7, 18]. While this approximation approach has yielded considerable success, to our knowledge, there haven't been significant advancements in the area of detecting important dimension interactions from these results.

We bring both concepts, the ANOVA decomposition and function reconstruction with rank-1 lattice, together to solve the combined problem of finding an approximation while simultaneously being able to understand the dimension interactions of the function.

4.1 Active Set Construction and Error Bounds

The Problem 4.1 can be approached from different directions. We always have the chosen superposition dimension d_s and therefore the approximate ANOVA model $\mathcal{T}_{d_s} f$, see Definition 3.27, as a foundation. The first goal is to determine an approximation for the function $\mathcal{T}_{d_s} f$ that allows us to find the important dimension interactions.

We start by considering an initial index set $I \subset \mathbb{Z}^d$ and the corresponding Fourier partial sum. With regard to the interpretability of the results, the index set cannot be chosen arbitrary since ANOVA terms $f_{\boldsymbol{u}}$, see (3.6), of the same order should be supported on isomorphic low-dimensional index sets. Otherwise, one cannot expect to find good approximate values for the global sensitivity indices $\varrho(\boldsymbol{u}; f)$.

This means that we do not choose I directly, but rather index sets $I_1 \subset \mathbb{Z} \setminus \{0\}, I_2 \subset (\mathbb{Z} \setminus \{0\})^2, \dots, I_{d_s} \subset (\mathbb{Z} \setminus \{0\})^{d_s}$ for every order of the ANOVA terms in our approximate model. We collect all frequency indices as

$$I = \bigcup_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} I_{\boldsymbol{u}}^{(d)} \tag{4.1}$$

where $I_{\boldsymbol{u}}^{(d)} = \{\boldsymbol{k}\in\mathbb{Z}^d\colon \boldsymbol{k}_{\boldsymbol{u}}\in I_{|\boldsymbol{u}|}, \boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}\}$. This union is disjoint by Lemma 3.5. Now, we consider the cutoff-error for the index set I.

Lemma 4.2 *Let $f\in\mathcal{A}_w(\mathbb{T}^d)$ with weight function $w:\mathbb{Z}^d\to[1,\infty)$, $I_i = \{\boldsymbol{k}\in(\mathbb{Z}\setminus\{0\})^i\colon w(\boldsymbol{k})\le N_i\}$ for $N_i\in\mathbb{N}$, $i=1,2,\dots,d_s$ and I the union* (4.1)*. Then*

$$\|\mathcal{T}_{d_s}f - S_I\mathcal{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \le \frac{1}{\min_{i=1,2,\dots,d_s} N_i}\|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.$$

Here, S_I is the Fourier partial sum operator, see (2.3)*.*

Proof. By Parseval's identity (2.2) and the disjointness of the index sets, we get

$$\begin{aligned}
\|\mathcal{T}_{d_s}f - S_I\mathcal{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} &= \sup_{\boldsymbol{x}\in\mathbb{T}^d}|\mathcal{T}_{d_s}f(\boldsymbol{x}) - S_I\mathcal{T}_{d_s}f(\boldsymbol{x})|\\
&\le \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}}\sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}}\left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|\\
&= \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}}\sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}}\frac{w_{\boldsymbol{u}}(\boldsymbol{\ell})}{w_{\boldsymbol{u}}(\boldsymbol{\ell})}\left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|\\
&\le \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}}\frac{1}{N_{|\boldsymbol{u}|}}\sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}} w_{\boldsymbol{u}}(\boldsymbol{\ell})\left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right|\\
&\le \frac{1}{\min_{i=1,2,\dots,d_s} N_i}\|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.
\end{aligned}$$

■

We use Lemma 4.2 to get an upper bound on the error if we approximate f with $S_I\mathcal{T}_{d_s}f$, i.e., the approximate ANOVA model supported on I.

Theorem 4.3 *Let $f\in\mathcal{A}_w(\mathbb{T}^d)$ with weight function $w:\mathbb{Z}^d\to[1,\infty)$, $I_i = \{\boldsymbol{k}\in(\mathbb{Z}\setminus\{0\})^i\colon w(\boldsymbol{k})\le N_i\}$ for $N_i\in\mathbb{N}$, $i=1,2,\dots,d_s$ and I as in* (4.1)*. Then*

$$\|f - S_I\mathcal{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \le \left(\frac{1}{\inf_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\subset\mathcal{D},|\boldsymbol{u}|>d_s}\mathbb{F}_{\boldsymbol{u}}^d} w(\boldsymbol{k})} + \frac{1}{\min_{i=1,2,\dots,d_s} N_i}\right)\|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.$$

Proof. We start with the triangle inequality in $\mathrm{L}_\infty(\mathbb{T}^d)$

$$\|f - S_I\mathcal{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \le \|f - \mathcal{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} + \|\mathcal{T}_{d_s}f - S_I\mathcal{T}_{d_s}f\|_{\mathrm{L}_\infty(\mathbb{T}^d)}.$$

For the first summand on the right-hand-side, we use Theorem 3.35 and for the second, Lemma 4.2. We get

$$\begin{aligned}\|f - S_I \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} &\leq \frac{1}{\inf_{\boldsymbol{k} \in \bigcup_{\boldsymbol{u} \subset \mathcal{D}, |\boldsymbol{u}| > d_s} \mathbb{F}_{\boldsymbol{u}}^d} w(\boldsymbol{k})} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)} + \frac{1}{\min_{i=1,2,\dots,d_s} N_i} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)} \\ &= \left(\frac{1}{\inf_{\boldsymbol{k} \in \bigcup_{\boldsymbol{u} \subset \mathcal{D}, |\boldsymbol{u}| > d_s} \mathbb{F}_{\boldsymbol{u}}^d} w(\boldsymbol{k})} + \frac{1}{\min_{i=1,2,\dots,d_s} N_i} \right) \|f\|_{\mathcal{A}_w(\mathbb{T}^d)} .\end{aligned}$$

■

Now we proceed with reconstructing the Fourier coefficients of the ANOVA terms on the index set I. We approximate f by the Fourier partial sum w.r.t. I of $\mathcal{T}_{d_s} f$

$$f(\boldsymbol{x}) \approx S_I \mathcal{T}_{d_s} f(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} \sum_{\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}} \hat{f}_{\boldsymbol{\ell}, \boldsymbol{u}} \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}}. \tag{4.2}$$

Given a rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ with $\boldsymbol{z} \in \mathbb{Z}^d$ and $M \in \mathbb{N}$, see Definition 2.13, we rewrite the approximation of (4.2) as a system of linear equations

$$\begin{aligned}\boldsymbol{f} &= \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} \boldsymbol{F}_{\boldsymbol{u}} \hat{\boldsymbol{f}}_{\boldsymbol{u}} \qquad (4.3) \\ &= (\boldsymbol{F}_{\boldsymbol{u}_1} \, \boldsymbol{F}_{\boldsymbol{u}_2} \, \cdots \, \boldsymbol{F}_{\boldsymbol{u}_n}) \begin{pmatrix} \hat{\boldsymbol{f}}_{\boldsymbol{u}_1} \\ \hat{\boldsymbol{f}}_{\boldsymbol{u}_2} \\ \vdots \\ \hat{\boldsymbol{f}}_{\boldsymbol{u}_n} \end{pmatrix} \\ &= \boldsymbol{F} \hat{\boldsymbol{f}}. \qquad (4.4)\end{aligned}$$

Note that $\boldsymbol{u}_j$ for $j = 1, 2, \dots, n$ with

$$n = \sum_{i=0}^{d_s} \binom{d}{i}$$

is an ordering for the ANOVA terms that can be arbitrary but has to be consistent. Moreover, we have $\boldsymbol{f} = (f(\boldsymbol{x}))_{\boldsymbol{x} \in \Lambda(\boldsymbol{z}, M)}$, $\hat{\boldsymbol{f}}_{\boldsymbol{u}} = (\hat{f}_{\boldsymbol{\ell}, \boldsymbol{u}})_{\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}}$ and the Fourier matrices

$$\boldsymbol{F}_{\boldsymbol{u}} = \left(\mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}} \right)_{\boldsymbol{x} \in \Lambda(\boldsymbol{z}_{\boldsymbol{u}}, M), \boldsymbol{\ell} \in I_{|\boldsymbol{u}|}} . \tag{4.5}$$

We can use the dimension-reduced lattice $\Lambda(\boldsymbol{z}_{\boldsymbol{u}}, M)$ in the matrices $\boldsymbol{F}_{\boldsymbol{u}}$ because of Lemma 2.14. Furthermore, we have the Fourier block matrix $\boldsymbol{F} \coloneqq (\boldsymbol{F}_{\boldsymbol{u}_1} \, \boldsymbol{F}_{\boldsymbol{u}_2} \, \cdots \, \boldsymbol{F}_{\boldsymbol{u}_n})$ and the vector

$$\hat{\boldsymbol{f}} \coloneqq \begin{pmatrix} \hat{\boldsymbol{f}}_{\boldsymbol{u}_1} \\ \hat{\boldsymbol{f}}_{\boldsymbol{u}_2} \\ \vdots \\ \hat{\boldsymbol{f}}_{\boldsymbol{u}_n} \end{pmatrix} .$$

Finding an approximation for $S_I \mathcal{T}_{d_s} f$, i.e., its Fourier coefficients, requires us to solve the least squares problem that belongs to (4.4) for $\hat{\boldsymbol{f}}$. This is only uniquely solvable if the Fourier matrix $\boldsymbol{F}$ has full rank. A condition for this is delivered by the following lemma.

Lemma 4.4 *Let $\boldsymbol{F} \in \mathbb{C}^{M,|I|}$ be the Fourier matrix generated by the rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ with blocks as in* (4.5) *and I the finite index set* (4.1) *with $M \geq |I|$. Then*

$$\operatorname{rank} \boldsymbol{F} = |I|$$

if $\Lambda(\boldsymbol{z_u}, M, I_{|\boldsymbol{u}|})$ is a reconstructing rank-1 lattice for each $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$ and

$$\boldsymbol{\ell} \cdot \boldsymbol{z_u} \not\equiv \boldsymbol{h} \cdot \boldsymbol{z_v} \pmod{M} \tag{4.6}$$

for $\boldsymbol{u}, \boldsymbol{v} \subset \mathcal{D}$, $|\boldsymbol{u}|, |\boldsymbol{v}| \leq d_s$ and $\boldsymbol{u} \neq \boldsymbol{v}$ with $\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}$ and $\boldsymbol{h} \in I_{|\boldsymbol{v}|}$. Moreover, in this case

$$\boldsymbol{F}^H \boldsymbol{F} = M \mathbf{I}_{|I|}. \tag{4.7}$$

Proof. First, we consider the structure of $\boldsymbol{F}^H \boldsymbol{F}$

$$\begin{aligned}
\boldsymbol{F}^H \boldsymbol{F} &= \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{u}_1}^H \\ \boldsymbol{F}_{\boldsymbol{u}_2}^H \\ \vdots \\ \boldsymbol{F}_{\boldsymbol{u}_n}^H \end{pmatrix} \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{u}_1} & \boldsymbol{F}_{\boldsymbol{u}_2} & \cdots & \boldsymbol{F}_{\boldsymbol{u}_n} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{u}_1}^H \boldsymbol{F}_{\boldsymbol{u}_1} & \boldsymbol{F}_{\boldsymbol{u}_1}^H \boldsymbol{F}_{\boldsymbol{u}_2} & \cdots & \boldsymbol{F}_{\boldsymbol{u}_1}^H \boldsymbol{F}_{\boldsymbol{u}_n} \\ \boldsymbol{F}_{\boldsymbol{u}_2}^H \boldsymbol{F}_{\boldsymbol{u}_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \boldsymbol{F}_{\boldsymbol{u}_n}^H \boldsymbol{F}_{\boldsymbol{u}_1} & \cdots & \cdots & \boldsymbol{F}_{\boldsymbol{u}_n}^H \boldsymbol{F}_{\boldsymbol{u}_n} \end{pmatrix}.
\end{aligned}$$

Theorem 2.16 tells us that $\boldsymbol{F}_{\boldsymbol{u}_i}^H \boldsymbol{F}_{\boldsymbol{u}_i} = M\mathbf{I}$. Furthermore, we know by (2.5) that

$$\left(\boldsymbol{F}_{\boldsymbol{u}_i}^H \boldsymbol{F}_{\boldsymbol{u}_j}\right)_{\boldsymbol{\ell} \in I_{|\boldsymbol{u}_i|}, \boldsymbol{h} \in I_{|\boldsymbol{u}_j|}} = M \cdot \delta_{\boldsymbol{\ell} \cdot \boldsymbol{z_u} \equiv \boldsymbol{h} \cdot \boldsymbol{z_v} \pmod{M}}$$

for $i \neq j$. In this case, the condition (4.6) ensures that each entry is 0. Therefore our matrix has full rank and is of the form

$$\boldsymbol{F}^H \boldsymbol{F} = \begin{pmatrix} M\mathbf{I}_{|I_{|\boldsymbol{u}_1|}|} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & M\mathbf{I}_{|I_{|\boldsymbol{u}_2|}|} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & M\mathbf{I}_{|I_{|\boldsymbol{u}_n|}|} \end{pmatrix} = M\mathbf{I}_{|I|}.$$

■

The solution to the least squares problem (4.4) is given by the **Moore-Penrose-Inverse** which reads as follows if we apply Lemma 4.4

$$\boldsymbol{F}^{\dagger} = (\boldsymbol{F}^H \boldsymbol{F})^{-1} \boldsymbol{F}^H \overset{\text{Lemma 4.4}}{=} (M \mathbf{I}_{|I|})^{-1} \boldsymbol{F}^H = \frac{1}{M} \boldsymbol{F}^H.$$

Moreover, multiplication with $\boldsymbol{F}^H$ is reduced to multiplication with the blocks since the solution to (4.4) now reads as

$$\frac{1}{M} \boldsymbol{F}^H \boldsymbol{f} = \begin{pmatrix} \frac{1}{M} \boldsymbol{F}_{\boldsymbol{u}_1}^H \boldsymbol{f} \\ \frac{1}{M} \boldsymbol{F}_{\boldsymbol{u}_2}^H \boldsymbol{f} \\ \vdots \\ \frac{1}{M} \boldsymbol{F}_{\boldsymbol{u}_n}^H \boldsymbol{f} \end{pmatrix} = \begin{pmatrix} \tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}_1} \\ \tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}_2} \\ \vdots \\ \tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}_n} \end{pmatrix} =: \tilde{\hat{\boldsymbol{f}}}.$$

with $\tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}}$ being **approximate Fourier coefficients** for the ANOVA term $f_{\boldsymbol{u}}$ in (4.2). This means that we have to perform one matrix-vector-multiplication per ANOVA term. Those multiplications can be done efficiently using the adjoint LFFT algorithm 2.2.

The previous considerations give rise to Algorithm 4.1.

Algorithm 4.1 Function reconstruction over special rank-1 lattice

Input:	$d \in \mathbb{N}$	spatial dimension of $f : \mathbb{T}^d \to \mathbb{R}$
	$d_s \in \mathbb{N}$	superposition dimension with $d_s < d$
	$I_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \ldots, d_s$	finite frequency index sets
	$\boldsymbol{z} \in \mathbb{Z}^d$	generating vector satisfying the conditions in Lemma 4.4
	$M \in \mathbb{N}$	lattice size
	f	black box function

1: $\boldsymbol{f} \leftarrow \left(f(\frac{1}{M}(j\boldsymbol{z} \mod M))\right)_{j=0}^{M-1}$
2: **for** $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$ **do**
3: $\quad \tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}} \leftarrow \frac{1}{M} \boldsymbol{F}_{\boldsymbol{u}}^H \boldsymbol{f}$ $\qquad \triangleleft$ $|\boldsymbol{u}|$-variate adjoint LFFT, see Algorithm 2.2
4: **end for**

Output:	$\tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}} \in \mathbb{C}^{\lvert I_{\lvert \boldsymbol{u} \rvert} \rvert}, \boldsymbol{u} \subset \mathcal{D}, \lvert \boldsymbol{u} \rvert \leq d_s$	approximate Fourier coefficients of ANOVA terms $f_{\boldsymbol{u}}$, $\boldsymbol{u} \subset \mathcal{D}$, $\boldsymbol{u} \leq d_s$

Arithmetic cost: $nM \log M + \sum_{i=1}^{d_s} \binom{d}{i} \cdot i \, |I_i| +$ (M eval. of black box function)

Remark 4.5 *The matrix-vector products in line 3 of Algorithm 4.1 are independent of each other and can therefore be computed simultaneously using multi-core parallelization. Furthermore, the first step of the adjoint LFFT is an FFT, see Algorithm 2.2. Since this FFT is always the same, it can be computed outside of the loop which would improve the first term in the arithmetic cost from $nM \log M$ to $M \log M$.*

We keep the nomenclature in [18, Chapter 8] and define the **approximate Fourier sum**

$$\left(S_I^\Lambda \mathcal{T}_{d_s} f\right)(\boldsymbol{x}) \coloneqq \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} \sum_{\boldsymbol{\ell}\in I_{|\boldsymbol{u}|}} \tilde{\hat{f}}_{\boldsymbol{\ell},\boldsymbol{u}} \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{x}_{\boldsymbol{u}}}, \tag{4.8}$$

where $\tilde{\hat{f}}_{\boldsymbol{\ell},\boldsymbol{u}}$ are the Fourier coefficients calculated by Algorithm 4.1. Subsequently we consider the error for approximating f by $S_I^\Lambda \mathcal{T}_{d_s} f$.

Lemma 4.6 *Let $f \in \mathcal{A}_w(\mathbb{T}^d)$ with weight function $w : \mathbb{Z}^d \to [1,\infty)$, $I_i = \{\boldsymbol{k} \in (\mathbb{Z}\setminus\{0\})^i \colon w(\boldsymbol{k}) \le N_i\}$ for $N_i \in \mathbb{N}$, $i = 1,2,\ldots,d_s$ and I the union* (4.1)*. Moreover, let $\Lambda(\boldsymbol{z}, M, I)$ be a reconstructing rank-1 lattice for I. Then*

$$\left\|S_I\mathcal{T}_{d_s} f - S_I^\Lambda \mathcal{T}_{d_s} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \le \frac{1}{\min_{i=1,2,\ldots,d_s} N_i} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.$$

Proof. Similar to the proof of Lemma 4.2, we estimate

$$\begin{aligned}\left\|S_I\mathcal{T}_{d_s} f - S_I^\Lambda \mathcal{T}_{d_s} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)} &= \sup_{\boldsymbol{x}\in\mathbb{T}^d}\left|S_I\mathcal{T}_{d_s} f(\boldsymbol{x}) - S_I^\Lambda \mathcal{T}_{d_s} f(\boldsymbol{x})\right| \\ &\le \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} \left|\sum_{\boldsymbol{\ell}\in I_{|\boldsymbol{u}|}} \left(\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}} - \tilde{\hat{f}}_{\boldsymbol{\ell},\boldsymbol{u}}\right)\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\right|.\end{aligned}$$

Employing (2.6) and then [18, Lemma 8.13], i.e., $\{\boldsymbol{\ell}+\boldsymbol{h}\colon \boldsymbol{\ell}\in I_{|\boldsymbol{u}|}, \boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z}_{\boldsymbol{u}}, M)\setminus\{\boldsymbol{0}\}\} \subset (\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}$, yields

$$\begin{aligned}\left\|S_I\mathcal{T}_{d_s} f - S_I^\Lambda \mathcal{T}_{d_s} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)} &\le \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} \sum_{\boldsymbol{\ell}\in I_{|\boldsymbol{u}|}} \sum_{\boldsymbol{h}\in\Lambda^\perp(\boldsymbol{z}_{\boldsymbol{u}},M)\setminus\{\boldsymbol{0}\}} \left|\hat{f}_{\boldsymbol{\ell}+\boldsymbol{h},\boldsymbol{u}}\right| \\ &\le \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}} \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right| \\ &= \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}} \frac{w_{\boldsymbol{u}}(\boldsymbol{\ell})}{w_{\boldsymbol{u}}(\boldsymbol{\ell})} \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right| \\ &\le \frac{1}{\min_{i=1,2,\ldots,d_s} N_i} \sum_{\substack{\boldsymbol{u}\subset\mathcal{D}\\|\boldsymbol{u}|\le d_s}} \sum_{\boldsymbol{\ell}\in(\mathbb{Z}\setminus\{\boldsymbol{0}\})^{|\boldsymbol{u}|}\setminus I_{|\boldsymbol{u}|}} w_{\boldsymbol{u}}(\boldsymbol{\ell}) \left|\hat{f}_{\boldsymbol{\ell},\boldsymbol{u}}\right| \\ &\le \frac{1}{\min_{i=1,2,\ldots,d_s} N_i} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)}.\end{aligned}$$

■

Theorem 4.7 *Let $f \in \mathcal{A}_w(\mathbb{T}^d)$ with weight function $w : \mathbb{Z}^d \to [1, \infty)$, $I_i = \{\boldsymbol{k} \in (\mathbb{Z} \setminus \{0\})^i \colon w(\boldsymbol{k}) \leq N_i\}$ for $N_i \in \mathbb{N}$, $i = 1, 2, \ldots, d_s$ and I the union (4.1). Moreover, let $\Lambda(\boldsymbol{z}, M, I)$ be a reconstructing rank-1 lattice for I. Then*

$$\left\|f - S_I^\Lambda \mathcal{T}_{d_s} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq \left(\frac{1}{\inf_{\boldsymbol{k} \in \bigcup_{\boldsymbol{u} \subset \mathcal{D}, |\boldsymbol{u}| > d_s} \mathbb{F}_{\boldsymbol{u}}^d} w(\boldsymbol{k})} + \frac{2}{\min_{i=1,2,\ldots,d_s} N_i} \right) \|f\|_{\mathcal{A}_w(\mathbb{T}^d)} .$$

Proof. We apply the triangle inequality and get

$$\left\|f - S_I^\Lambda \mathcal{T}_{d_s} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \leq \|f - S_I \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} + \left\|S_I \mathcal{T}_{d_s} f - S_I^\Lambda \mathcal{T}_{d_s} f\right\|_{\mathrm{L}_\infty(\mathbb{T}^d)}$$

The statement now follows immediately from Theorem 4.3 and Lemma 4.6. ■

Now, we use Algorithm 4.1 to obtain a first approximation of a function f. However, the number of terms we have to consider is $n = \sum_{i=0}^{d_s} \binom{d}{i}$ which has polynomial growth in d for fixed d_s, but is still large in general. This forces us to decrease the size of the index sets I_i, $i = 1, 2, \ldots, d_s$ and therefore the quality of the approximation if we want the algorithm to finish in a reasonable amount of time. However, we only use this first approximation to obtain an **active set**.

Given small index sets I_i, $i = 1, 2, \ldots, d_s$ and the resulting set I as in (4.1), we use Algorithm 4.1 to compute an approximation

$$\tilde{f}_1 \coloneqq S_I^\Lambda \mathcal{T}_{d_s} f, \tag{4.9}$$

see (4.8).

Remark 4.8 *We deliberately do not clearly specify any exact proportion of the small index sets I_i mentioned before. The goal is to find this first approximation in a comparably short time, i.e., the component-by-component search Algorithm 4.2, presented in Section 4.2, has to find a generating vector $\boldsymbol{z}$ fast. This depends on the dimension d of the function, the structure of the index set and the chosen superposition dimension d_s.*

Now, we calculate approximations to the global sensitivity indices $\varrho(\boldsymbol{u}; f)$, see (3.15), by using the first approximation $\tilde{f}_1$. For this purpose we consider the global sensitivity indices of $\tilde{f}_1$

$$\varrho(\boldsymbol{u}; \tilde{f}_1) = \frac{\left\|\tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}}\right\|_2^2}{\left\|\tilde{\hat{\boldsymbol{f}}}\right\|_2^2 - \tilde{\hat{f}}_{\boldsymbol{0}}^2}$$

and assume them to be a good approximation to the global sensitivity indices of f, i.e., $\varrho(\boldsymbol{u}; \tilde{f}_1) \approx \varrho(\boldsymbol{u}; f)$. This can of course only be true for sets $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$ since $\varrho(\boldsymbol{u}; \tilde{f}_1) = 0$ otherwise.

In order to identify the important sets we introduce a **threshold vector** $\boldsymbol{\theta} \in (0,1)^{d_s}$ and the **active set**

$$U(\boldsymbol{\theta}; \tilde{f}_1) \coloneqq \{\boldsymbol{u} \subset \mathcal{D} \colon 1 \leq |\boldsymbol{u}| \leq d_s, \varrho(\boldsymbol{u}; \tilde{f}_1) > \theta_{|\boldsymbol{u}|}\} \cup \{\emptyset\}. \tag{4.10}$$

The active set $U(\boldsymbol{\theta}; \tilde{f}_1)$ contains the sets $\boldsymbol{u}$ such that the terms $f_{\boldsymbol{u}}$ contribute most to the variance $\sigma^2(\tilde{f}_1)$ with respect to the threshold vector $\boldsymbol{\theta}$.

4.2 Lattice Construction

In the previous section we worked with the assumption of having a rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ that satisfies two conditions for a given $d_s < d$

(a) $\Lambda(\boldsymbol{z}_{\boldsymbol{u}}, M, I_{|\boldsymbol{u}|})$ is a reconstructing rank-1 lattice for each $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$

(b) $\boldsymbol{\ell} \cdot \boldsymbol{z}_{\boldsymbol{u}} \neq \boldsymbol{h} \cdot \boldsymbol{z}_{\boldsymbol{v}} \mod M$ for $\boldsymbol{u}, \boldsymbol{v} \subset \mathcal{D}$, $|\boldsymbol{u}|, |\boldsymbol{v}| \leq d_s$ and $\boldsymbol{u} \neq \boldsymbol{v}$ with $\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}$ and $\boldsymbol{h} \in I_{|\boldsymbol{v}|}$,

see Lemma 4.4. Now, we develop an algorithm to find such a rank-1 lattice. First, we can prove that the property to be a reconstructing rank-1 lattice for an index set I as in (4.1) is equivalent to the conditions (a) and (b).

Lemma 4.9 *Let I be an index set as in* (4.1) *and $d_s < d$ the superposition dimension. Then $\Lambda(\boldsymbol{z}, M, I)$ is a reconstructing rank-1 lattice if and only if it satisfies the conditions (a) and (b).*

Proof. The result follows directly from Lemma 4.4 and specifically equation (4.7). ■

The existence of such a reconstructing rank-1 lattice has been discussed in [5, 6] and [18, Chapter 8.4], see especially Theorem 8.16. The component-by-component algorithm proposed therein can be used for construction. This covers the existence of a lattice satisfying both conditions. In the following we use the special structure of the index set I to propose a new version of the component-by-component algorithm that is specifically tuned for our problem and setting.

Algorithm 4.2 computes a vector $\boldsymbol{z}$ satisfying the conditions a) and b) from above by checking them for each component. There are some improvements to increase performance in comparison to a naive check of the conditions which we explain in the following.

The first important observation is that we only need to check the conditions for subsets that involve the component we are currently updating. This means that for each component of $\boldsymbol{z}$ with index $i \in \{2, 3, \dots, d\}$, we have to check the sets $\boldsymbol{u} \subset \{1, 2, \dots, i\}$ with $i \in \boldsymbol{u}$ and $|\boldsymbol{u}| \leq d_s$. Note that the condition (b) has to be checked against all other sets $\boldsymbol{u} \subset \{1, 2, \dots, i\}$. Here, the implementation of the algorithm can be accelerated if one avoids symmetries, i.e., checking the condition for two sets $\boldsymbol{u} \subset \{1, 2, \dots, i\}$ with $i \in \boldsymbol{u}$ and $\boldsymbol{v} \subset \{1, 2, \dots, i\}$ with $i \in \boldsymbol{v}$ twice. This can of course only happen for sets that contain the current index i.

There are multiple reasons why using the ANOVA variant of the component by component search 4.2 is better in our context. First, it is unnecessary to ever construct and save the entire index set I. Since I has such a special structure here, it is completely sufficient to construct and save the smaller index sets I_i, $i = 1, 2, \dots, d_s$.

One could get the idea that increasing the dimension d of the function therefore plays no role if we don't construct the index set which is not the case since we have to check more sets $\boldsymbol{u}$ for each component in return. However, the influence of this effect can be lowered since one can immediately stop to check the conditions as soon as we detect that it is not satisfied for one set.

Algorithm 4.2 ANOVA component-by-component lattice search

Input:

$d \in \mathbb{N}$	spatial dimension of $f : \mathbb{T}^d \to \mathbb{R}$
$d_s \in \mathbb{N}$	superposition dimension with $d_s < d$
$I_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \ldots, d_s$	finite frequency index sets
$M \in \mathbb{N}$	lattice size

$\boldsymbol{z} \leftarrow \mathbf{0}$
$z_1 \leftarrow 1$
for $i = 2, 3, \ldots, d$ **do**
  for $q = 1, 2, \ldots, M-1$ **do**
    $z_i \leftarrow q$
    for $\boldsymbol{u} \subset \{1, 2, \ldots, i\}$ with $i \in \boldsymbol{u}$ and $|\boldsymbol{u}| \leq d_s$ **do**
      if $\Lambda(\boldsymbol{z}_{\boldsymbol{u}}, M, I_{|\boldsymbol{u}|})$ is not a reconstructing rank-1 lattice **then**
        skip to next q
      end if
      for $\boldsymbol{v} \subset \{1, 2, \ldots, i\}$ **do**
        if $\boldsymbol{\ell} \cdot \boldsymbol{z}_{\boldsymbol{u}} \equiv \boldsymbol{h} \cdot \boldsymbol{z}_{\boldsymbol{v}} \mod M$ for any $\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}$ and $\boldsymbol{h} \in I_{|\boldsymbol{v}|}$ **then**
          skip to next q
        end if
      end for
    end for
  end for
end for

Output: $\boldsymbol{z} \in \mathbb{N}^d$ generating vector for rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ satisfying conditions (a) and (b)

Note that it is possible to increase the performance of an implementation even further by storing vectors $(\boldsymbol{k} \cdot \boldsymbol{z}_{\boldsymbol{u}})_{\boldsymbol{k} \in I_{|\boldsymbol{u}|}} \mod M$ if the components of $\boldsymbol{z}_{\boldsymbol{u}}$ have already been determined and therefore won't change in further iterations. The vector can be stored using an integer datatype that is able to hold values in $\{1, 2, \ldots, M-1\}$. The downside of this is of course the additional memory needed.

Remark 4.10 *For the choice of the lattice size $M \in \mathbb{N}$ we refer to [18, Lemma 8.16], i.e., choosing M larger than $|I|^2 - |I| + 1$ yields a generating vector $\boldsymbol{z}$. Note that this result is for general index sets. In special cases it can suffice to choose M much smaller than this bound.*

4.3 Approximation with Active Set

In Section 4.1 we obtained an approximation $\tilde{f}_1$ and defined the active set $U(\boldsymbol{\theta}; \tilde{f}_1)$. Now, we choose a threshold vector $\boldsymbol{\theta}$ and use the resulting set to construct a new approximation. For this purpose we modify the Algorithms 4.2 and 4.1 to work with a given active set of terms.

Algorithm 4.3 ANOVA component-by-component lattice search with active set

Input:	$d \in \mathbb{N}$	spatial dimension of $f : \mathbb{T}^d \to \mathbb{R}$
	$d_s \in \mathbb{N}$	superposition dimension with $d_s < d$
	$U \subset \mathcal{P}(\mathcal{D})$	active set (4.10)
	$J_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \ldots, d_s$	finite frequency index sets
	$M \in \mathbb{N}$	lattice size

$z_1 \leftarrow 1$
for $i = 2, 3, \ldots, d$ **do**
    for $q = 1, 2, \ldots, M-1$ **do**
        $z_i \leftarrow q$
        for $\boldsymbol{u} \in U$ with $\boldsymbol{u} \subset \{1, 2, \ldots, i\}$ and $i \in \boldsymbol{u}$ **do**
            if $\Lambda(\boldsymbol{z}_{\boldsymbol{u}}, M, J_{|\boldsymbol{u}|})$ is not a reconstructing rank-1 lattice **then**
                skip to next q
            end if
            for $\boldsymbol{u} \in U$ with $\boldsymbol{u} \subset \{1, 2, \ldots, i\}$ **do**
                if $\boldsymbol{\ell} \cdot \boldsymbol{z}_{\boldsymbol{u}} \equiv \boldsymbol{h} \cdot \boldsymbol{z}_{\boldsymbol{v}} \mod M$ for any $\boldsymbol{\ell} \in J_{|\boldsymbol{u}|}$ and $\boldsymbol{h} \in J_{|\boldsymbol{v}|}$ **then**
                    skip to next q
                end if
            end for
        end for
    end for
end for

Output: $\boldsymbol{z} \in \mathbb{N}^d$ gen. vector for rank-1 lattice $\Lambda(\boldsymbol{z}, M)$ satisfying a) and b)

We choose new index sets $J_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \ldots, d_s$, that can be significantly larger than before because the number of elements in $U(\boldsymbol{\theta}; \tilde{f}_1)$ will be a lot smaller than considering all terms of order up to d_s. A generating vector $\boldsymbol{z}$ can be found using Algorithm 4.3 and the second approximation is given by

$$\tilde{f}_2(\boldsymbol{x}) := \sum_{\boldsymbol{u} \in U(\boldsymbol{\theta}; \tilde{f}_1)} \sum_{\boldsymbol{\ell} \in J_{|\boldsymbol{u}|}} \hat{\tilde{f}}_{\boldsymbol{\ell}, \boldsymbol{u}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}} \tag{4.11}$$

with $J = \bigcup_{\boldsymbol{u} \in U} J_{|\boldsymbol{u}|}^{(d)}$, see (4.1), using Algorithm 4.4.

Remark 4.11 *It is possible to reduce the lattice size M by taking the resulting vector $\boldsymbol{z}$ and checking for each number from $|J|$ to $M-1$ if the two conditions (a) and (b) are satisfied.*

Algorithm 4.4 Function reconstruction over special rank-1 lattice with active set

Input:	$d \in \mathbb{N}$	spatial dimension of $f : \mathbb{T}^d \to \mathbb{R}$
	$d_s \in \mathbb{N}$	superposition dimension with $d_s < d$
	$U \subset \mathcal{P}(\mathcal{D})$	active set (4.10)
	$J_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \dots, d_s$	finite frequency index sets
	$\boldsymbol{z} \in \mathbb{Z}^d$	generating vector satisfying the conditions in Lemma 4.4
	$M \in \mathbb{N}$	lattice size
	f	black box function

1: $\boldsymbol{f} \leftarrow \left(f(\frac{1}{M}(j\boldsymbol{z} \mod M))\right)_{j=0}^{M-1}$
2: **for** $\boldsymbol{u} \in U$ **do**
3: $\quad \tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}} \leftarrow \frac{1}{M} \boldsymbol{F}_{\boldsymbol{u}}^H \boldsymbol{f}$ $\qquad \triangleleft$ $|\boldsymbol{u}|$-variate adjoint LFFT, see Algorithm 2.2
4: **end for**

Output: $\tilde{\hat{\boldsymbol{f}}}_{\boldsymbol{u}} \in \mathbb{C}^{\left|J_{|\boldsymbol{u}|}\right|}, \boldsymbol{u} \in U$ approximate Fourier coefficients of ANOVA term $f_{\boldsymbol{u}}$
Arithmetic cost: $|U|\, M \log M + \sum_{\boldsymbol{u} \in U} |\boldsymbol{u}| \left|J_{|\boldsymbol{u}|}\right| + M$ eval. of black box function

This reduces the function evaluations in return. One has to weigh if the lattice reduction makes sense since it is costly in arithmetic operations and doing more evaluations might be cheaper.

Remark 4.12 *Algorithms 4.3 and 4.4 are generalized versions of the Algorithms 4.2 and 4.1 respectively. Choosing $\boldsymbol{U} = \{\boldsymbol{u} \subset \{1, 2, \dots, d\} : |\boldsymbol{u}| \le d_s\}$ in the former algorithms yields the same behavior as the latter ones. Furthermore, the first term in the arithmetic cost can be proved from $|U|\, M \log M$ to $M \log M$ analogously to the way described in Remark 4.5*

4.4 Numerical Results

We now apply the previously presented algorithms to an example function of a class that has already been considered in [19]. It is a sum of products of univariate functions

$$B_r\colon \mathbb{T} \to \mathbb{R}, x \mapsto B_r(x) := C_r \sum_{k\in\mathbb{Z}} \operatorname{sinc}\left(\frac{\pi}{r}k\right)^r \cos(\pi k)\mathrm{e}^{2\pi\mathrm{i}kx}. \tag{4.12}$$

B_r is a shifted, scaled and dilated B-spline of order r and $C_r > 0$ such that $\|B_r\|_{\mathrm{L}_2(\mathbb{T})} = 1$. We denote the Fourier-coefficients of B_r with

$$\hat{b}_k^{(r)} = C_r \operatorname{sinc}\left(\frac{\pi}{r}k\right)^r \cos(\pi k).$$

In the following, we consider the specific 9-dimensional example

$$f(\boldsymbol{x}) = \prod_{i\in\{1,3,8\}} B_2(x_i) + \prod_{i\in\{2,5,6\}} B_4(x_i) + \prod_{i\in\{4,7,9\}} B_6(x_i) \tag{4.13}$$

with Fourier coefficients

$$\hat{f}_{\boldsymbol{k}} = \delta_{\boldsymbol{k}_{\mathcal{D}\setminus\{1,3,8\}},\boldsymbol{0}} \prod_{i\in\{1,3,8\}} \hat{b}_{k_i}^{(2)} + \delta_{\boldsymbol{k}_{\mathcal{D}\setminus\{2,5,6\}},\boldsymbol{0}} \prod_{i\in\{2,5,6\}} \hat{b}_{k_i}^{(4)} + \delta_{\boldsymbol{k}_{\mathcal{D}\setminus\{4,7,9\}},\boldsymbol{0}} \prod_{i\in\{4,7,9\}} \hat{b}_{k_i}^{(6)}. \tag{4.14}$$

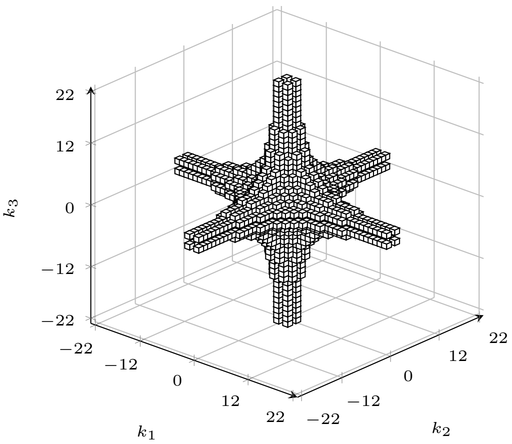
Furthermore, for the norm of f we get

$$\|f\|^2_{\mathrm{L}_2(\mathbb{T}^9)} = 3 + \sum_{i=1}^{2} \sum_{j=i+1}^{3} 2(\hat{b}_0^{(2i)})^3(\hat{b}_0^{(2j)})^3.$$

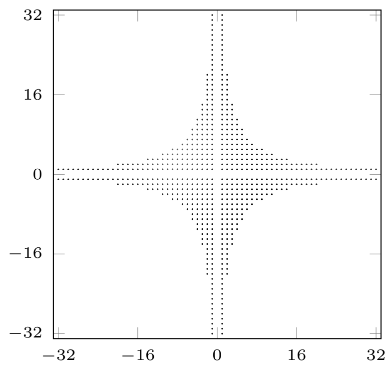
From (4.14) it immediately follows that $\hat{f}_{\boldsymbol{k}}$ is 0 for $\|\boldsymbol{k}\|_0 > 3$ which is why we choose $d_s = 3$. For the three finite index sets I_1, I_2 and I_3 we use

$$I_i = \left\{\boldsymbol{k} \in (\mathbb{Z}\setminus\{0\})^i : w_{s,\mathrm{mix}}^{(i)}(\boldsymbol{k}) \leq N_i\right\}, i = 1, 2, 3,$$

with a vector $\boldsymbol{N} \in \mathbb{N}^3$ and the weight functions $w_{s,\mathrm{mix}}^{(i)}(\boldsymbol{k}) = \prod_{j=1}^{i}(1 + |k_j|^2)^s$. Since $f \in \mathrm{H}_{\mathrm{mix}}^{3/2-\varepsilon}(\mathbb{T}^d)$, $\varepsilon > 0$, we take $s = \frac{3}{2}$. Figure 4.1 shows examples of the two- and three-dimensional index sets I_2 and I_3 for our chosen weight function $w_{3/2,\mathrm{mix}}^{(i)}(\boldsymbol{k})$.



(a) I_3 with $N = 10^5$



(b) I_2 with $N = 10^5$

Figure 4.1: The index sets I_3 and I_2 with weight function $w^{(i)}_{3/2,\mathrm{mix}}(\boldsymbol{k})$.

All calculations are performed on a computer with 4 Intel Xeon E5-4640 2.40GHz 8-core. We used 30 parallel workers for multi-core parallelization of the component-by-component algorithm and the calculation of the Fourier coefficients.

Active Set Construction

We proceed by choosing an $\boldsymbol{N} \in \mathbb{N}^3$ as cutoff vector to generate the index sets I_1, I_2 and I_3. Applying Algorithm 4.2 with an M yields the generating vector $\boldsymbol{z}$. Using Algorithm 4.1 we construct the first approximation $\tilde{f}_1$, see (4.9), and consider the errors

$$\varepsilon_{L_2} = \frac{\left\| f - \tilde{f}_1 \right\|_{\mathrm{L}_2(\mathbb{T}^9)}}{\|f\|_{\mathrm{L}_2(\mathbb{T}^9)}} \quad \text{and} \quad \varepsilon_2 = \frac{\left\| \boldsymbol{f} - \tilde{\boldsymbol{f}}_1 \right\|_2}{\|\boldsymbol{f}\|_2}$$

with $\tilde{\boldsymbol{f}}_1 = (\tilde{f}_1(\boldsymbol{x}))_{\boldsymbol{x}\in\Lambda(\boldsymbol{z},M)}$. Note that our goal in this step is to identify the important ANOVA terms. In the case of the specific function f from (4.13) that means all sets $\boldsymbol{u} \subset \mathcal{D}$ such that $\boldsymbol{u} \subset \{1,3,8\}$, $\boldsymbol{u} \subset \{2,5,6\}$ or $\boldsymbol{u} \subset \{4,7,9\}$ give a corresponding ANOVA term, i.e., 22 terms in total. This follows immediately from the formula for the Fourier coefficients (4.14) and the definition of the sets $\mathbb{F}^{(d)}_{\boldsymbol{u}}$ in (3.7).

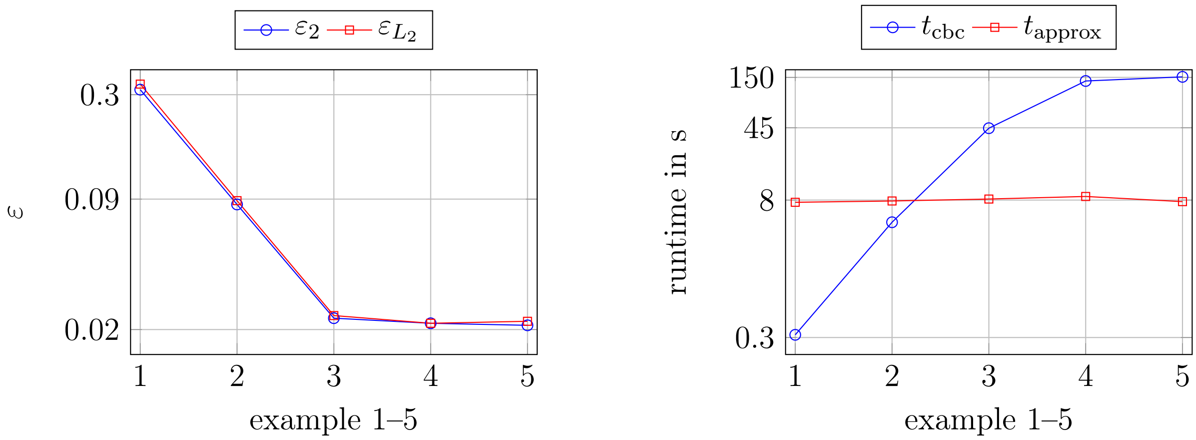
In Table 4.1 we present the results of the active set construction step for a variety of vectors $\boldsymbol{N}$ with fixed $M = 10^6 + 3$. The last column contains one threshold vector $\boldsymbol{\theta} \in (0,1)^3$ for every $\boldsymbol{N}$ that allowed us to identify the active ANOVA terms. Furthermore, t_{cbc} is the runtime for Algorithm 4.2 and t_{approx} is the runtime of Algorithm 4.1.

For each example vector $\boldsymbol{N}$ except the smallest $[10, 10, 10]$ we were able to find a $\boldsymbol{\theta}$ that allowed us to identify the ANOVA terms of f that are nonzero. Even for the small index sets generated by $[10^2, 10^2, 10^2]$, we achieve this goal. In Figure 4.2 we have displayed the errors and runtimes for the examples 1–5 from Table 4.1.

Now, we calculate the global sensitivity indices $\varrho(\boldsymbol{u}; \tilde{f}_1)$ and use them and the vector $\boldsymbol{\theta}$ from Table 4.1 to threshold the sets $\boldsymbol{u}$. Figure 4.3 shows the behavior of the global

ex.	$\boldsymbol{N}$	t_{cbc}	t_{approx}	ε_2	ε_{L_2}	$\boldsymbol{\theta}$
1	$[10, 10, 10]$	0.32s	7.59s	0.32	0.34	$[0.04, 0.02, \sim]$
2	$[10^2, 10^2, 10^2]$	4.71s	7.83s	$8.46 \cdot 10^{-2}$	$8.84 \cdot 10^{-2}$	$[0.04, 0.01, 0.006]$
3	$[10^3, 10^3, 10^3]$	44.55s	8.2s	$2.28 \cdot 10^{-2}$	$2.35 \cdot 10^{-2}$	$[0.04, 0.01, 0.006]$
4	$[10^4, 10^4, 10^3]$	137.60s	8.72s	$2.15 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$[0.04, 0.01, 0.006]$
5	$[10^5, 10^4, 10^3]$	151.88s	7.71s	$2.10 \cdot 10^{-2}$	$2.20 \cdot 10^{-2}$	$[0.04, 0.01, 0.006]$

Table 4.1: Active set construction step for examples (ex.) 1–5 with different cutoff vectors $\boldsymbol{N}$. Each runtime is the mean of three runs.



(a) errors ε_{L_2} and ε_2

(b) runtimes t_{cbc} and t_{approx}

Figure 4.2: Errors ε_{L_2} and ε_2 with runtimes for Algorithms 4.2 and 4.1 for examples from Table 4.1.

sensitivity indices $\varrho(\boldsymbol{u}; \tilde{f}_1)$. The gap between the 21 relevant terms and the other terms is crucial for the choice of $\boldsymbol{\theta}$. We observe that the gap widens with increasing size of the index sets and the $\varrho(\boldsymbol{u}; \tilde{f}_1)$ are smaller for the irrelevant terms.

Approximation with Active Set

We proceed by choosing new index sets

$$J_i = \left\{ \boldsymbol{k} \in (\mathbb{Z} \setminus \{0\})^i : w_{s,\text{mix}}^{(i)}(\boldsymbol{k}) \leq N_i \right\}, i = 1, 2, 3,$$

with a larger cutoff vector $\boldsymbol{N}$. We then use the modified algorithms from Section 4.3 with an active set to obtain the next approximation $\tilde{f}_2$, see (4.11). Table 4.2 contains the results for different cutoff vectors $\boldsymbol{N}$ and fixed $M = 10^7$ while Figure 4.4 displays the times and errors. Observe that we are able to reach an approximation error of about $2.9 \cdot 10^{-4}$ in reasonable time. Furthermore, both errors are behaving similarly and are always close for the same vector $\boldsymbol{N}$.

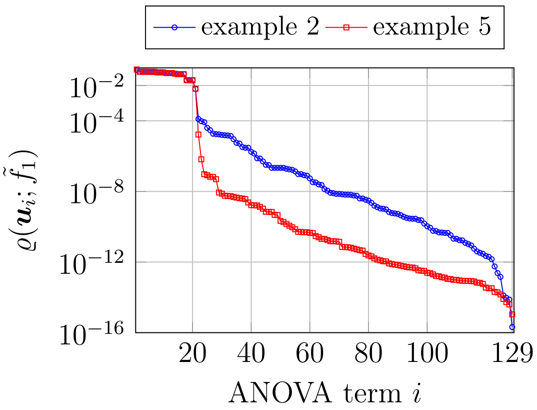
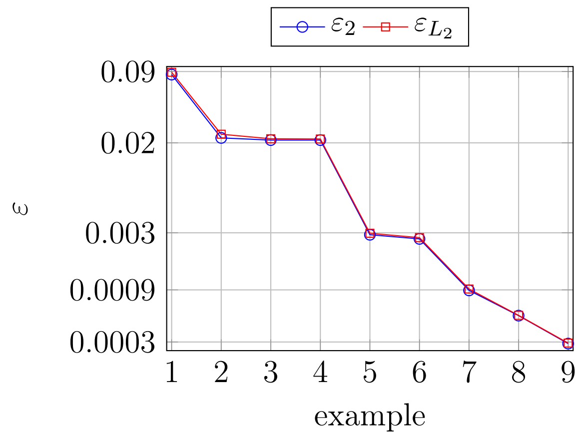
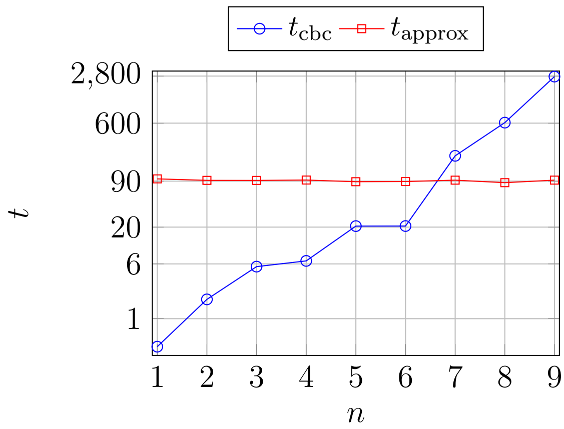


Figure 4.3: Behavior of the global sensitivity indices $\varrho(\boldsymbol{u}; \tilde{f}_1)$ for the examples 2 and 5 from Table 4.1.



(a) errors ε_{L_2} and ε_2



(b) runtimes t_{cbc} and t_{approx}

Figure 4.4: Errors ε_{L_2} and ε_2 with runtimes for Algorithms 4.2 and 4.1 for examples from Table 4.2.

Summarizing, we were able to reach the goal of detecting the significant dimension interactions, i.e., the ANOVA terms, for the example function f from (4.13) in a short amount of time with relatively small index sets. Using these results, we improved the approximation by only taking these terms into account which allowed us to find a good quality approximation.

A similar function to f from (4.13) was considered in [19, Section 3.3]. While a direct comparison is not possible because the function there is 10-dimensional, there is a noticeable large difference in the number of required function evaluations. We achieved a relative l_2-error of about $2.9 \cdot 10^{-4}$ with 10^7 samples (without lattice size decreasing) in comparison to the best approach in [19] with l_2-errors of $5.1 \cdot 10^{-4}$ with 20968000 samples and $1.8 \cdot 10^{-4}$ with 73500131 samples.

ex.	$\boldsymbol{N}$	t_{cbc}	t_{approx}	ε_2	ε_{L_2}
1	$[10^2, 10^2, 10^2]$	0.40s	96.87s	$8.44 \cdot 10^{-2}$	$8.86 \cdot 10^{-2}$
2	$[10^3, 10^3, 10^3]$	1.88s	92.32s	$2.22 \cdot 10^{-2}$	$2.40 \cdot 10^{-2}$
3	$[10^4, 10^4, 10^3]$	5.49s	92.16s	$2.12 \cdot 10^{-2}$	$2.18 \cdot 10^{-2}$
4	$[10^5, 10^4, 10^3]$	6.62s	93.27s	$2.12 \cdot 10^{-2}$	$2.17 \cdot 10^{-2}$
5	$[10^4, 10^4, 10^4]$	20.67s	88.34s	$2.88 \cdot 10^{-3}$	$2.97 \cdot 10^{-3}$
6	$[10^5, 10^4, 10^4]$	20.71s	88.88s	$2.64 \cdot 10^{-3}$	$2.71 \cdot 10^{-3}$
7	$[10^5, 10^5, 10^5]$	205.98s	92.92s	$8.91 \cdot 10^{-4}$	$9.17 \cdot 10^{-4}$
8	$[10^6, 10^6, 10^5]$	610.30s	85.58s	$5.23 \cdot 10^{-4}$	$5.26 \cdot 10^{-4}$
9	$[10^6, 10^6, 10^6]$	2753.41s	93.03s	$2.90 \cdot 10^{-4}$	$2.93 \cdot 10^{-4}$

Table 4.2: Obtaining the approximation $\tilde{f}_2$, see (4.11), for different cutoff vectors $\boldsymbol{N}$. Each time is the mean of three runs.

5 Approximation with Scattered Data

In this section we consider a modified version of the problem from Section 4.

Problem 5.1 *Let $f : \mathbb{T}^d \to \mathbb{R}$ be a function in $C(\mathbb{T}^d) \subset \mathrm{L}_2(\mathbb{T}^d)$ with $d \in \mathbb{N}$. We have a set of sampling nodes $X \subset \mathbb{T}^d$ with $|X| = M$, $M \in \mathbb{N}$, and samples of f at those points $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x} \in X}$. Furthermore, a superposition dimension $d_s \in \mathbb{N}$ with $d_s \leq d$ is given.*
We want to find an approximation for f based on the approximate ANOVA model $\mathcal{T}_{d_s} f$. Furthermore, we are looking for important dimension interactions, i.e., the ANOVA terms that contribute significantly to f, in other words, sets $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$ whose global sensitivity index $\varrho(\boldsymbol{u}; f)$ is large.

Problem 5.1 is of great significance in practical applications. While we were able to achieve good results for Problem 4.1, the ability to evaluate a function at any point in $\mathbb{T}^d$ is not present in many applications. This might be the case because evaluations are expensive or because the way that the data is obtained provides us only with a certain range of data.

As mentioned before, this is a very significant and active topic, in research and application alike. While state-of-the-art deep learning approaches provide us with a powerful tool to tackle such problems, a large part of the mathematical background has yet to be explored.

The approach presented in the following section rests on the ANOVA decomposition on the torus and the NFFT algorithm [10]. The basic idea is to solve a least-squares system, similar to the one for Problem 4.1.

5.1 Approximation Scheme

We approach Problem 5.1 as before by first choosing a superposition dimension d_s and using the approximate ANOVA model $\mathcal{T}_{d_s} f$. This will be used to identify important ANOVA terms.

I. Active Set Construction

First, we consider a finite index $I \subset \mathbb{Z}^d$ and the Fourier partial sum $S_I \mathcal{T}_{d_s} f$, respectively. We want to find approximations for the global sensitivity indices $\varrho(\boldsymbol{u}; f)$. To this end, the ANOVA terms $f_{\boldsymbol{u}}$ of the same order have to be supported on isomorphic low-dimensional index sets as before, meaning we choose index sets $I_1 \subset \mathbb{Z}, I_2 \subset \mathbb{Z}^2, \ldots, I_{d_s} \subset \mathbb{Z}^{d_s}$. Then we have a disjoint union

$$I = \bigcup_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} I_{\boldsymbol{u}}^{(d)} \tag{5.1}$$

with $I_{\boldsymbol{u}}^{(d)} = \{\boldsymbol{k} \in \mathbb{Z}^d \colon \boldsymbol{k}_{\boldsymbol{u}} \in I_{|\boldsymbol{u}|}, \boldsymbol{k}_{\mathcal{D}\setminus\boldsymbol{u}} = \boldsymbol{0}\}$. Instead of choosing the index sets with an arbitrary weight function, we use specific index sets

$$I_i = \{\boldsymbol{k} \in \mathbb{Z}^i \colon -N_i/2 \le k_j \le N_i/2 - 1, j = 1, 2, \dots, i\}, i = 1, 2, \dots, d_s,$$

i.e., full grids with cutoff vector $\boldsymbol{N} \in \mathbb{N}^{d_s}$.

Using Lemma 4.2 we have

$$\|f - S_I \mathcal{T}_{d_s} f\|_{\mathrm{L}_\infty(\mathbb{T}^d)} \le \frac{1}{\min_{i=1,2,\dots,d_s}(N_i/2 - 1)} \|f\|_{\mathcal{A}_w(\mathbb{T}^d)}$$

as an estimation for the cutoff error.

Contrary to Section 4, we use equation (3.17), i.e.,

$$\mathcal{T}_{d_s} f = \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \le d_s}} c(|\boldsymbol{u}|, d, d_s) P_{\boldsymbol{u}} f,$$

as a basis for our approach. This means that we are working with the projections $P_{\boldsymbol{u}} f$ and therefore index sets including zeros. We approximate the projections $P_{\boldsymbol{u}} f$ by a trigonometric polynomial supported on $I_{|\boldsymbol{u}|}$ yielding

$$f(\boldsymbol{x}) \approx S_I \mathcal{T}_{d_s} f(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \le d_s}} c(|\boldsymbol{u}|, d, d_s) \sum_{\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}} \hat{p}_{\boldsymbol{\ell},\boldsymbol{u}} \, \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{x}_{\boldsymbol{u}}}.$$

Taking the given function evaluations $\boldsymbol{y} \in \mathbb{R}^M$, see Problem 5.1, we use matrix-vector products to obtain

$$\boldsymbol{y} = \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \le d_s}} c(|\boldsymbol{u}|, d, d_s) \boldsymbol{P}_{\boldsymbol{u}} \hat{\boldsymbol{p}}_{\boldsymbol{u}} \tag{5.2}$$

with Fourier matrices

$$\boldsymbol{P}_{\boldsymbol{u}} = \left(\mathrm{e}^{2\pi\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{x}_{\boldsymbol{u}}}\right)_{\boldsymbol{x} \in X, \boldsymbol{\ell} \in I_{|\boldsymbol{u}|}}$$

and coefficient vectors $\hat{\boldsymbol{p}}_{\boldsymbol{u}} = (\hat{p}_{\boldsymbol{\ell},\boldsymbol{u}})_{\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}}$. Now, we write (5.2) as a block matrix times a vector

$$\boldsymbol{y} = \begin{pmatrix} c_{|\boldsymbol{u}_1|} \boldsymbol{P}_{\boldsymbol{u}_1} & c_{|\boldsymbol{u}_2|} \boldsymbol{P}_{\boldsymbol{u}_2} & \cdots & c_{|\boldsymbol{u}_n|} \boldsymbol{P}_{\boldsymbol{u}_n} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{p}}_{\boldsymbol{u}_1} \\ \hat{\boldsymbol{p}}_{\boldsymbol{u}_2} \\ \vdots \\ \hat{\boldsymbol{p}}_{\boldsymbol{u}_n} \end{pmatrix} =: \boldsymbol{P}\hat{\boldsymbol{p}} \tag{5.3}$$

with $c_{|\boldsymbol{u}|} \coloneqq c(|\boldsymbol{u}|, d, d_s)$. Here, $\boldsymbol{u}_j$ for $j = 1, 2, \dots, n$ with

$$n = \sum_{i=0}^{d_s} \binom{d}{i}$$

has to be an ordering for the ANOVA terms.

We find an approximate solution for the Fourier coefficients $\hat{\boldsymbol{p}}$ as a solution of the least-squares problem

$$\tilde{\hat{\boldsymbol{p}}} = \underset{\hat{\boldsymbol{p}} \in \mathbb{C}^{1+\sum_{i=1}^{d_s} N_i^i \binom{d}{i}}}{\arg\min} \|\boldsymbol{y} - \boldsymbol{P}\hat{\boldsymbol{p}}\|_2^2 . \tag{5.4}$$

We solve the problem using a matrix-free variant of any iterative least-squares solve which requires us to be able to multiply with $\boldsymbol{P}$ and $\boldsymbol{P}^H$ efficiently. This can be achieved using the NFFT as denoted in Algorithm 5.1 for the product and Algorithm 5.2 for the product with the adjoint matrix.

Algorithm 5.1 Multiplication of a vector $\hat{\boldsymbol{p}}$ with block matrix $\boldsymbol{P}$, see (5.3)

Input:	$d \in \mathbb{N}$	spatial dimension of $f : \mathbb{T}^d \to \mathbb{R}$
	$d_s \in \mathbb{N}$	superposition dimension with $d_s < d$
	$I_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \dots, d_s$	finite frequency index sets
	$X \subset \mathbb{T}^d$	sampling node set with $\lvert X\rvert = M \in \mathbb{N}$
	$\hat{\boldsymbol{p}} \in \mathbb{C}^{1+\sum_{i=1}^{d_s} N_i^i \binom{d}{i}}$	coefficient vector, see (5.3)

1: $\boldsymbol{p} \leftarrow [0, 0, \dots, 0]$
2: **for** $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \le d_s$ **do**
3: $\quad \boldsymbol{p} \leftarrow \boldsymbol{p} + c_{|\boldsymbol{u}|} \boldsymbol{P}_{\boldsymbol{u}} \hat{\boldsymbol{p}}_{\boldsymbol{u}}$ $\quad \triangleleft$ $|\boldsymbol{u}|$-variate NFFT, see [18, Algorithm 7.1]
4: **end for**

Output: $\boldsymbol{p} \in \mathbb{C}^M$ result of multiplication $\boldsymbol{p} = \boldsymbol{P}\hat{\boldsymbol{p}}$
Arithmetic cost: $\sum_{i=1}^{d_s} \binom{d}{i} (N_i^i \log N_i + \text{const.} \cdot M)$

Algorithm 5.2 Multiplication of a vector $\boldsymbol{p}$ with adjoint block matrix $\boldsymbol{P}^H$

Input:	$d \in \mathbb{N}$	spatial dimension of $f : \mathbb{T}^d \to \mathbb{R}$
	$d_s \in \mathbb{N}$	superposition dimension with $d_s < d$
	$I_i \in (\mathbb{Z} \setminus \{0\})^i$, $i = 1, 2, \dots, d_s$	finite frequency index sets
	$X \subset \mathbb{T}^d$	sampling node set with $\lvert X\rvert = M \in \mathbb{N}$
	$\boldsymbol{p} \in \mathbb{C}^M$	vector to be multiplied with

1: **for** $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \le d_s$ **do**
2: $\quad \hat{\boldsymbol{p}}_{\boldsymbol{u}} \leftarrow c_{|\boldsymbol{u}|} \boldsymbol{P}_{\boldsymbol{u}}^H \boldsymbol{p}$ $\quad \triangleleft$ $|\boldsymbol{u}|$-variate adjoint NFFT, see [18, Algorithm 7.3]
3: **end for**
4: $\hat{\boldsymbol{p}} \leftarrow [\hat{\boldsymbol{p}}_{\boldsymbol{u}_1}, \hat{\boldsymbol{p}}_{\boldsymbol{u}_2}, \dots, \hat{\boldsymbol{p}}_{\boldsymbol{u}_n}]$

Output: $\hat{\boldsymbol{p}} \in \mathbb{C}^{1+\sum_{i=1}^{d_s} N_i^i \binom{d}{i}}$ result of multiplication $\hat{\boldsymbol{p}} = \boldsymbol{P}^H \boldsymbol{p}$
Arithmetic cost: $\sum_{i=1}^{d_s} \binom{d}{i} (N_i^i \log N_i + \text{const.} \cdot M)$

Remark 5.2 *The least-squares problem in* (5.4) *only has a unique solution if the matrix* $\boldsymbol{P}$ *has full rank. A necessary condition is therefore that we have more sampling nodes than*

Fourier coefficients, i.e., $M > 1 + \sum_{i=1}^{d_s} N_i$. *Using an oversampling factor* $\sigma > 1$ *we write*

$$M = \sigma \left(1 + \sum_{i=1}^{d_s} \binom{d}{i} (N_i)^i \right).$$

The grid, i.e., the N_i, *has to be chosen such that* σ *is large enough and we may assume the matrix* $\boldsymbol{P}$ *to have full rank.*

We use the solution in (5.4) as an approximation for f and define

$$\tilde{f}_1(\boldsymbol{x}) \coloneqq S_I^X \mathcal{T}_{d_s} f(\boldsymbol{x}) \coloneqq \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} c(|\boldsymbol{u}|, d, d_s) \sum_{\boldsymbol{\ell} \in I_{|\boldsymbol{u}|}} \tilde{\hat{p}}_{\boldsymbol{\ell},\boldsymbol{u}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}}. \tag{5.5}$$

Now, we calculate approximations to the global sensitivity indices $\varrho(\boldsymbol{u}; f)$, see (3.15), using the global sensitivity indices $\varrho(\boldsymbol{u}; \tilde{f}_1)$ of $\tilde{f}_1$. We assume that they are a good approximation, i.e., $\varrho(\boldsymbol{u}; \tilde{f}_1) \approx \varrho(\boldsymbol{u}; f)$ for the sets $\boldsymbol{u} \subset \mathcal{D}$ with $|\boldsymbol{u}| \leq d_s$ since $\varrho(\boldsymbol{u}; \tilde{f}_1) = 0$ otherwise.

As in Section 4 we use a **threshold vector** $\boldsymbol{\theta} \in (0,1)^{d_s}$ and the **active set**

$$U(\boldsymbol{\theta}; \tilde{f}_1) = \{\boldsymbol{u} \subset \mathcal{D} \colon 1 \leq |\boldsymbol{u}| \leq d_s, \varrho(\boldsymbol{u}; \tilde{f}_1) > \theta_{|\boldsymbol{u}|}\} \cup \{\emptyset\}.$$

The active set $U(\boldsymbol{\theta}; \tilde{f}_1)$ contains the sets $\boldsymbol{u}$ such that the terms $f_{\boldsymbol{u}}$ contribute most to the variance $\sigma^2(\tilde{f}_1)$ with respect to the threshold vector $\boldsymbol{\theta}$. Since we are working with projections, the condition

$$\boldsymbol{u} \in U(\boldsymbol{\theta}; \tilde{f}_1), \boldsymbol{v} \subset \boldsymbol{u} \Longrightarrow \boldsymbol{v} \in U(\boldsymbol{\theta}; \tilde{f}_1)$$

has to be satisfied in order to be able to determine the approximation of $f_{\boldsymbol{u}}$.

II. Approximation with Active Set

In order to obtain a better approximation, we choose larger index sets than before, i.e., $J_i = \{\boldsymbol{k} \in \mathbb{Z}^i \colon -N_i/2 \leq k_j \leq N_i/2 - 1, j = 1, 2, \ldots, i\}, i = 1, 2, \ldots, d_s$, with a larger cutoff vector $\boldsymbol{N} = [N_1, N_2, \ldots, N_{d_s}] \in \mathbb{N}^{d_s}$. The full index set J is formed as I in (5.1).

Using the active set $U(\boldsymbol{\theta}; \tilde{f}_1)$ we consider the new least-squares problem

$$\tilde{\hat{\boldsymbol{p}}} = \operatorname*{arg\,min}_{\hat{\boldsymbol{p}} \in \mathbb{C}^{1+\sum_{\boldsymbol{u} \in U(\boldsymbol{\theta}; \tilde{f}_1) \setminus \{\emptyset\}} N_{|\boldsymbol{u}|}^{|\boldsymbol{u}|}}} \left\| \boldsymbol{y} - \tilde{\boldsymbol{P}} \hat{\boldsymbol{p}} \right\|_2^2 \tag{5.6}$$

with the Fourier block matrix

$$\tilde{\boldsymbol{P}} = \left(c_{|\boldsymbol{u}_1|} \tilde{\boldsymbol{P}}_{\boldsymbol{u}_1} \; c_{|\boldsymbol{u}_2|} \tilde{\boldsymbol{P}}_{\boldsymbol{u}_2} \; \cdots \; c_{|\boldsymbol{u}_m|} \tilde{\boldsymbol{P}}_{\boldsymbol{u}_m} \right).$$

and the sets $\boldsymbol{u}_j \in U$, $j = 1, 2, \ldots, m$ with $m \coloneqq |U|$. The matrices are defined as

$$\tilde{\boldsymbol{P}}_{\boldsymbol{u}} = \left(\mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}} \right)_{\boldsymbol{x} \in X, \boldsymbol{\ell} \in J_{|\boldsymbol{u}|}}.$$

Again, we use a matrix-free variant of an iterative least-squares solver and slightly modify Algorithms 5.1 and 5.2 by adjusting the for-loop to just go over the sets in U. We obtain a new approximation

$$\tilde{f}_2(\boldsymbol{x}) \coloneqq \sum_{\boldsymbol{u} \in U(\boldsymbol{\theta}; \tilde{f}_1)} c(|\boldsymbol{u}|, d, d_s) \sum_{\boldsymbol{\ell} \in J_{|\boldsymbol{u}|}} \tilde{\hat{p}}_{\boldsymbol{\ell},\boldsymbol{u}} \, \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{\ell} \cdot \boldsymbol{x}_{\boldsymbol{u}}}. \tag{5.7}$$

5.2 Numerical Results

Now, we apply the approximation scheme from Section 5.1 to a test function. Let B_r be the univariate, shifted, scaled and dilated B-spline from (4.12). We consider the 9-dimensional function

$$f(\boldsymbol{x}) = \prod_{i\in\{1,3,8\}} B_2(x_i) + \prod_{i\in\{2,5,6\}} B_4(x_i) + \prod_{i\in\{4,7,9\}} B_6(x_i)$$

as in Section 4.4. Since we are working with scattered data, we assume to have a set $X \subset \mathbb{T}^d$ of $|X| = M = 5 \cdot 10^6$ sampling nodes and access to the corresponding evaluations of f.

The nonzero ANOVA terms of f are those terms $f_{\boldsymbol{u}}$ with $\boldsymbol{u} \subset \{1,3,8\}, \boldsymbol{u} \subset \{2,5,6\}$ or $\boldsymbol{u} \subset \{4,7,9\}$ - a total of 22 terms. This implies that $d_s = 3$ is a good choice for the superposition dimension.

All calculations were performed on the Bull HPC-Cluster Taurus at the ZIH of the TU Dresden utilizing 25 cores in parallel.

Active Set Construction

We have the given vector of function evaluations $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in X}$ and choose the cutoff vector $\boldsymbol{N} \in \mathbb{N}^3$ for the full grids I_1, I_2 and I_3. We then solve the least-squares problem (5.4) to obtain a coefficient vector $\tilde{\boldsymbol{p}}$ and the approximation $\tilde{f}_1$, see (5.5), and consider the error on the sampling nodes

$$\varepsilon_2 = \frac{\left\|\boldsymbol{f} - \tilde{\boldsymbol{f}}_1\right\|_2}{\|\boldsymbol{f}\|_2}.$$

Furthermore, we take the L_2 error

$$\varepsilon_{L_2} = \frac{\left\|f - \tilde{f}_1\right\|_{\mathrm{L}_2(\mathbb{T}^9)}}{\|f\|_{\mathrm{L}_2(\mathbb{T}^9)}}$$

as a representation of the generalization error. Note that one has to calculate the Fourier coefficients of $\tilde{f}_1$, see (5.5), according to (3.20).

ex.	$\boldsymbol{N}$	t_{approx}	ε_2	ε_{L_2}	$\boldsymbol{\theta}$
1	$[512, 16, 16]$	16832.04s	$1.43 \cdot 10^{-2}$	$1.56 \cdot 10^{-2}$	$[0.04, 0.01, 0.006]$
2	$[2048, 32, 16]$	17102.29s	$5.24 \cdot 10^{-2}$	$5.71 \cdot 10^{-2}$	$[0.04, 0.01, 0.005]$
3	$[4096, 32, 32]$	19168.53s	$8.05 \cdot 10^{-2}$	0.13	$[0.04, 0.01, 0.003]$

Table 5.1: Construction of the active set for different cutoff vectors $\boldsymbol{N}$ with corresponding runtime, errors and one possible threshold vector $\boldsymbol{\theta}$. Each runtime is the mean of three runs.

Table 5.1 contains the results for the active set construction step with t_{approx} being the time to solve the least squares problem (5.4). We used the LSQR solver from the Julia package IterativeSolvers with an iteration limit of 20 iterations and 20 parallel workers. Note that it is possible for each example to find a threshold vector $\boldsymbol{\theta}$ such that the active set $U(\boldsymbol{\theta}; \tilde{f}_1)$ coincides with the nonzero terms of the test function. Table 5.1 contains one example for $\boldsymbol{\theta}$ while Figure 5.1 shows the global sensitivity indices $\varrho(\boldsymbol{u}_i; \tilde{f}_1)$ of the 129 ANOVA terms. The gap between the important terms, i.e., the terms which are actually nonzero for the test function, and the other terms is shrinking with increasing cardinality of the index sets. This behavior can be explained by the fact that we have an increasing number of Fourier coefficients for the same number M of samples, i.e., the oversampling parameter σ gets smaller. The gap itself represents the interval of choice for the threshold parameter (with regard to the dimension of the corresponding ANOVA term). It is therefore sufficient to work with very small index sets in order to obtain a good active set in this case. Contrary to the observations in Section 4, we notice that the global sensitivity indices are almost constant for a large number of the insignificant terms.

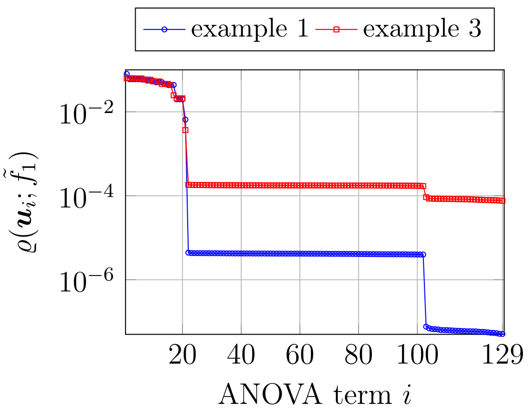


Figure 5.1: Behavior of the global sensitivity indices $\varrho(\boldsymbol{u}; \tilde{f}_1)$ for the examples 1 and 3 from Table 5.1.

Approximation

Since we identified an active set we now increase the size of the low-dimensional index sets

$$J_i = \{\boldsymbol{k} \in (\mathbb{Z} \setminus \{0\})^i : w^{(i)}_{s,\text{mix}}(\boldsymbol{k}) \leq N_i\}, i = 1, 2, 3,$$

with a larger cutoff vector $\boldsymbol{N}$. One should keep Remark 5.2 in mind when doing this since the oversampling parameter σ has an influence on the quality of the error. Solving the least squares problem (5.6) with the active set $U(\boldsymbol{\theta}; \tilde{f}_1)$ now yields the approximation $\tilde{f}_2$ as in (5.7).

Table 5.2 contains the results for different cutoff vectors $\boldsymbol{N}$ and Figure 5.2 displays the errors ε_2 and ε_{L_2}. Note that the error decreases from the first to the second example and then increases again. This behavior can be explained by the fact that the oversampling parameter σ decreases since the number of Fourier coefficients increases while the test function is smooth and the Fourier coefficients itself decrease rapidly anyway.

ex.	$\boldsymbol{N}$	t_{approx}	ε_2	ε_{L_2}
1	$[512, 16, 16]$	3038.86s	$5.05 \cdot 10^{-3}$	$5.06 \cdot 10^{-3}$
2	$[2048, 32, 16]$	3090.33s	$2.75 \cdot 10^{-3}$	$2.77 \cdot 10^{-3}$
3	$[2048, 32, 32]$	3511.59s	$4.83 \cdot 10^{-3}$	$4.98 \cdot 10^{-3}$
4	$[4096, 64, 32]$	3549.58s	$1.35 \cdot 10^{-2}$	$1.40 \cdot 10^{-2}$
5	$[4096, 64, 64]$	3569.92s	$3.19 \cdot 10^{-2}$	$4.02 \cdot 10^{-2}$

Table 5.2: Errors and runtimes for obtaining the approximation (5.5) with different cutoff vectors $\boldsymbol{N}$. Each time is the mean of three runs.

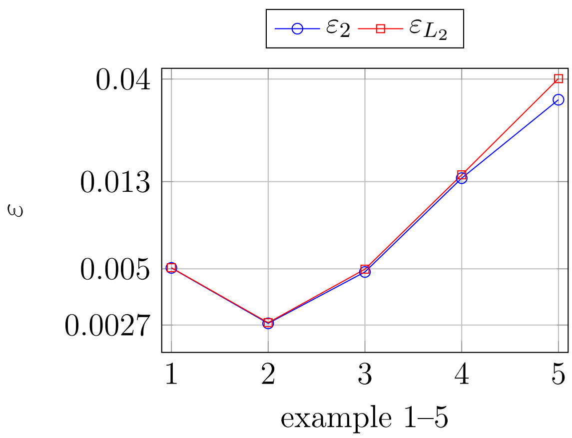


Figure 5.2: Errors ε_{L_2} and ε_2 for examples 1–5 from Table 5.2.

The results show that with as little as 5 million samples we are able to find an approximation where the error on the samples as well as the generalization error behave similarly with a best error of $\varepsilon_{L_2} \approx 2.7 \cdot 10^{-3}$. Furthermore, it is possible with this method to identify ANOVA terms that contribute most to the variance of the function which corresponds to finding the important dimension interactions.

6 Conclusion

In this thesis we studied the ANOVA decomposition

$$f(\boldsymbol{x}) = f_\emptyset + \sum_{i=1}^{d} f_{\{i\}}(\boldsymbol{x}) + \sum_{\substack{i,j=1 \\ i<j}}^{d} f_{\{i,j\}}(\boldsymbol{x}) + \cdots + f_{\{1,2,\ldots,d\}}(\boldsymbol{x}),$$

its properties and relations to Fourier analysis on the d-variate torus $\mathbb{T}^d$. We found means to describe the projections $P_{\boldsymbol{u}} f$, see Definition 3.3, and the ANOVA terms $f_{\boldsymbol{u}}$ with index sets $\mathbb{P}_{\boldsymbol{u}}^{(d)}$, see (3.5), and $\mathbb{F}_{\boldsymbol{u}}^{(d)}$, see (3.7), and how they decompose the frequency domain.

Furthermore, we considered the notion of inheritance of smoothness from [3], i.e., how the smoothness of a function f is inherited by its ANOVA terms $f_{\boldsymbol{u}}$, and translated results to the torus while also generalizing them in this context to Sobolev type spaces $\mathrm{H}^{\boldsymbol{w}}(\mathbb{T}^d)$, see Theorem 3.19, and the weighted Wiener algebra $\mathcal{A}_w(\mathbb{T}^d)$, see Theorem 3.22. We subsequently considered the variance of a function $\sigma^2(f)$ and its ANOVA terms $\sigma^2(f_{\boldsymbol{u}})$, e.g. as in [4, Chapter 2.1], as well as the global sensitivity indices $\varrho(\boldsymbol{u}; f)$ and related those terms to our previous findings in connection with Fourier analysis.

These considerations lead to the approximate ANOVA model

$$\mathcal{T}_{d_s} f \coloneqq \sum_{\substack{\boldsymbol{u} \subset \mathcal{D} \\ |\boldsymbol{u}| \leq d_s}} f_{\boldsymbol{u}}$$

as a special case of a proposed model in [4, Chapter 3.2]. For a fixed superposition dimension d_s, we have polynomial as opposed to exponential growth in the number of ANOVA terms. We subsequently found a formula to express the model $\mathcal{T}_{d_s} f$ through the projections $P_{\boldsymbol{u}} f$, see Theorem 3.29, as well as a formula for the Fourier coefficients. Furthermore, we proved error bounds for approximation with $\mathcal{T}_{d_s} f$ in $\mathrm{L}_2(\mathbb{T}^d)$ for Sobolev spaces, see Theorem 3.32, and $\mathrm{L}_\infty(\mathbb{T}^d)$ for the Wiener algebra, see Theorem 3.35.

The goal was to use the approximate model to detect the important dimensions and dimension interactions of a function $f : \mathbb{T}^d \to \mathbb{R}$ and use this information to obtain an approximation for the function. We started by assuming to have a function with black-box-access and used rank-1 lattice as sampling schemes. We developed an ANOVA version of the component-by-component algorithm to generate a reconstructing rank-1 lattice, see Algorithm 4.2, as well as an algorithm for the approximation, see Algorithm 4.1. In a first step we were using a Fourier partial sum with a small index set to obtain an approximation $\tilde{f}_1$ that allowed us to calculate approximate global sensitivity indices $\varrho(\boldsymbol{u}; \tilde{f}_1)$. With those sensitivity indices we constructed an active set of important ANOVA terms $U(\boldsymbol{\theta}; \tilde{f}_1)$. Taking into account only the terms contained in $U(\boldsymbol{\theta}; \tilde{f}_1)$, we arrived at a better approximation $\tilde{f}_2$. Numerical tests with a sum of products of B-splines, see (4.13), showed that the method works well. We were able to exactly detect the actual active set of a 9-dimensional test function and furthermore construct a good approximation from these results with a best error of

$$\frac{\left\| f - \tilde{f}_2 \right\|_{\mathrm{L}_2(\mathbb{T}^9)}}{\| f \|_{\mathrm{L}_2(\mathbb{T}^9)}} \approx 2.9 \cdot 10^{-4}.$$

Subsequently, we considered a scenario where only scattered data of the function is given. We adjusted the previously used strategy to obtain the approximation $\tilde{f}_1$ with the use of projections. This allowed us to solve a large least-squares system efficiently with the use of the NFFT. Following the same ideas as before, we calculated the approximate global sensitivity indices and constructed an active set $U(\boldsymbol{\theta}; \tilde{f}_1)$. We proceeded to use the active set to obtain the approximation $\tilde{f}_2$. Utilizing the same test function as before, we were again able to detect the actual active set using only 5 million samples. The best error of the subsequently calculated approximation $\tilde{f}_2$ achieved

$$\frac{\left\| f - \tilde{f}_2 \right\|_{\mathrm{L}_2(\mathbb{T}^9)}}{\| f \|_{\mathrm{L}_2(\mathbb{T}^9)}} \approx 2.7 \cdot 10^{-3}$$

Summarizing, one can say that the method developed in this thesis achieved success for a known benchmark function. Yet, for scattered data problems in applications working with the torus bears the disadvantage of having an induced periodicity. To counter this effect we are interested in translating the results to the unit cube $[0, 1]^d$ using the Chebyshev polynomials as an orthonormal basis instead of the Fourier basis.

References

[1] G. B. Folland. *Fourier Analysis and its Applications*. Brooks/Cole, Albany, 1992.

[2] 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(5):A3240–A3266, 2018.

[3] M. Griebel, F. Y. Kuo, and I. H. Sloan. The smoothing effect of the ANOVA decomposition. *J. Complexity*, 26(5):523–551, 2010.

[4] M. Holtz. *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance (Lecture Notes in Computational Science and Engineering)*. Springer, 2010.

[5] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along rank-1 lattices. *SIAM J. Numer. Anal.*, 51:2773–2796, 2013.

[6] L. Kämmerer. Reconstructing multivariate trigonometric polynomials from samples along rank-1 lattices. In G. E. Fasshauer and L. L. Schumaker, editors, *Approximation Theory XIV: San Antonio 2013*, pages 255–271. Springer International Publishing, 2014.

[7] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate periodic functions by trigonometric polynomials based on rank-1 lattice sampling. *J. Complexity*, 31:543–576, 2015.

[8] L. Kämmerer, D. Potts, and T. Volkmer. High-dimensional sparse FFT based on sampling along multiple rank-1 lattices. *ArXiv e-prints 1711.05152*, 2017.

[9] L. Kämmerer and T. Volkmer. Approximation of multivariate periodic functions based on sampling along multiple rank-1 lattices. *ArXiv e-prints 1802.06639*, 2018.

[10] J. Keiner, S. Kunis, and D. Potts. NFFT 3.5, C subroutine library. `http://www.tu-chemnitz.de/~potts/nfft`. Contributors: F. Bartel, M. Fenn, T. Görner, M. Kircheis, T. Knopp, M. Quellmalz, M. Schmischke, T. Volkmer, A. Vollrath.

[11] T. Kühn, S. Mayer, and T. Ullrich. Counting via entropy: New preasymptotics for the approximation numbers of sobolev embeddings. *SIAM J. Numer. Anal.*, 54(6):3625–3647, 2016.

[12] F. Kuo, D. Nuyens, L. Plaskota, I. Sloan, and G. Wasilkowski. Infinite-dimensional integration and the multivariate decomposition method. *J. Comput. Appl. Math.*, 326:217–234, dec 2017.

[13] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski. On decompositions of multivariate functions. *Math. Comp.*, 79(270):953–966, 2009.

[14] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2, 1990.

[15] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Math. Comput. Simulation*, 55(1-3):271–280, 2001.

[16] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.

[17] E. Novak and H. Wozniakowski. *Tractability of Multivariate Problems: Linear Information (Ems Tracts in Mathematics)*. European Mathematical Society, 2008.

[18] G. Plonka, D. Potts, G. Steidl, and M. Tasche. *Numerical Fourier Analysis (Applied and Numerical Harmonic Analysis)*. Birkhäuser, 2019.

[19] D. Potts and T. Volkmer. Sparse high-dimensional FFT based on rank-1 lattice sampling. *Appl. Comput. Harmon. Anal.*, 41:713–748, 2016.

[20] I. H. Sloan and P. J. Kachoyan. Lattice methods for multiple integration: Theory, error analysis and examples. *SIAM J. Numer. Anal.*, 24(1):116–128, 1987.

[21] H. Tyagi and J. Vybiral. Learning non-smooth sparse additive models from point queries in high dimensions. *arXiv e-prints*, 2018. arXiv:1801.08499 [math.NA], to appear in Constr. Approx.