LEARNING MULTIVARIATE FUNCTIONS WITH LOW-DIMENSIONAL STRUCTURES USING POLYNOMIAL BASES

A PREPRINT

Daniel Potts
Faculty of Mathematics
Chemnitz University of Technology
09107 Chemnitz
potts@math.tu-chemnitz.de

Michael Schmischke
Faculty of Mathematics
Chemnitz University of Technology
09107 Chemnitz
potts@math.tu-chemnitz.de

ABSTRACT

In this paper we propose a method for the approximation of high-dimensional functions over finite intervals with respect to complete orthonormal systems of polynomials. An important tool for this is the multivariate classical analysis of variance (ANOVA) decomposition. For functions with a low-dimensional structure, i.e., a low superposition dimension, we are able to achieve a reconstruction from scattered data and simultaneously understand relationships between different variables.

Keywords ANOVA decomposition · high-dimensional approximation · Chebyshev polynomials · orthogonal polynomials

1 Introduction

The approximation of high-dimensional functions is an active research topic and of high relevance in numerous applications. We assume a setting where we are given scattered data about an unknown function. The related approximation problem is generally referred to as scattered data approximation. Classical methods suffer from the curse of dimensionality in this setting, i.e., the amount of required data increases exponentially with the spatial dimension. Finding ways to circumvent the curse poses the main challenge in this high-dimensional setting. Besides finding an approximation there is the ever more important question of interpretability. In many application one wishes to understand how important the different dimensions and dimension interactions are in order to interpret the results.

In this paper we consider functions $f\colon [-1,1]^d \to \mathbb{R}$ defined over the cube with a high spatial dimension $d \in \mathbb{N}$. Given scattered data about f, i.e., a finite sampling set $\mathcal{X} \subseteq [-1,1]^d$ and evaluations $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$, we aim to construct an approximation of f and simultaneously understand its structure, i.e., how important variables and their interactions are. As opposed to black-box approximation or active learning, we may not choose the location of the nodes in $\mathcal{X}$. This prohibits us from using well-established spatial discretizations such as sparse grids, see [1, 2], or rank-1 lattices, see [3, 4, 5], that use low-dimensional structures in the node set. Our approach to circumvent the curse of dimensionality is to assume sparsity in the (analysis of variance) ANOVA decomposition of the function, i.e., we assume that f is dominated by a small number of low-complexity interactions. This may also be referred to as sparsity-of-effects, see e.g. [6].

We focus on complete orthonormal systems $\{\varphi_{\boldsymbol{k}}\}$ in $\mathrm{L}_2([-1,1]^d,\omega)$ where the functions are tensor products of univariate polynomials, e.g., the Chebyshev polynomials. Any function from the weighted Lebesgue space $\mathrm{L}_2([-1,1]^d,\omega)$ can then be written as a series $f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} c_{\boldsymbol{k}}\,\varphi_{\boldsymbol{k}}(\boldsymbol{x})$ with coefficients $c_{\boldsymbol{k}} \in \mathbb{R}$, $\boldsymbol{k} \in \mathbb{N}_0^d$. Our method focuses on approximations using partial sums of the type $S_I f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in I} c_{\boldsymbol{k}}\,\varphi_{\boldsymbol{k}}(\boldsymbol{x})$, with grouped finite index sets $I \subseteq \mathbb{N}_0^d$ that reflect the low-dimensional structure of f. Determining a frequency index set I that yields a good approximation while not scaling exponentially in d poses one of the main challenges.

The method presented here uses the classical ANOVA decomposition, see [7, 8, 9, 2], as a main tool. The decomposition is important in the analysis of the dimensions for multivariate, high-dimensional functions. It has also been used in

understanding the reason behind the success of certain quadrature methods for high-dimensional integration [10, 11, 12] and also infinite-dimensional integration [13, 14, 15]. The unique and orthogonal ANOVA decomposition decomposes a d-variate function in 2^d ANOVA terms where each term belongs to a subset of $\{1, 2, \dots, d\}$. The terms depends only on the variables in the corresponding subset and the number of these variables is the order of the ANOVA term.

Our method assumes sparsity by restricting the number of possible simultaneous dimension interactions. The knowledge that the function f has a structure such that it can be well approximated using this sparsity assumption is the only information we require a-priori. The approach allows us to learn the basis coefficients by solving a least-squares problem. The problem is hard to solve in general since we are dealing with a large system matrix, but we are able to apply the concept of grouped transformation, see [16], to tackle this issue. In summary, we present a method for the approximation of high-dimensional functions with a low-dimensional structure using possibly noisy scattered data.

The outline of the paper is as follows. In Section 2 we introduce some necessary preliminaries for weighted Lebesgue spaces with complete orthonormal systems of polynomials. Moreover, we discuss the non-equispaced fast cosine transform and the fast polynomial transform for the evaluation of Chebyshev partial sums and computing the basis exchange from any polynomial bases to the Chebyshev system, respectively. In Section 3 we consider the properties of the ANOVA decomposition in the previously explained setting of weighted Lebesgue spaces. The approximation method itself is discussed in Section 4 with numerical examples in Section 5.

2 Prerequisites, Notation and orthogonal Polynomials

Let $\tilde{\omega}\colon (-1,1) \to \mathbb{R}$ be a non-negative weight function with $\int_{-1}^{1} \tilde{\omega}(x)\ \mathrm{d}x\ =\ 1$ then we define the weighted Lebesgue space

$$\mathrm{L}_2([-1,1],\tilde{\omega}) \coloneqq \left\{ f\colon [-1,1] \to \mathbb{R}\colon\ \|f\|_{\mathrm{L}_2([-1,1],\tilde{\omega})} = \sqrt{\int_{-1}^{1} |f(x)|^2\ \tilde{\omega}(x)\,\mathrm{d}x} \right\}$$

with the inner product

$$\langle f, g\rangle \coloneqq \int_{-1}^{1} f(x)g(x)\,\omega(x)\,\mathrm{d}x.$$

Moreover, we consider a complete orthonormal system of polynomials $\{\varphi_k\}_{k\in\mathbb{N}_0}$ in $\mathrm{L}_2([-1,1],\tilde{\omega})$. Here, we have $\varphi_k \in \Pi_k$ with Π_k denoting the set of polynomials of degree $\leq k$. Taking the products $\varphi_{\boldsymbol{k}}(\boldsymbol{x}) \coloneqq \prod_{j=1}^{d} \varphi_{k_j}(x_j)$ we find that the system $\{\varphi_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^d}$ is an orthonormal basis in the tensor product space $\mathrm{L}_2([-1,1]^d,\omega)$ and the functions $f \in \mathrm{L}_2([-1,1]^d,\omega)$ have a unique representation with respect to the system $\{\varphi_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^d}$ as series

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}),$$

where $\mathrm{c}_{\boldsymbol{k}}(f) \coloneqq \int_{[-1,1]^d} f(\boldsymbol{x})\,\varphi_{\boldsymbol{k}}(\boldsymbol{x})\,\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \in \mathbb{R}$, $\boldsymbol{k} \in \mathbb{N}_0^d$, are the basis coefficients of f. The density ω is a product density, i.e.,

$$\omega(\boldsymbol{x}) \coloneqq \prod_{s\in\mathcal{D}} \tilde{\omega}(x_s).$$

For a finite index set $\mathcal{I} \subseteq \mathbb{N}_0^d$, we call

$$S(\mathcal{I})f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{I}} \mathrm{c}_{\boldsymbol{k}}(f)\,\varphi_{\boldsymbol{k}}(\boldsymbol{x}), \tag{1}$$

the partial sum of f with respect to the index set $\mathcal{I}$. In this paper we make use of the fact, that we are able to compute the sum (1) for arbitrary nodes $\boldsymbol{x}_j \in [-1,1]^d$, $j = 1, 2, \dots, M$, $M \in \mathbb{N}$, in an efficient manner. We realize this fast evaluation as follows:

Consider the univariate polynomial

$$P \coloneqq \sum_{k=0}^{N} c_k\,\varphi_k \in \Pi_N$$

with known real coefficients c_k. Our concern is the realization of a the basis exchange from $\{\varphi_k\}_{k=0}^N$ to $\{T_k\}_{k=0}^N$ in Π_N that produces the Chebyshev coefficients $\tilde{c}_k$ in

$$P\ =\ \sum_{k=0}^{N} \tilde{c}_k\,T_k\,.$$

By $T_k := \sqrt{2}^{1-\delta_{k,0}} \cos(k \arccos \cdot)$, we denote the normed Chebyshev polynomials of first kind. Note that $\arccos : [-1,1] \to [0,\pi)$ is the inverse function of $\cos$ restricted to $[0,\pi)$. As known, the Chebyshev polynomials form a complete orthonormal system in $\mathrm{L}_2([-1,1],\tilde{\omega})$ with the special Chebyshev density $\tilde{\omega}(x) := \pi^{-1} \cdot (1-x^2)^{-1/2}$. For $m,n \in \mathbb{N}_0$ we have

$$\langle T_m, T_n \rangle = \begin{cases} 1 & m = n, \\ 0 & m \neq n\,. \end{cases}$$

An algorithm, that realize the fast evaluation of $\tilde{c}_k$ from c_k is known as discrete polynomial transform and was developed in [17], see also the approach of Driscoll and Healy for the transposed problem developed in [18]. Our approach computes the basis exchange with $\mathcal{O}(N \log^2 N)$ arithmetical operations by a divide–and–conquer technique combined with fast polynomial multiplications. The algorithm was designed for arbitrary polynomials P_n satisfying a three–term recurrence relation, see [5, Section 6.5]. We introduce the notation $T_{\boldsymbol{k}}(\boldsymbol{x}) := \prod_{j=1}^d T_{k_j}(x_j)$ and observe that this algorithm can be straightforward generalized to the tensor product case, such that we realize the basis exchange, i.e., compute the Chebyshev coefficients $\tilde{c}_{\boldsymbol{k}} \in \mathbb{R}$ from the coefficients $c_{\boldsymbol{k}} \in \mathbb{R}$,

$$P = \sum_{\boldsymbol{k} \in \{0,1,\ldots,N\}^d} c_{\boldsymbol{k}} \varphi_{\boldsymbol{k}} = \sum_{\boldsymbol{k} \in \{0,1,\ldots,N\}^d} \tilde{c}_{\boldsymbol{k}} T_{\boldsymbol{k}},$$

in $\mathcal{O}(N^d \log^{2d} N)$ arithmetical operations. Knowing the Chebyshev coefficients $\tilde{c}_{\boldsymbol{k}}$, the values $P(\boldsymbol{x}_j)$, $j = 0,\ldots,M$, can be computed by the non-equidistant cosine transform at the nodes $\arccos(\boldsymbol{x}_j)$ by [5, Algorithm 7.10] in the complexity of $\mathcal{O}\big(N^d \log N + M\big)$ arithmetical operations. In summary we are able to compute the polynomial P at all arbitrary nodes $\boldsymbol{x}_j$, $j = 0,\ldots,M$

$$P(\boldsymbol{x}_j) = \sum_{\boldsymbol{k} \in \{0,1,\ldots,N\}^d} c_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}(\boldsymbol{x}_j), \tag{2}$$

in only $\mathcal{O}(N^d \log^{2d} N + M)$ arithmetical operations. For the special case of Chebyshev polynomials, i.e., $\varphi_{\boldsymbol{k}} = T_{\boldsymbol{k}}$ we need only $\mathcal{O}(N^d \log N + M)$ arithmetical operations, since the discrete polynomial transform is not necessary. We stress on the fact, that a fast algorithm implies the factorization of the transform matrix $\mathbf{P} := (\varphi_{\boldsymbol{k}}(\boldsymbol{x}_j))_{j=0,\ldots,M,\boldsymbol{k}\in\{0,1,\ldots,N\}^d}$ into a product of sparse matrices. Consequently, once a fast algorithm for (2) is known, a fast algorithm for the "transposed" problem

$$c_{\boldsymbol{k}} = \sum_{j=0}^{M} f_j \varphi_{\boldsymbol{k}}(\boldsymbol{x}_j), \quad \boldsymbol{k} \in \{0,1,\ldots,N\}^d \tag{3}$$

with the transform matrix $\mathbf{P}^{\mathrm{T}}$ and the same arithmetical complexity is also available by transposing the sparse matrix product. The algorithms are part of the software package [19].

In order to overcome the high complexity with growing dimensions d, we focus on models with a low superposition dimension, see Section 3. To this end we assume, that the effects of degree interactions among the input variables weaken rapidly or vanish altogether.

3 Classical Analysis of Variance Decomposition on the Interval

In this section we introduce the ANOVA decomposition in the setting of weighted Lebesgue spaces with orthonormal polynomials als bases. See also [7, 9, 20, 2, 21]. For a given spatial dimension d we denote with $\mathcal{D} = \{1,2,\ldots,d\}$ the set of coordinate indices and subsets as bold small letters, e.g., $\boldsymbol{u} \subseteq \mathcal{D}$. The complement of those subsets are always with respect to $\mathcal{D}$, i.e., $\boldsymbol{u}^{\mathrm{c}} = \mathcal{D} \setminus \boldsymbol{u}$. For a vector $\boldsymbol{x} \in \mathbb{C}^d$ we define $\boldsymbol{x}_{\boldsymbol{u}} = (x_i)_{i\in\boldsymbol{u}} \in \mathbb{C}^{|\boldsymbol{u}|}$. Furthermore, we use the p-norm (or quasi norm) of a vector which is defined as

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

for $\boldsymbol{x} \in \mathbb{R}^d$. The space $\mathrm{L}_2([-1,1]^d,\omega)$ with product density ω and complete orthonormal system $\{\varphi_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^d}$ consisting of tensor product functions, see Section 1, is fixed.

We start by defining the integral projection operator

$$\mathrm{P}_{\boldsymbol{u}} f(\boldsymbol{x}_{\boldsymbol{u}}) \coloneqq \int_{[-1,1]^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \omega(\boldsymbol{x}_{\boldsymbol{u}^{\mathrm{c}}})\, \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}^{\mathrm{c}}} \tag{4}$$

that integrates over the variables $\boldsymbol{x}_{\boldsymbol{u}^c}$. Clearly, the image $\mathrm{P}_{\boldsymbol{u}} f$ depends only on the variables $\boldsymbol{x}_{\boldsymbol{u}} \in [-1,1]^{|\boldsymbol{u}|}$. Furthermore, we define the index set

$$\mathbb{P}_{\boldsymbol{u}}^{(d)} \coloneqq \left\{\boldsymbol{k} \in \mathbb{N}_0^d \colon \boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0}\right\} \tag{5}$$

which can be identified with $\mathbb{N}_0^{|\boldsymbol{u}|}$ using the mapping $\boldsymbol{k} \mapsto \boldsymbol{k}_{\boldsymbol{u}}$ as well as the index set

$$\mathbb{F}_{\boldsymbol{u}}^{(d)} \coloneqq \left\{\boldsymbol{k} \in \mathbb{N}_0^d \colon \operatorname{supp} \boldsymbol{k} = \boldsymbol{u}\right\}$$

which can be identified with $\mathbb{N}^{|\boldsymbol{u}|}$ using the mapping $\boldsymbol{k} \mapsto \boldsymbol{k}_{\boldsymbol{u}}$. Moreover, we use the convention $\mathbb{N}_0^{|\emptyset|} = \{0\}$ and $\mathbb{N}^{|\emptyset|} = \{0\}$. The **ANOVA term** for $\boldsymbol{u} \subseteq \mathcal{D}$ is recursively defined as

$$f_{\boldsymbol{u}} \coloneqq \mathrm{P}_{\boldsymbol{u}} f - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} f_{\boldsymbol{v}}. \tag{6}$$

We now prove a relationship between the basis coefficients of $\mathrm{P}_{\boldsymbol{u}} f$, $f_{\boldsymbol{u}}$ and f.

Lemma 3.1. *Let $f \in \mathrm{L}_2([-1,1]^d, \omega)$ and $\boldsymbol{\ell} \in \mathbb{N}_0^{|\boldsymbol{u}|}$. Then*

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

and

$$\mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = \begin{cases} \mathrm{c}_{\boldsymbol{k}}(f) & : \boldsymbol{\ell} \in \mathbb{N}^{|\boldsymbol{u}|} \\ \delta_{\boldsymbol{u},\emptyset} \cdot \mathrm{c}_{\boldsymbol{0}}(f) & : \boldsymbol{\ell} = \boldsymbol{0} \\ 0 & : \textit{otherwise} \end{cases}$$

for $\boldsymbol{k} \in \mathbb{N}_0^d$ with $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0}$. Moreover, $\mathrm{P}_{\boldsymbol{u}} f, f_{\boldsymbol{u}} \in \mathrm{L}_2([-1,1]^{|\boldsymbol{u}|}, \omega)$.

Proof. We prove the formula for $\mathrm{c}_{\boldsymbol{\ell}}(\mathrm{P}_{\boldsymbol{u}} f)$, consolidate the two integrals and derive

$$\begin{aligned} \mathrm{c}_{\boldsymbol{\ell}}(\mathrm{P}_{\boldsymbol{u}} f) &= \int_{[-1,1]^{|\boldsymbol{u}|}} \int_{[-1,1]^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \omega(\boldsymbol{x}_{\boldsymbol{u}^c}) \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}^c}\, \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}) \omega(\boldsymbol{x}_{\boldsymbol{u}}) \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\ &= \int_{[-1,1]^d} f(\boldsymbol{x})\, \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}) \omega(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\ &= \int_{[-1,1]^d} f(\boldsymbol{x})\, \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \omega(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \mathrm{c}_{\boldsymbol{k}}(f) \end{aligned}$$

for $\boldsymbol{k} \in \mathbb{N}_0^d$ with $\boldsymbol{k}_{\boldsymbol{u}} = \boldsymbol{\ell}$ and $\boldsymbol{k}_{\boldsymbol{u}^c} = \boldsymbol{0}$. Then $\mathrm{P}_{\boldsymbol{u}} f \in \mathrm{L}_2([-1,1]^{|\boldsymbol{u}|}, \omega)$ is clear due to Parseval's identity.

In order to prove the formula for $\mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}})$, we employ the direct formula for the ANOVA terms $f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) = \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{P}_{\boldsymbol{v}} f(\boldsymbol{x}_{\boldsymbol{v}})$ to obtain

$$\begin{aligned} \mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) &= \int_{\mathbb{T}^{|\boldsymbol{u}|}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}) \omega(\boldsymbol{x}_{\boldsymbol{u}}) \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\ &= \int_{\mathbb{T}^{|\boldsymbol{u}|}} \left[\sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{P}_{\boldsymbol{v}} f(\boldsymbol{x}_{\boldsymbol{v}}) \right] \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}) \omega(\boldsymbol{x}_{\boldsymbol{u}}) \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\ &= \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \int_{\mathbb{T}^{|\boldsymbol{u}|}} \mathrm{P}_{\boldsymbol{v}} f(\boldsymbol{x}_{\boldsymbol{v}}) \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}) \omega(\boldsymbol{x}_{\boldsymbol{u}}) \,\mathrm{d}\boldsymbol{x}_{\boldsymbol{u}} \\ &= \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{c}_{\boldsymbol{k}_{\boldsymbol{v}}} (\mathrm{P}_{\boldsymbol{v}} f)\, \delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}}. \end{aligned}$$

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

$$\begin{aligned} \mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) &= \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{c}_{\boldsymbol{k}_{\boldsymbol{v}}} (\mathrm{P}_{\boldsymbol{v}} f)\, \delta_{\boldsymbol{k}_{\boldsymbol{u} \setminus \boldsymbol{v}}, \boldsymbol{0}} = \mathrm{c}_{\boldsymbol{0}}(f) \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \\ &= \mathrm{c}_{\boldsymbol{0}}(f) \sum_{n=0}^{|\boldsymbol{u}|} \binom{|\boldsymbol{u}|}{n} (-1)^{|\boldsymbol{u}|-n} = \mathrm{c}_{\boldsymbol{0}}(f) \cdot \delta_{\boldsymbol{u},\emptyset}. \end{aligned}$$

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

$$\begin{aligned}
\mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) &= \sum_{\boldsymbol{v}\subseteq\boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{c}_{\boldsymbol{k}_{\boldsymbol{v}}}\left(\mathrm{P}_{\boldsymbol{v}} f\right) \delta_{\boldsymbol{k}_{\boldsymbol{u}\setminus\boldsymbol{v}},\mathbf{0}} = \sum_{\overline{\boldsymbol{v}}^{\mathrm{c}}\subseteq\boldsymbol{v}\subseteq\boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{c}_{\boldsymbol{k}_{\boldsymbol{v}}}\left(\mathrm{P}_{\boldsymbol{v}} f\right) \\
&= \mathrm{c}_{\boldsymbol{k}}(f) \sum_{\overline{\boldsymbol{v}}^{\mathrm{c}}\subseteq\boldsymbol{v}\subseteq\boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} = \mathrm{c}_{\boldsymbol{k}}(f) \sum_{n=|\overline{\boldsymbol{v}}^{\mathrm{c}}|}^{|\boldsymbol{u}|} \binom{|\boldsymbol{u}| - |\overline{\boldsymbol{v}}^{\mathrm{c}}|}{n - |\overline{\boldsymbol{v}}^{\mathrm{c}}|} (-1)^{|\boldsymbol{u}|-n} \\
&= \mathrm{c}_{\boldsymbol{k}}(f) \sum_{m=0}^{|\boldsymbol{u}|-|\overline{\boldsymbol{v}}^{\mathrm{c}}|} \binom{|\boldsymbol{u}| - |\overline{\boldsymbol{v}}^{\mathrm{c}}|}{m} (-1)^{|\boldsymbol{u}|-|\overline{\boldsymbol{v}}^{\mathrm{c}}|-m} = 0.
\end{aligned}$$

For the case where the entries of $\boldsymbol{\ell}$ are all nonzero, only the addend where $\boldsymbol{v} = \boldsymbol{u}$ is nonzero, i.e., $\mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}}) = \mathrm{c}_{\boldsymbol{k}}(f)$ and $f_{\boldsymbol{u}} \in \mathrm{L}_2([-1,1]^{|\boldsymbol{u}|}, \omega)$ is clear due to Parseval's identity. □

Using Lemma 3.1, we are able to write $\mathrm{P}_{\boldsymbol{u}} f$ and $f_{\boldsymbol{u}}$ as both, d-dimensional

$$\mathrm{P}_{\boldsymbol{u}} f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{P}_{\boldsymbol{u}}^{(d)}} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}),\ f_{\boldsymbol{u}}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x})$$

and $|\boldsymbol{u}|$-dimensional series

$$\mathrm{P}_{\boldsymbol{u}} f(\boldsymbol{x}_{\boldsymbol{u}}) = \sum_{\boldsymbol{\ell}\in\mathbb{N}_0^{|\boldsymbol{u}|}} \mathrm{c}_{\boldsymbol{\ell}}(\mathrm{P}_{\boldsymbol{u}} f)\ \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}),\ f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) = \sum_{\boldsymbol{\ell}\in\mathbb{N}^{|\boldsymbol{u}|}} \mathrm{c}_{\boldsymbol{\ell}}(f_{\boldsymbol{u}})\ \varphi_{\boldsymbol{\ell}}(\boldsymbol{x}_{\boldsymbol{u}}).$$

This directly implies that $\langle f_{\boldsymbol{u}}, f_{\boldsymbol{v}} \rangle = 0$ for $\boldsymbol{u} \neq \boldsymbol{v}$. With the ANOVA terms we are able to introduce the ANOVA decomposition.

Theorem 3.2. *Let $f \in \mathrm{L}_2([-1,1]^d, \omega)$, the ANOVA terms $f_{\boldsymbol{u}}$ as in* (6) *and the set of coordinate indices $\mathcal{D} = \{1, 2, \ldots, d\}$. Then f can be uniquely decomposed as*

$$f(\boldsymbol{x}) = f_{\emptyset} + \sum_{i=1}^{d} f_{\{i\}}(x_i) + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} f_{\{i,j\}}(\boldsymbol{x}_{\{i,j\}}) + \cdots + f_{\mathcal{D}}(\boldsymbol{x}) = \sum_{\boldsymbol{u}\subseteq\mathcal{D}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) \tag{7}$$

*which we call **analysis of variance (ANOVA) decomposition**. Moreover, $\bigcup_{\boldsymbol{u}\subseteq\mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)} = \mathbb{N}_0^d$ and the union is disjoint.*

Proof. We use that $\mathbb{N}_0^d$ is clearly the disjoint union of the sets $\mathbb{F}_{\boldsymbol{u}}^{(d)}$ for $\boldsymbol{u} \subseteq \mathcal{D}$. With this fact we obtain

$$\begin{aligned}
\sum_{\boldsymbol{u}\subseteq\mathcal{D}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) &= \sum_{\boldsymbol{u}\subseteq\mathcal{D}} \sum_{\boldsymbol{k}\in\mathbb{F}_{\boldsymbol{u}}^{(d)}} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\bigcup_{\boldsymbol{u}\subseteq\mathcal{D}} \mathbb{F}_{\boldsymbol{u}}^{(d)}} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \\
&= \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}) = f(\boldsymbol{x}).
\end{aligned}$$

Since the union is disjoint, the decomposition is unique. □

In order to get a notion of the importance of single terms compared to the entire function, we define the **variance of a function**

$$\sigma^2(f) \coloneqq \int_{[-1,1]^d} \left(f(\boldsymbol{x}) - \mathrm{c}_{\mathbf{0}}(f)\right)^2 \omega(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x}$$

and the equivalent formulation

$$\sigma^2(f) = \|f\|^2_{\mathrm{L}_2([-1,1]^d,\omega)} - \left|\mathrm{c}_{\mathbf{0}}(f)\right|^2 . \tag{8}$$

For the ANOVA terms $f_{\boldsymbol{u}}$ with $\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}$, we have $\mathrm{c}_{\mathbf{0}}(f_{\boldsymbol{u}}) = 0$ and therefore $\sigma^2(f_{\boldsymbol{u}}) = \|f_{\boldsymbol{u}}\|^2_{\mathrm{L}_2([-1,1]^{|\boldsymbol{u}|},\omega)}$. For $f \in \mathrm{L}_2([-1,1]^d, \omega)$ we obtain the property

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

for the variance by Parsevel's identity. In order to measure the importance of a term $f_{\boldsymbol{u}}$ in relation to the function, we use global sensitivity indices, cf. [22, 23, 9],

$$\varrho(\boldsymbol{u}, f) \coloneqq \frac{\sigma^2(f_{\boldsymbol{u}})}{\sigma^2(f)} \in [0,1] \tag{9}$$

for $\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}$. They have the property $\sum_{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D}} \varrho(\boldsymbol{u}, f) = 1$.

The global sensitivity indices motivate the notion of **effective dimensions** as proposed in [7]. Given a fixed $\delta \in [0,1]$, the **superposition dimension**, one notion of effective dimension, is defined as

$$\min \left\{ s \in \mathcal{D} \colon \sum_{\substack{\emptyset \neq \boldsymbol{u} \subseteq \mathcal{D} \\ |\boldsymbol{u}| \leq s}} \sigma^2(f_{\boldsymbol{u}}) \geq \delta \sigma^2(f) \right\} \tag{10}$$

for accuracy δ. In other words, the proportion δ of the variance $\sigma^2(f)$ is explained by ANOVA terms of order less or equal to the superposition dimension.

The number of ANOVA terms in a full decomposition is $|\mathcal{P}(\mathcal{D})| = 2^d$ and therefore grows exponentially in d. This reflects the curse of dimensionality and poses a problem in high-dimensional approximation. In order to circumvent that, we make use of sparsity in the ANOVA decomposition. Specifically, we focus on truncating the ANOVA decomposition, i.e., removing certain terms $f_{\boldsymbol{u}}$. We therefore define a **subset of ANOVA terms** as a subset of the power set of $\mathcal{D}$, i.e., $U \subseteq \mathcal{P}(D)$, such that it is downward closed, i.e., the inclusion condition

$$\boldsymbol{u} \in U \implies \forall \boldsymbol{v} \subseteq U \colon \boldsymbol{v} \in U \tag{11}$$

holds, cf. [2, Chapter 3.2]. This fits with the recursive definition of the ANOVA terms, see (6). For any subset of ANOVA terms U we then define the **truncated ANOVA decomposition** as

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

This truncation can be done with the superposition concept in mind, cf. (10). For a superposition threshold $d_s \in \mathcal{D}$ we define $U_{d_s} \coloneqq \{\boldsymbol{u} \subseteq \mathcal{D} \colon |\boldsymbol{u}| \leq d_s\}$ and $\mathrm{T}_{d_s} \coloneqq \mathrm{T}_{U_{d_s}}$. This reduces the number of ANOVA terms to grow polynomially in d for fixed d_s since

$$|U_{d_s}| \leq \left(\frac{d \cdot \mathrm{e}}{d_s} \right)^{d_s}, \tag{12}$$

cf. [21]. The basis coefficients of the truncated ANOVA decomposition are then

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

which means that $\mathrm{c}_{\boldsymbol{k}}(\mathrm{T}_{d_s} f)$ is nonzero only for at most d_s-sparse frequencies.

The approximation method introduced in Section 4 uses partial sums where the frequency index sets have a grouped structure related to the ANOVA terms in a set $U \subseteq \mathcal{P}(\mathcal{D})$. Every finite index set $\mathcal{I}_{\boldsymbol{u}} \subseteq \mathbb{N}^d$ corresponds to one ANOVA term $f_{\boldsymbol{u}}$, $\boldsymbol{u} \in U$, i.e.,

$$\mathcal{I}_{\boldsymbol{u}} \subseteq \left\{ \boldsymbol{k} \in \mathbb{N}^d \colon \operatorname{supp} \boldsymbol{k} = \boldsymbol{u} \right\},$$

with $\mathcal{I}_\emptyset = \{\boldsymbol{0}\}$ and for the disjoint union we have

$$\mathcal{I}(U) \coloneqq \bigcup_{\boldsymbol{u} \in U} I_{\boldsymbol{u}} \subseteq \mathbb{N}_0^d. \tag{13}$$

It is also possible to choose the frequencies only based on the order of the ANOVA term $|\boldsymbol{u}|$, i.e., we have for the projections

$$\{\boldsymbol{k}_{\boldsymbol{u}} \in \mathbb{N}^{|\boldsymbol{u}|} \colon \boldsymbol{k} \in \mathcal{I}_{\boldsymbol{u}}\} = \{\boldsymbol{\ell}_{\boldsymbol{v}} \in \mathbb{N}^{|\boldsymbol{v}|} \colon \boldsymbol{\ell} \in \mathcal{I}_{\boldsymbol{v}}\}$$

for every pair of sets $\boldsymbol{u}, \boldsymbol{v} \subseteq \mathcal{D}$ with $|\boldsymbol{v}| = |\boldsymbol{u}|$.

4 Approximation Method

In this section, we present a method for the approximation of functions $f\colon [-1,1]^d \to \mathbb{C}$ with a high spatial dimension $d \in \mathbb{N}$ such that $f \in \mathrm{L}_2([-1,1]^d, \omega)$. In scattered data approximation, the data consists of a finite set of sampling nodes $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_M\} \subseteq [-1,1]^d$ and a vector of values $\boldsymbol{y} \in \mathbb{R}^M$. Now, we assume that $y_i \approx f(\boldsymbol{x}_i)$, i.e., the entries of $\boldsymbol{y}$ are noisy evaluations of the function. Here, it is especially important that we cannot choose the location of the nodes $\boldsymbol{x}_i$. The space $\mathrm{L}_2([-1,1]^d, \omega)$ and the corresponding complete orthonormal system $\{\varphi_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^d}$ is fixed. Moreover, we focus on functions with a low-dimensional structure, i.e., a low superposition dimension, cf. (10). This implies that choosing a low threshold $d_s \in \mathcal{D}$ will yield a good approximation $\mathrm{T}_{d_s} f(\boldsymbol{x}) \approx f(\boldsymbol{x})$. It has been speculated that functions in many applications consist of a low-dimensional structure and therefore belong to our class. This is referred to as sparsity-of-effects or the Pareto principle, see e.g. [7, 24, 6]. From a theoretical standpoint, functions of specific smoothness classes also have a low-dimensional structure. In [21] it was shown that functions of certain isotropic and dominating-mixed smoothness belong to this class. In particular, a POD (or *product and order-dependent*) weight structure was considered which is motivated by the application of quasi-Monte Carlo methods for PDEs with random coefficients, cf. [25, 26, 27, 28].

The idea of the method is to exploit sparsity in the ANOVA decomposition by considering only terms up to order d_s, i.e., $\mathrm{T}_{d_s} f$. The immediate benefit is that the number of terms is reduced from being exponential in the spatial dimension d to being polynomial, see (12). This assumption also provides us with a way of efficiently calculating an initial least-squares approximation on the basis coefficients. From there we focus on understanding the structure of the function regarding the importance of dimensions and dimension interactions, i.e., the importance of ANOVA terms $f_{\boldsymbol{u}}$. We measure the importance of a term $f_{\boldsymbol{u}}$ using the global sensitivity indices $\varrho(\boldsymbol{u}, f)$, see (9). In order to reduce the number of basis coefficients and subsequently the model complexity further, we use this knowledge to reduce the number of involved ANOVA terms to certain subset $U \subseteq U_{d_s}$. This simplifies our model function and reduces effects of overfitting.

In Section 4.1 we consider how to obtain an approximation on the function given a set of ANOVA terms $U \subseteq \mathcal{P}(\mathcal{D})$. This may be the set U_{d_s} for the initial approximation with threshold $d_s \in \mathcal{D}$ or an active set $U \subseteq U_{d_s}$. The detection of the active set will be addressed in Section 4.2.

4.1 Least Squares and Grouped Transformations

In this section we explain the optimization problem for obtaining an approximation on the basis coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$ of our function f given a set of terms $U \subseteq \mathcal{P}(\mathcal{D})$. Here, $U = U_{d_s}$ for the initial approximation and $U \subseteq U_{d_s}$ after the active set detection. We focus on the Chebyshev system

$$T_{\boldsymbol{k}}(\boldsymbol{x}) \coloneqq \sqrt{2}^{\|\boldsymbol{k}\|_0} \prod_{s\in\operatorname{supp}\boldsymbol{k}} \cos(k_s \arccos x_s)$$

which is a complete orthonormal system in $\mathrm{L}_2([-1,1]^d, \omega)$ with the Chebyshev density

$$\omega(\boldsymbol{x}) = \prod_{s\in\mathcal{D}} \frac{1}{\pi\sqrt{1-x_s^2}} \tag{14}$$

since we may use the FPT to compute the Chebyshev coefficients from other polynomials, see Section 1. Now, we approximate f by a finite partial sum $S(\mathcal{I}(U))f(\boldsymbol{x})$, see (1), with corresponding index set $\mathcal{I}(U)$ of a grouped structure, cf. (13) for a superposition threshold $d_s \in \mathcal{D}$. The index set for every ANOVA term $f_{\boldsymbol{u}}$, $\boldsymbol{u} \in U \setminus \emptyset$, is given by

$$\mathcal{I}_{\boldsymbol{u}} = \{\boldsymbol{k} \in \mathbb{N}^d \colon \operatorname{supp}\boldsymbol{k} = \boldsymbol{u}, \|\boldsymbol{k}_{\boldsymbol{u}}\|_\infty \le N_{\boldsymbol{u}} - 1\} \tag{15}$$

with parameters $N_{\boldsymbol{u}} \in \mathbb{N}$ and $\mathcal{I}_\emptyset = \{\boldsymbol{0}\}$. We then have

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \approx \mathrm{T}_U f(\boldsymbol{x}) \approx S(\mathcal{I}(U))f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{I}(U)} \mathrm{c}_{\boldsymbol{k}}(f)\ \varphi_{\boldsymbol{k}}(\boldsymbol{x}). \tag{16}$$

The coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$ are unknown and have to be determined from the given scattered data $\mathcal{X}$ and $\boldsymbol{y}$.

Here, we distinguish between two different cases. The first case being that the nodes $\mathcal{X}$ are distributed i.i.d. according to the Chebyshev probability density ω in (14) and the second case being that $\mathcal{X}$ is distributed uniformly in $[-1,1]^d$.

4.1.1 Chebyshev Distributed Nodes

Here, we assume that the nodes $\mathcal{X}$ are distributed i.i.d. according to the Chebyshev probability density ω. We aim to determine approximations for the basis coefficients by solving the minimization problem

$$\hat{\boldsymbol{f}}_{\text{sol}} = \underset{\hat{\boldsymbol{f}} \in \mathbb{R}^{|\mathcal{I}(U)|}}{\arg\min} \left\| \boldsymbol{y} - \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)) \hat{\boldsymbol{f}} \right\|_2^2 \tag{17}$$

with system matrix $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)) = (\varphi_{\boldsymbol{k}}(\boldsymbol{x}))_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{k} \in \mathcal{I}(U)}$. Solving the problem (17) is equivalent to solving the normal equation

$$\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U)) \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)) \hat{\boldsymbol{f}}_{\text{sol}} = \boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U)) \boldsymbol{y}. \tag{18}$$

The properties of this system have been considered in [29]. To summarize, we get from [29, Section 5] that the expected value of the matrix product $\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U)) \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ is a diagonal matrix and the singular values of $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ are between $\sqrt{|\mathcal{X}|/2}$ and $\sqrt{3|\mathcal{X}|/2}$. This yields an upper bound for the norm of the Moore-Penrose inverse

$$\left\| \left(\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U)) \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)) \right)^{-1} \boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U)) \right\|_2 < \sqrt{\frac{|\mathcal{X}|}{2}}.$$

This holds with a probability of $1 - \delta$ if

$$|I(U)| \leq \frac{1}{2^{d_s} \cdot 48(\sqrt{2} - \log \delta)} \cdot \frac{|X|}{\log(2|\mathcal{X}|)}.$$

We also gain that the matrix $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)$ has full rank and our problem a unique solution $\hat{\boldsymbol{f}}_{\text{sol}}$ with this probability.

Problem (17) is difficult to solve in general since we have a large matrix and need an efficient matrix-vector multiplication. However, we have a grouped index set $\mathcal{I}(U)$ which allow us to use the Grouped Transformations idea from [16]. If we have an order $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$, $n = |U|$, on the ANOVA terms, we may write

$$\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)) = (\boldsymbol{F}_1 \ \boldsymbol{F}_2 \ \cdots \ \boldsymbol{F}_n)$$

with $\boldsymbol{F}_i = (\varphi_{\boldsymbol{k}}(\boldsymbol{x}))_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{k} \in \mathcal{I}_{\boldsymbol{u}_i}}$. Therefore, a multiplication of $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ with the vector $\hat{\boldsymbol{f}} = (\hat{\boldsymbol{f}}_1, \hat{\boldsymbol{f}}_2, \ldots, \hat{\boldsymbol{f}}_n) \in \mathbb{R}^{|\mathcal{I}(U)|}$ can be written as

$$\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U)) \hat{\boldsymbol{f}} = \sum_{j=1}^{n} \boldsymbol{F}_j \hat{\boldsymbol{f}}_j.$$

An efficient way of performing the matrix-vector multiplication $\boldsymbol{F}_j \hat{\boldsymbol{f}}_j$ can then be realized using a NFCT, see Section 1. The same holds true for the adjoint problem, i.e., the multiplication of $\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U))$ with a vector $\boldsymbol{f} \in \mathbb{R}^{|\mathcal{X}|}$. Here, we have

$$\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U)) \boldsymbol{f} = \begin{pmatrix} \boldsymbol{F}_1^\top \boldsymbol{f} \\ \boldsymbol{F}_2^\top \boldsymbol{f} \\ \vdots \\ \boldsymbol{F}_n^\top \boldsymbol{f} \end{pmatrix}.$$

In this case we can use an adjoint NFCT for efficient multiplications $\boldsymbol{F}_j^\top \boldsymbol{f}$. We then proceed to solve (17) using iterative LSQR, see [30], in a matrix-free variant, i.e., $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ is not explicitly required by providing the fast grouped transformations algorithm for multiplication of vectors with matrices of type $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ and $\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U))$.

Remark 4.1. *The multiplications $\boldsymbol{F}_j \hat{\boldsymbol{f}}_j$ and $\boldsymbol{F}_j^\top \boldsymbol{f}$, $j = 1, 2, \ldots, n$, are all independent of each other which allows us to use parallelization for a fast multiplication. If the computer allows for it, it is possible to calculate all n products and n adjoint products simultaneously.*

The elements of the solution vector $\hat{\boldsymbol{f}}_{\text{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathcal{I}(U)}$ are the unique least-squares approximation to the basis coefficients, i.e., $\hat{f}_{\boldsymbol{k}} \approx \mathrm{c}_{\boldsymbol{k}}(f)$, with respect to $\mathcal{X}$ and $\boldsymbol{y}$. We then have an approximation by the approximate partial sum

$$S(\mathcal{X}, \mathcal{I}(U)) f(\boldsymbol{x}) \coloneqq \sum_{\boldsymbol{k} \in \mathcal{I}(U)} \hat{f}_{\boldsymbol{k}} \, \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \approx S(\mathcal{I}(U)) f(\boldsymbol{x}). \tag{19}$$

4.1.2 Uniformly Distributed Nodes

In this section we assume that the nodes $\mathcal{X}$ are uniformly i.i.d distributed in $[-1,1]^d$. If we would proceed in the same way as before, the expected value of the matrix $\boldsymbol{F}^\top(\mathcal{X},\mathcal{I}(U))\boldsymbol{F}(\mathcal{X},\mathcal{I}(U))$ from the normal equation (18) would not be the identity. In other words, our system would not be stable. However, this can be fixed by scaling and preconditioning.

We choose a padding parameter $\vartheta \in (0,1)$ and scale the nodes $\mathcal{X}$ such that

$$\tilde{\mathcal{X}} \coloneqq \left\{ \begin{pmatrix} (1-\vartheta)\, x_1 \\ \vdots \\ (1-\vartheta)\, x_d \end{pmatrix} : \boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathcal{X} \right\},$$

i.e., we have uniformly distributed nodes in $[-1+\vartheta, 1-\vartheta]^d$. Now, we choose our preconditioner as the diagonal matrix

$$\boldsymbol{W} = \operatorname{diag}\left(\sqrt{\omega(\boldsymbol{x})}\right)_{\boldsymbol{x}\in\mathcal{X}}$$

such that we have the minimization problem

$$\hat{\boldsymbol{f}}_{\text{sol}} = \underset{\hat{\boldsymbol{f}}\in\mathbb{R}^{|\mathcal{I}(U)|}}{\arg\min} \left\| \boldsymbol{W}\boldsymbol{y} - \boldsymbol{W}\boldsymbol{F}(\mathcal{X},\mathcal{I}(U))\hat{\boldsymbol{f}} \right\|_2^2.$$

The normal equation (18) transforms to

$$\boldsymbol{F}^\top(\mathcal{X},\mathcal{I}(U))\boldsymbol{W}^2\boldsymbol{F}(\mathcal{X},\mathcal{I}(U))\hat{\boldsymbol{f}}_{\text{sol}} = \boldsymbol{F}^\top(\mathcal{X},\mathcal{I}(U))\boldsymbol{W}\boldsymbol{y} \tag{20}$$

with $\boldsymbol{W}^2 = \boldsymbol{W}\cdot\boldsymbol{W}$. We denote

$$\boldsymbol{H} \coloneqq \boldsymbol{F}^\top(\mathcal{X},\mathcal{I}(U))\boldsymbol{W}^2\boldsymbol{F}(\mathcal{X},\mathcal{I}(U)). \tag{21}$$

In the following we consider the properties of this system and prove that with this preconditioner we are able to achieve a stable system under certain conditions.

Lemma 4.2. *Let $\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}_0^d$, $\boldsymbol{k}\neq\boldsymbol{\ell}$, with $\|\boldsymbol{k}\|_0, \|\boldsymbol{\ell}\|_0 \leq d_s \in \mathcal{D}$ and $\vartheta\in(0,1)$. Then*

$$\left| \int_{[-1,-1+\vartheta]^d} T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right| \leq 4^{d_s} \left(\frac{\arccos(1-\vartheta)}{\pi} \right)^d$$

and

$$\left| \int_{[1-\vartheta,1]^d} T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right| \leq 4^{d_s} \left(\frac{\arccos(1-\vartheta)}{\pi} \right)^d.$$

Proof. We define $C(k,x) = \cos(k\arccos(x))$ and set $M_1(\boldsymbol{k},\boldsymbol{\ell}) = \{s\in\mathcal{D}\colon k_s\neq 0, \ell_s\neq 0\}$, $M_2(\boldsymbol{k},\boldsymbol{\ell}) = \{s\in\mathcal{D}\colon \text{either } k_s=0 \text{ or } \ell_s=0\}$, $M_3(\boldsymbol{k},\boldsymbol{\ell}) = \{s\in\mathcal{D}\colon k_s=\ell_s=0\}$. The first integral can be separated as follows

$$\begin{aligned} \int_{[-1,-1+\vartheta]^d} T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \prod_{s\in M_1} \underbrace{\frac{2}{\pi}\int_{-1}^{-1+\vartheta} C(k_s,x)C(\ell_s,x)\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x}_{I_1} \\ \times \prod_{s\in M_2} \underbrace{\frac{\sqrt{2}}{\pi}\int_{-1}^{-1+\vartheta} C(\max\{k_s,\ell_s\},x)\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x}_{I_2} \\ \times \prod_{s\in M_3} \underbrace{\frac{1}{\pi}\int_{-1}^{-1+\vartheta} \frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x}_{I_3}. \end{aligned}$$

We have

$$I_1 = -\left[\frac{\sin((k_s-\ell_s)\arccos(x))}{\pi(k_s-\ell_s)} + \frac{\sin((k_s+\ell_s)\arccos(x))}{\pi(k_s+\ell_s)} \right]_{-1}^{-1+\vartheta}$$

and since $\arccos(-1)=\pi$ this becomes

$$I_1 = -\frac{\sin((k_s-\ell_s)\arccos(-1+\vartheta))}{\pi(k_s-\ell_s)} - \frac{\sin((k_s+\ell_s)\arccos(-1+\vartheta))}{\pi(k_s+\ell_s)}.$$

Writing $\arccos(-1+\vartheta) = \pi - \rho$ yields

$$|\sin((k_s - \ell_s)(\pi - \rho))| = |\sin((k_s - \ell_s)\rho)| \leq |(k_s - \ell_s)\rho|\,.$$

Therefore, we get

$$|I_1| \leq \frac{\rho}{\pi} + \frac{\rho}{\pi} = \frac{2\rho}{\pi} = \frac{2\arccos(1-\vartheta)}{\pi}.$$

For the second integral we have w.l.o.g.

$$I_2 = \left[-\frac{\sqrt{2}\sin(k_s \arccos(x))}{k_s \pi}\right]_{\pi}^{-1+\vartheta} = -\frac{\sqrt{2}\sin(k_s \arccos(-1+\vartheta))}{k_s \pi}$$

and

$$|I_2| \leq \frac{\sqrt{2}\rho}{\pi}.$$

For the last intergral we deduce

$$I_3 = \frac{1}{\pi}\left(\arcsin(-1+\vartheta) - \arcsin(-1)\right) = \frac{1}{\pi}\arccos(1-\vartheta).$$

The final result follows by

$$2^{|M_1|}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{|M_1|} \cdot \sqrt{2}^{|M_2|}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{|M_2|} \tag{22}$$

$$\cdot \left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{|M_3|} \leq 2^{d_s}\sqrt{2}^{2d_s}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{d}. \tag{23}$$

The steps work analogously for the second integral. □

Lemma 4.3. *Let $\boldsymbol{k} \in \mathbb{N}_0^d$ with $\|\boldsymbol{k}\|_0 \leq d_s \in \mathcal{D}$ and $\vartheta > 0$. Then*

$$\left|\int_{[-1,-1+\vartheta]^d} T_{\boldsymbol{k}}^2(\boldsymbol{x})\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \leq 2^{d_s}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{d}$$

and

$$\left|\int_{[1-\vartheta,1]^d} T_{\boldsymbol{k}}^2(\boldsymbol{x})\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \leq 2^{d_s}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{d}.$$

Proof. We define $C(k,x) = \cos(k\arccos(x))$ and write the first integral as the product

$$\int_{[-1,-1+\vartheta]^d} T_{\boldsymbol{k}}^2(\boldsymbol{x})\omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \prod_{s\in\operatorname{supp}\boldsymbol{k}} \underbrace{\frac{2}{\pi}\int_{-1}^{-1+\vartheta} C^2(k_s,x)\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x}_{I_1}$$

$$\prod_{s\in\mathcal{D}\setminus\operatorname{supp}\boldsymbol{k}} \underbrace{\frac{1}{\pi}\int_{-1}^{-1+\vartheta}\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x}_{I_2}.$$

We then have

$$I_1 = -\frac{2}{\pi}\left[\frac{2k_s\arccos x + \sin(2k_s\arccos x)}{4k}\right]_{-1}^{-1+\vartheta}$$

$$= 1 - \frac{\arccos(\vartheta-1)}{\pi} - \frac{\sin(2k_s\arccos(\vartheta-1))}{2\pi k_s}$$

and estimate the absolute value

$$|I_1| \leq \frac{2\arccos(1-\vartheta)}{\pi}.$$

The integral I_2 was considered in the proof of Lemma 4.2 which results in

$$\left(\frac{2\arccos(1-\vartheta)}{\pi}\right)^{d_s}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{d-d_s} = 2^{d_s}\left(\frac{\arccos(1-\vartheta)}{\pi}\right)^{d}.$$

The steps work analogously for the second integral. □

Using the previous two lemmas, we are able to prove that the expected value of the matrix $\boldsymbol{H}$ from (21) is close to a diagonal matrix.

Theorem 4.4. *Let $\mathcal{X} \subseteq [-1+\vartheta, 1-\vartheta]^d$, $0 < \vartheta < 1$, be a set of uniformly distributed i.i.d. nodes, $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ the basis matrix for the Chebyshev polynomials with respect to an index set $\mathcal{I}(U)$ of type* (15) *such that $U \subseteq U_{d_s}$ for superposition threshold $d_s \in \mathcal{D}$, and $\boldsymbol{H}$ as in* (21)*. Then for the entries of the expected value matrix*

$$\boldsymbol{E} \coloneqq \mathbb{E}\left[\frac{1}{|\mathcal{X}|}\boldsymbol{H}\right] \in \mathbb{R}^{|I(U)|,|I(U)|} \tag{24}$$

we have

$$|\delta_{\boldsymbol{k},\boldsymbol{\ell}} - (\boldsymbol{E})_{\boldsymbol{k},\boldsymbol{\ell}}| \leq 2 \cdot 4^{d_s} \cdot \left(\frac{\arccos(1-\vartheta)}{\pi}\right)^d.$$

Proof. The entries of $\boldsymbol{E}$ are given by

$$(\boldsymbol{E})_{\boldsymbol{k},\boldsymbol{\ell}} = \int_{[-1+\vartheta,1-\vartheta]^d} \omega(\boldsymbol{x}) T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}.$$

We may rewrite this integral as

$$\begin{aligned}(\boldsymbol{E})_{\boldsymbol{k},\boldsymbol{\ell}} &= \int_{[-1,1]^d} \omega(\boldsymbol{x}) T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} - \int_{[-1,-1+\vartheta]^d} \omega(\boldsymbol{x}) T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\ &\qquad - \int_{[1-\vartheta,1]^d} \omega(\boldsymbol{x}) T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\ &= \delta_{\boldsymbol{k},\boldsymbol{\ell}} - \int_{[-1,-1+\vartheta]^d} \omega(\boldsymbol{x}) T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} - \int_{[1-\vartheta,1]^d} \omega(\boldsymbol{x}) T_{\boldsymbol{k}}(\boldsymbol{x}) T_{\boldsymbol{\ell}}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}\end{aligned}$$

and apply Lemma 4.2, and Lemma 4.3 to obtain the result. □

It remains to consider the eigenvalues of this expected value matrix $\boldsymbol{E}$ from (24).

Lemma 4.5. *Let $\mathcal{X} \subseteq [-1+\vartheta, 1-\vartheta]^d$, $0 < \vartheta < 1$, be a set of uniformly distributed i.i.d. nodes, $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ the basis matrix for the Chebyshev polynomials with respect to an index set $\mathcal{I}(U)$ of type* (15) *such that $U \subseteq U_{d_s}$ for superposition threshold $d_s \in \mathcal{D}$, and $\boldsymbol{E}$ as in* (24)*. Then for every eigenvalue $\lambda \in \mathbb{R}$ of $\boldsymbol{E}$ it holds*

$$|1-\lambda| < 4^{d_s} \left(\frac{\arccos(1-\vartheta)}{\pi}\right)^d |I(U)|.$$

Proof. Theorem 4.4 tells us that we may write $\boldsymbol{E} = \boldsymbol{I} + \boldsymbol{P}$ with the identity matrix $\boldsymbol{I}$ and a perturbation matrix $\boldsymbol{P}$. Applying the theorem of Bauer and Fike, we immediately obtain the result. □

Corollary 4.6. *Let $\mathcal{X} \subseteq [-1+\vartheta, 1-\vartheta]^d$, $0 < \vartheta < 1$, be a set of uniformly distributed i.i.d. nodes, $\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ the basis matrix for the Chebyshev polynomials with respect to an index set $\mathcal{I}(U)$ of type* (15) *such that $U \subseteq U_{d_s}$ for superposition threshold $d_s \in \mathcal{D}$, and $\boldsymbol{H}$ as in* (21)*. Then for every eigenvalue $\lambda \in \mathbb{R}$ of $\frac{1}{|\mathcal{X}|}\boldsymbol{H}$ we have*

$$|1-\lambda| < \frac{1}{2} + 4^{d_s} \kappa(\delta,\vartheta) \left(\frac{\arccos(1-\vartheta)}{\pi}\right)^d \frac{|\mathcal{X}|}{\log(2\,|X|)}$$

with probability $1-\delta$ if

$$|\mathcal{I}(U)| \leq \kappa(\delta,\vartheta) \frac{|\mathcal{X}|}{\log(2\,|\mathcal{X}|)}, \quad \kappa(\delta,\vartheta) \coloneqq \frac{\left(2\vartheta - \vartheta^2\right)^{\frac{d}{2}}}{2^{d_s} \cdot 48(\sqrt{2} - \log\delta)}.$$

Proof. We apply the concentration inequality [29, Proposition 4.1] and note that

$$M^2 \leq \sup_{\boldsymbol{x}\in\mathcal{X}} \sum_{\boldsymbol{k}\in\mathcal{I}(U)} \left|\sqrt{\omega(\boldsymbol{x})} T_{\boldsymbol{k}}(\boldsymbol{x})\right|^2 \leq \frac{2^{d_s}\,|\mathcal{I}(U)|}{\sqrt{2\vartheta - \vartheta^2}}.$$

Setting $t \coloneqq 0.5$ in the inequality yields the result after rearranging. □

Corollary 4.6 now tells us that the singular values τ_i, $i = 1, \dots, |I(U)|$, of $\boldsymbol{W}\boldsymbol{F}(\mathcal{X}, \mathcal{I}(U))$ are bounded

$$\sqrt{|\mathcal{X}| \left(\frac{1}{2} + \gamma\right)} \leq \tau_i \leq \sqrt{|\mathcal{X}| \left(\frac{3}{2} + \gamma\right)}$$

with

$$\gamma \coloneqq 4^{d_s} \kappa(\delta, \vartheta) \left(\frac{\arccos(1 - \vartheta)}{\pi}\right)^d \frac{|\mathcal{X}|}{\log(2\,|X|)}.$$

Moreover, the norm of the Moore-Penrose inverse is also bounded

$$\left\|\boldsymbol{H}^{-1}\boldsymbol{F}^\top(\mathcal{X}, \mathcal{I}(U))\boldsymbol{W}\right\|_2 < \sqrt{|\mathcal{X}| \left(\frac{1}{2} + \gamma\right)}.$$

4.2 Active Set Detection

We determine an initial approximation on the function f by the partial sum $S(\mathcal{X}, \mathcal{I}(U_{d_s}))f(\boldsymbol{x})$, cf. (19), for a chosen superposition threshold $d_s \in \mathcal{D}$ using the method described in Section 4.1. The set $\mathcal{I}(U_{d_s})$ is formed with the sets (15) and order-dependent parameters $N_{|\boldsymbol{u}|} \in \mathbb{N}$ such that $N_{\boldsymbol{u}} = N_{|\boldsymbol{u}|}$. In order to understand the structure of f, i.e., find the important ANOVA terms $f_{\boldsymbol{u}}$, we perform a sensitivity analysis using the global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}, \mathcal{I}(U_{d_s}))f)$. A sensitivity analysis with the indices from the approximation $S(\mathcal{X}, \mathcal{I}(U_{d_s}))f$ of course only makes sense if they behave similarly to the function, i.e., the assumption

$$\varrho(\boldsymbol{u}_1, S(\mathcal{X}, \mathcal{I}(U_{d_s}))f) \leq \varrho(\boldsymbol{u}_2, S(\mathcal{X}, \mathcal{I}(U_{d_s}))f) \implies \varrho(\boldsymbol{u}_1, f) \leq \varrho(\boldsymbol{u}_2, f). \tag{25}$$

holds for $\boldsymbol{u}_1, \boldsymbol{u}_2 \in U_{d_s}$. The accuracy of this assumption may depend on multiple factors like the size of the index set $\mathcal{I}(U_{d_s})$, the underlying function and the number of samples, but numerical experiments suggest that we can rely on these indices even for small index sets $\mathcal{I}(U_{d_s})$. We then use a threshold vector $\boldsymbol{\varepsilon} \in (0, 1)^{d_s}$ to form an active set of ANOVA terms

$$U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}) \coloneqq \emptyset \cup \left\{\boldsymbol{u} \subseteq \mathcal{D} \setminus \emptyset \colon \varrho(\boldsymbol{u}, S(\mathcal{X}, \mathcal{I}(U_{d_s}))f) > \varepsilon_{|\boldsymbol{u}|}\right\}. \tag{26}$$

The inclusion condition (11) is fulfilled if we assume that for all $\boldsymbol{v} \subseteq \boldsymbol{u}$ with $\boldsymbol{u} \in U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ and $\boldsymbol{v} \notin U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ we have $f_{\boldsymbol{v}} \equiv 0$.

This active set $U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ is then used to build a corresponding grouped index set $\mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))$, see (13), with finite frequency sets $\mathcal{I}_{\boldsymbol{u}}$ and parameters $N_{\boldsymbol{u}}$ as in (15). Depending on the information from the global sensitivity indices one may choose to vary the number of frequencies for terms of the same order, i.e., choose different parameters $N_{\boldsymbol{u}}$ and $N_{\boldsymbol{v}}$ for two sets $\boldsymbol{u}$ and $\boldsymbol{v}$ with $|\boldsymbol{u}| = |\boldsymbol{v}|$.

We obtain the approximate partial sum $S(\mathcal{X}, \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})))f(\boldsymbol{x})$ by applying the method from Section 4.1 again. The benefit of this second approximation is that through the smaller number of ANOVA terms we have a reduced model complexity and we may use more frequencies per remaining ANOVA term in our set $\mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))$. We obtain the final approximation

$$f(\boldsymbol{x}) \approx S(\mathcal{X}, \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})))f(\boldsymbol{x}) \coloneqq \sum_{\boldsymbol{k} \in \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))} \hat{f}_{\boldsymbol{k}}\, \varphi_{\boldsymbol{k}}(\boldsymbol{x}).$$

Algorithm 1 and Algorithm 2 summarize the proposed method.

Using the iterative least-squares method LSQR, cf. [30], the arithmetic complexity of an iteration is determined by the matrix-vector multiplications. The following results show the precise complexity of one iteration if we focus on the Chebyshev system.

Theorem 4.7. *Let $\mathrm{L}_2([-1, 1]^d, \omega)$ be the weighted Lebesgue space with Chebyshev product weight ω and the Chebyshev system $\{T_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \mathbb{N}_0^d}$ as orthonormal basis. Moreover, let $\mathcal{I}(U)$ for $U \subseteq \mathcal{P}(\mathcal{D})$ be formed with the sets* (15) *and parameters $N_{\boldsymbol{u}} \in \mathbb{N}$, $\boldsymbol{u} \in U$. Then each iteration of the LSQR algorithm to solve the minimization problem* (17) *or* (20) *with node set $\mathcal{X}$ and evaluations $\boldsymbol{y} \in \mathbb{R}^{|\mathcal{X}|}$ has a complexity in*

$$\mathcal{O}\left(\sum_{\boldsymbol{u} \in U} N_{\boldsymbol{u}}^{|\boldsymbol{u}|} \log N_{\boldsymbol{u}} + |\mathcal{X}|\,|U|\right).$$

Proof. During each iteration of the least-squares algorithm [30] there are two matrix multiplications, one with $\boldsymbol{F}$ and one with $\boldsymbol{F}^\top$. We have to compute one nonequispaced fast cosine transform (NFCT) and one adjoined nonequispaced fast cosine transform for each ANOVA term $f_{\boldsymbol{u}}$, $\boldsymbol{u} \in U$, with complexity in $\mathcal{O}(N_{\boldsymbol{u}}^{|\boldsymbol{u}|} \log N_{\boldsymbol{u}})$ each. Summing over the complexities yields the result. □

Algorithm 1 ANOVA Approximation Method with nodes distributed according to the density ω

Input: $\mathcal{X} \subseteq [-1,1]^d$ finite node set distributed i.i.d. according to probability density ω
$\boldsymbol{y}$ evaluation vector
$d_s \in \mathcal{D}$ superposition threshold

1: Choose finite order-dependent parameters $N_i \in \mathbb{N}$, $i = 1, 2, \ldots, d_s$ as in (15).
2: Compute approximation $S(\mathcal{X}, \mathcal{I}(U(d_s)))f$ by solving

$$\hat{\boldsymbol{f}}_{\text{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{I}(U_{d_s})} \leftarrow \arg\min_{\hat{\boldsymbol{f}}} \left\| \boldsymbol{y} - \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U_{d_s}))\hat{\boldsymbol{f}} \right\|_2^2 .$$

3: Compute global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}, \mathcal{I}(U(d_s)))f)$ for approximation $S(\mathcal{X}, \mathcal{I}(U(d_s)))f$ using (8).
4: Choose threshold vector $\boldsymbol{\varepsilon} \in (0,1)^{d_s}$ and build active set $U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ according to (26).
5: Use information from global sensitivity indices to choose parameters $N_{\boldsymbol{u}} \in \mathbb{N}$ for every ANOVA term in $U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ to obtain $\mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))$.
6: Compute approximation $S(\mathcal{X}, \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})))f$ by solving

$$\hat{\boldsymbol{f}}_{\text{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))} \leftarrow \arg\min_{\hat{\boldsymbol{f}}} \left\| \boldsymbol{y} - \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})))\hat{\boldsymbol{f}} \right\|_2^2 .$$

Output: $\hat{f}_{\boldsymbol{k}} \in \mathbb{R}, \boldsymbol{k} \in \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))$ approximations to basis coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$

Algorithm 2 ANOVA Approximation Method with nodes distributed uniformly

Input: $\mathcal{X} \subseteq [-1,1]^d$ finite node set distributed uniformly
$\boldsymbol{y}$ evaluation vector
$d_s \in \mathcal{D}$ superposition threshold
$\vartheta \in (0,1)$ padding parameter

1: Choose finite order-dependent parameters $N_i \in \mathbb{N}$, $i = 1, 2, \ldots, d_s$ as in (15).
2: Scale nodes $\mathcal{X}$ into interval $[-1+\vartheta, 1-\vartheta]^d$.
3: Set $\boldsymbol{W} = \operatorname{diag}(\boldsymbol{w})$ with $\boldsymbol{w} = (\omega(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$.
4: Compute approximation $S(\mathcal{X}, \mathcal{I}(U(d_s)))f$ by solving

$$\hat{\boldsymbol{f}}_{\text{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{I}(U_{d_s})} \leftarrow \arg\min_{\hat{\boldsymbol{f}}} \left\| \boldsymbol{y} - \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U_{d_s}))\hat{\boldsymbol{f}} \right\|_{2,\boldsymbol{W}}^2 .$$

5: Compute global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}, \mathcal{I}(U(d_s)))f)$ for approximation $S(\mathcal{X}, \mathcal{I}(U(d_s)))f$ using (8).
6: Choose threshold vector parameter $\boldsymbol{\varepsilon} > 0$ and build active set $U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ according to (26).
7: Use information from global sensitivity indices to choose parameters $N_{\boldsymbol{u}} \in \mathbb{N}$ for every ANOVA term in $U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})$ to obtain $\mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))$.
8: Compute approximation $S(\mathcal{X}, \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})))f$ by solving

$$\hat{\boldsymbol{f}}_{\text{sol}} = (\hat{f}_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))} \leftarrow \arg\min_{\hat{\boldsymbol{f}}} \left\| \boldsymbol{y} - \boldsymbol{F}(\mathcal{X}, \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon})))\hat{\boldsymbol{f}} \right\|_{2,\boldsymbol{W}}^2 .$$

Output: $\hat{f}_{\boldsymbol{k}} \in \mathbb{R}, \boldsymbol{k} \in \mathcal{I}(U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}))$ approximations to basis coefficients $\mathrm{c}_{\boldsymbol{k}}(f)$

Corollary 4.8. *Let $\mathrm{L}_2([-1,1]^d, \omega)$ be the weighted Lebesgue space with Chebyshev product weight ω and the Chebyshev system $\{T_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^d}$ as orthonormal basis. Moreover, let $\mathcal{I}(U)$ for $U = U_{d_s} \subseteq \mathcal{P}(\mathcal{D})$ with superposition threshold $d_s \in \mathcal{D}$ be formed with the sets* (15) *and order-dependent parameters $N_{\boldsymbol{u}} = N_{|\boldsymbol{u}|} \in \mathbb{N}$, $\boldsymbol{u} \in U$. Then each iteration of the algorithm to solve the minimization problem* (17) *or* (20) *with node set $\mathcal{X}$ and evaluations $\boldsymbol{y} \in \mathbb{R}^{|\mathcal{X}|}$ has a complexity in*

$$\mathcal{O}\left(d^{d_s}\left(N_{d_s}^{d_s}\log N_{d_s} + |\mathcal{X}|\right)\right)$$

if $N_{d_s} = \max_{j=1,2,\ldots,d_s} N_j$.

Proof. This follows directly from the previous theorem and the estimate $|U_{d_s}| \leq (\mathrm{e} \cdot d/d_s)^{d_s}$ from (12). □

Remark 4.9. *The proposed method is in principle related to sparse polynomial approximation, see e.g. [31]. The first step of considering ANOVA terms of order up to the superposition dimension d_s is equal to considering the basis functions $\varphi_{\boldsymbol{k}}$ with $\|\boldsymbol{k}\|_0 \leq d_s$. We combine this with fast algorithms for the solution of the corresponding least-squares problems that are able to deal with scattered data. Our approach also differs in the fact that we use the importance of ANOVA terms with global sensitivity indices to characterize important basis functions.*

5 Numerical Experiments

In this section we apply the proposed approximation method to high-dimension benchmark functions. We start with an 8-dimensional function that is the sum of products of B-splines in Section 5.1. A similar function has been considered in [4]. In Section 5.2 we consider the well-known Friedman benchmark functions which have previously been used an example for a synthetic regression problem, cf. [32, 33, 34, 35]. The method has been implemented as a Julia package [36]. The padding parameter for uniformly distributed nodes is fixed as $\vartheta = 10^{-4}$.

5.1 B-Spline Function

We apply our method to the test function $f\colon [-1,1]^8 \to \mathbb{R}$,

$$f(\boldsymbol{x}) = B_2(x_1)B_4(x_5) + B_2(x_2)B_4(x_6) + B_2(x_3)B_4(x_7) + B_2(x_4)B_4(x_8), \tag{27}$$

where B_2 and B_4 are parts of shifted, scaled and dilated B-splines of order 2 and 4, respectively. In Figure 1 we have visualized the splines B_2 and B_4 which are elements of $\mathrm{L}_2([-1,1],\tilde{\omega})$ with weight $\tilde{\omega}(x) := \pi^{-1} \cdot (1-x^2)^{-1/2}$ such that $\|B_2\|_{\mathrm{L}_2([-1,1],\tilde{\omega})} = \|B_4\|_{\mathrm{L}_2([-1,1],\tilde{\omega})} = 1$. We remark that the basis coefficients $\mathrm{c}_k(B_2)$ and $\mathrm{c}_k(B_4)$ decay like $\sim k^{-3}$ and $\sim k^{-5}$, respectively. Moreover, f is an element of the tensor product space $\mathrm{L}_2([-1,1]^8,\omega)$. As basis we have the normed Chebyshev polynomials of first kind $\{T_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}_0^8}$.

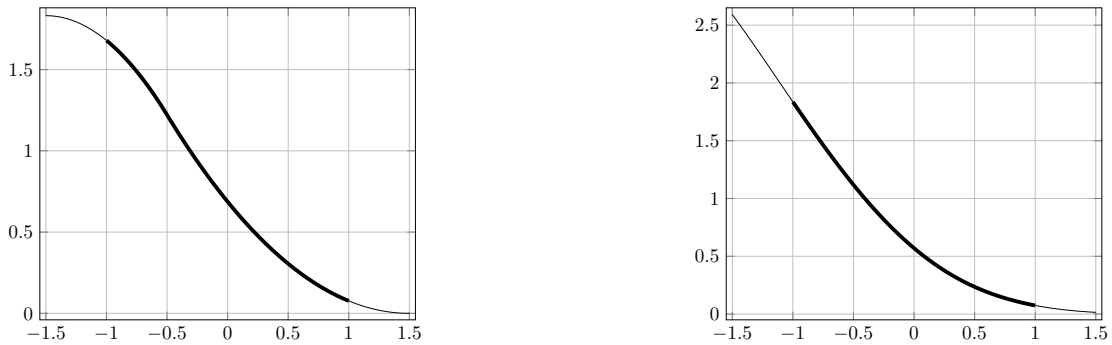


Figure 1: B-splines B_2 (left) and B_4 (right). The relevant part in $[-1,1]$ is highlighted.

The ANOVA terms $f_{\boldsymbol{u}}$ are nonzero for

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

which we call our active set of terms. The function f therefore has a superposition dimension 2 for the accuracy $\delta = 1$, cf. (10), i.e., $\mathrm{T}_3 f = f$. This leads to $d_s = 2$ being the optimal choice for the superposition threshold with no error caused by ANOVA truncation. In a scattered data scenario with an unknown function f this information is of course not known. For our numerical experiments we fix two sampling sets, $\mathcal{X}_{\text{uni}} \subseteq (-1,1)^9$ with uniformly distributed nodes and $\mathcal{X}_{\text{cheb}} \subseteq [-1,1]^9$ with nodes distributed according to ω. Moreover, we have $M := |\mathcal{X}_{\text{cheb}}| = |\mathcal{X}_{\text{uni}}| = 10000$, and an evaluation vector $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$. In the following, we always choose the superposition threshold $d_s = 2$.

Our first goal is to detect the ANOVA terms in U^*. To this end, we use the first step of our method and choose a frequency index set $\mathcal{I}(U_{d_s}) \subseteq \mathbb{N}_0^8$ through order-dependent sets $\mathcal{I}_\emptyset = \{\mathbf{0}\}$,

$$\mathcal{I}_{\boldsymbol{u}} = \{\boldsymbol{k} \in \mathbb{N}_0^d \colon \boldsymbol{k}_{\boldsymbol{u}^{\mathrm{c}}} = \mathbf{0}, k_j = 1, \ldots, N_{|\boldsymbol{u}|} - 1, j \in \boldsymbol{u}\}$$

with $N_1, N_2 \in \mathbb{N}$. We consider the two errors

$$\varepsilon_{\ell_2}^{\mathcal{X}}(f,\tilde{f}) = \frac{1}{\|\boldsymbol{y}\|_2} \sqrt{\sum_{\boldsymbol{x}\in\mathcal{X}} \left| f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x}) \right|^2}$$

and

$$\varepsilon_{\mathrm{L}_2}(f,\tilde{f}) = \frac{1}{\|f\|_{\mathrm{L}_2([-1,1]^8,\omega)}} \left\|f-\tilde{f}\right\|_{\mathrm{L}_2([-1,1]^8,\omega)}$$

where $\tilde{f}$ is an approximation on f. Here, the error $\varepsilon_{\ell_2}^{\mathcal{X}}$ can be regarded as a training error since it is taken at a given sampling set $\mathcal{X}$ and the error $\varepsilon_{\mathrm{L}_2}$ as a generalization error since it measures the error in the basis coefficients.

Our goal is to find the important ANOVA terms, i.e., the terms in U^*. In order to achieve this we expect to have intervals $I_j \subseteq (0,1)$, $j=1,2$, in which to choose the threshold vector $\boldsymbol{\varepsilon}$ with $\varepsilon_j \in I_j$ such that

$$U_{\mathcal{X},\boldsymbol{y}}(\boldsymbol{\varepsilon}) = U^*.$$

The results of our numerical experiments with the function f from (27) are displayed in Table 1.

Remark 5.1. *The norm occuring in the error $\varepsilon_{\mathrm{L}_2}$ can be calculated using Parselval's identity*

$$\|f - S(\mathcal{I}(U),\mathcal{X})f\|^2_{\mathrm{L}_2([-1,1]^d,\omega)} = \|f\|^2_{\mathrm{L}_2([-1,1]^d,\omega)} + \sum_{\boldsymbol{k}\in\mathcal{I}(U)} \left|\mathrm{c}_{\boldsymbol{k}}(f) - \hat{f}_{\boldsymbol{k}}\right|^2 - \sum_{\boldsymbol{k}\in\mathcal{I}(U)} \left|\mathrm{c}_{\boldsymbol{k}}(f)\right|^2.$$

This is of course only possible if the original coefficients and the norm of the function f is known.

size of index sets			relative errors $\mathcal{X}_{\text{cheb}}$		relative errors $\mathcal{X}_{\text{uni}}$	
N_1	N_2	$\lvert\mathcal{I}(U_2)\rvert$	$\varepsilon_{\ell_2}^{\mathcal{X}_{\text{cheb}}}(f,\tilde{f}_1)$	$\varepsilon_{\mathrm{L}_2}(f,\tilde{f}_1)$	$\varepsilon_{\ell_2}^{\mathcal{X}_{\text{uni}}}(f,\tilde{f}_2)$	$\varepsilon_{\mathrm{L}_2}(f,\tilde{f}_2)$
20	8	1525	$5.1\cdot10^{-4}$	$6.9\cdot10^{-4}$	$5.3\cdot10^{-4}$	$8.9\cdot10^{-4}$
20	12	3541	$1.5\cdot10^{-4}$	$4.1\cdot10^{-4}$	$3.2\cdot10^{-4}$	$5.1\cdot10^{-3}$
20	16	6453	$6.8\cdot10^{-5}$	$3.9\cdot10^{-4}$	$2.8\cdot10^{-3}$	$2.6\cdot10^{-1}$
20	20	10261	$3.3\cdot10^{-3}$	$1.6\cdot10^{-1}$	$2.8\cdot10^{-3}$	$5.4\cdot10^{-1}$
40	8	1685	$5.0\cdot10^{-4}$	$6.9\cdot10^{-4}$	$5.2\cdot10^{-4}$	$9.0\cdot10^{-4}$
40	12	3701	$1.4\cdot10^{-4}$	$4.0\cdot10^{-4}$	$3.8\cdot10^{-4}$	$6.7\cdot10^{-3}$
40	16	6613	$5.7\cdot10^{-5}$	$3.8\cdot10^{-4}$	$2.9\cdot10^{-3}$	$2.8\cdot10^{-1}$
40	20	10421	$1.3\cdot10^{-4}$	$2.0\cdot10^{-1}$	$2.6\cdot10^{-3}$	$5.7\cdot10^{-1}$

Table 1: Results of the detection step for important ANOVA terms of f with $M = 10000$ Chebyshev distributed nodes $\mathcal{X}_{\text{cheb}}$ and uniformly distributed nodes $\mathcal{X}_{\text{uni}}$. We define $\tilde{f}_1 \coloneqq S(\mathcal{X}_{\text{cheb}},\mathcal{I}(U_2))f$ and $\tilde{f}_2 \coloneqq S(\mathcal{X}_{\text{uni}},\mathcal{I}(U_2))f$.

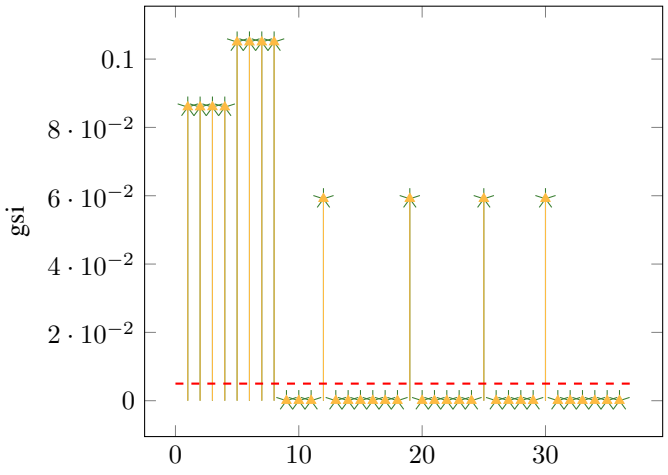


Figure 2: Behavior of the global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}_{\text{cheb}},\mathcal{I}(U_2))f)$ in green and $\varrho(\boldsymbol{u}, S(\mathcal{X}_{\text{uni}},\mathcal{I}(U_2))f)$ in orange for parameters $N_1 = 20$, $N_2 = 8$.

We observe that increasing the parameters N_1, N_2 will ultimately lead to an increasing generalization error $\varepsilon_{\mathrm{L}_2}$. This is consistent with our results in Section 4.1 since we cannot guarantee with enough certainty that the system matrix in the normal equations has good properties if the index set size increases beyond a certain point. This effect appears sooner with the uniform nodes $\mathcal{X}_{\text{uni}}$ which is connected to the necessary preconditioning, see Section 4.1.2. In Figure 2 we have visualized the global sensitivity indices for the parameters $N_1 = 20$ and $N_2 = 8$. The terms in U^* are clearly separated from the terms in its complement, i.e., the active set detection worked well for this example. Moreover, we observe that the gsi obtained by approximation with uniform nodes and with Chebyshev nodes are very close together.

There clearly exist N_1, N_2, and $\boldsymbol{\varepsilon}$, e.g. $N_1 = 20$, $N_2 = 8$, $\boldsymbol{\varepsilon} = (0.005, 0.005)$, such that we are able to recover the set of ANOVA terms, i.e., $U_{\mathcal{X}_{\text{uni}},\boldsymbol{y}}(\boldsymbol{\varepsilon}) = U_{\mathcal{X}_{\text{cheb}},\boldsymbol{y}}(\boldsymbol{\varepsilon}) = U^*$. We aim to improve our approximation quality with the given data by solving the minimization problem with the active set U^*. Here, we could choose individual index sets for every ANOVA term in U^* to form $\mathcal{I}(U^*)$ based on the global sensitivity indices, but for our function order-dependence can be maintained. Table 2 shows the results of the approximation using the index set $\mathcal{I}(U^*)$.

size of index sets			relative errors $\mathcal{X}_{\text{cheb}}$		relative errors $\mathcal{X}_{\text{uni}}$	
N_1	N_2	$\lvert I(U_{d_s})\rvert$	$\varepsilon_{\ell_2}^{\mathcal{X}_{\text{cheb}}}(f,\tilde{f}_1)$	$\varepsilon_{\mathrm{L}_2}(f,\tilde{f}_1)$	$\varepsilon_{\ell_2}^{\mathcal{X}_{\text{uni}}}(f,\tilde{f}_2)$	$\varepsilon_{\mathrm{L}_2}(f,\tilde{f}_2)$
60	12	957	$1.6 \cdot 10^{-4}$	$3.8 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$4.1 \cdot 10^{-4}$
60	20	1917	$4.5 \cdot 10^{-5}$	$3.4 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$4.1 \cdot 10^{-4}$
60	28	3389	$1.8 \cdot 10^{-5}$	$3.4 \cdot 10^{-4}$	$6.9 \cdot 10^{-4}$	$6.9 \cdot 10^{-2}$
80	12	1117	$1.6 \cdot 10^{-4}$	$3.8 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$4.2 \cdot 10^{-4}$
80	20	2077	$4.5 \cdot 10^{-5}$	$3.4 \cdot 10^{-4}$	$7.1 \cdot 10^{-4}$	$7.2 \cdot 10^{-2}$
80	28	3549	$1.8 \cdot 10^{-5}$	$3.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-3}$	$1.9 \cdot 10^{-1}$

Table 2: Results of approximation for important ANOVA terms of f with $M = 10000$ Chebyshev distributed nodes $\mathcal{X}_{\text{cheb}}$ and uniformly distributed nodes $\mathcal{X}_{\text{uni}}$. We define $\tilde{f}_1 \coloneqq S(\mathcal{X}_{\text{cheb}}, \mathcal{I}(U^*))f$ and $\tilde{f}_2 \coloneqq S(\mathcal{X}_{\text{uni}}, \mathcal{I}(U^*))f$.

We observe that the reduction of the ANOVA terms to U^* yields a benefit with regard to the approximation quality. This results from the reduction in model complexity, i.e., we have larger oversampling factors for the same node set. The comparison of Chebyshev and uniform nodes yields a similar behavior as for the detection step. We achieve better errors for the Chebyshev distributed nodes without the addition of preconditioning which was to be expected since we choose a different sampling distribution for a weighted space.

5.2 Friedman Benchmark Functions

The Friedmann functions are a well-known example for the approximation of functions with scattered data, see e.g. [32, 33, 34]. We define the three non-periodic Friedmann functions as

$$\tilde{f}_1 \colon [0,1]^{10} \to \mathbb{R}, \qquad f_1(\boldsymbol{x}) = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5$$

$$\tilde{f}_2 \colon [0,1]^{4} \to \mathbb{R}, \qquad f_2(\boldsymbol{x}) = \sqrt{s_1^2(x_1) + \left(s_2(x_2)\cdot x_3 - \frac{1}{s_2(x_2)\cdot s_4(x_4)}\right)^2}$$

$$\tilde{f}_3 \colon [0,1]^{4} \to \mathbb{R}, \qquad f_3(\boldsymbol{x}) = \arctan\left(\frac{s_2(x_2)\cdot x_3 - (s_2(x_2)\cdot s_4(x_4))^{-1}}{s_1(x_1)}\right)$$

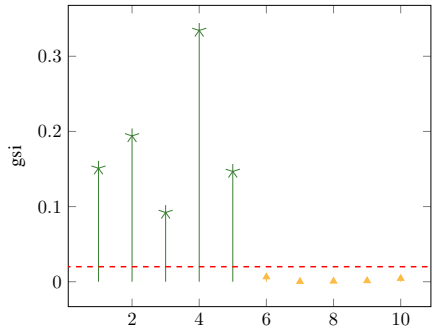
with variable scalings $s_1(x_1) = 100x_1$, $s_2(x_2) = 520\pi x_2 + 40\pi$, and $s_4(x_4) = 10x_4 + 1$. The function f_1 has spatial dimension 10. However, only five of the ten variables have any influence on the function. For f_1 and f_2 we also do not have more than two variables interact simultaneously, i.e., the superposition dimension for accuracy $\delta = 1$ is 2, cf. (10). For f_3 the superposition dimension for accuracy $\delta = 1$ would be equal to the spatial dimension 4. Since the original functions are given on the interval $[0,1]^d$ we define for our experiments

$$f_i \colon [-1,1]^{d^{(i)}} \to \mathbb{R},\ f_i(\boldsymbol{x}) = \tilde{f}_i(0.5(x_1+1), 0.5(x_2+1), \ldots, 0.5(x_{d_i}+1)),$$

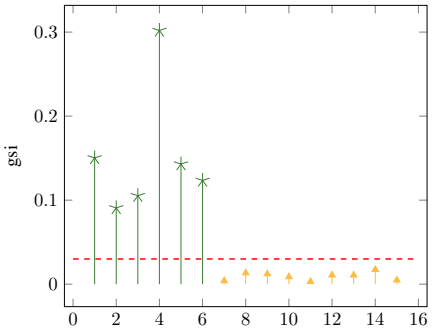
$i = 1, 2, 3$ with $d^{(1)} = 10$ and $d^{(2)} = d^{(3)} = 4$.

For calculation the approximations we use three sets of uniformly distributed nodes $\mathcal{X}_1 \subseteq (-1,1)^1 0$, $\mathcal{X}_2 \subseteq (-1,1)^4$, and $\mathcal{X}_3 \subseteq (-1,1)^4$ with $|\mathcal{X}_i| = 200$. Moreover, we evaluate the functions at those nodes and additionally add Gaussian noise, i.e.,

$$\boldsymbol{y}_i = (f(\boldsymbol{x}) + \eta_i)_{\boldsymbol{x}\in\mathcal{X}_i}, i = 1, 2, 3,$$



(a) Global sensitivity indices $\varrho(\{i\}, S(\mathcal{X}_1, \mathcal{I}(U_2))f_1)$, $i = 1, 2, 3, 4, 5$, in green and $\varrho(\{i\}, S(\mathcal{X}_1, \mathcal{I}(U_2))f_1)$, $i = 6, 7, 8, 9, 10$, in orange for parameters $N_1 = 4$, $N_2 = 2$.



(b) Global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}_1, \mathcal{I}(\tilde{U}_2))f_1)$, $\boldsymbol{u} \in U_1^*$ in green and $\boldsymbol{u} \in \tilde{U}_2 \setminus U_1^*$ in yellow with parameters $N_1 = N_2 = 4$.

Figure 3: Numerical experiments with the Friedman 1 function.

where the noise η_i has zero mean and variances $\sigma_1 = 1$, $\sigma_2 = 125$, $\sigma_3 = 0.1$, respectively. In order to measure the error of an approximation g_i on a Friedman function f_i, we consider the mean square error (mse)

$$\operatorname{mse}(f_i, g_i) = \frac{1}{1000} \sum_{\boldsymbol{x} \in \overline{\mathcal{X}}_i} |f_i(\boldsymbol{x}) + \eta_i - g_i(\boldsymbol{x})|^2 \tag{28}$$

on additional sets of nodes $\overline{\mathcal{X}}_1 \subseteq (-1, 1)^{10}$, $\overline{\mathcal{X}}_2 \subseteq (-1, 1)^4$, and $\overline{\mathcal{X}}_3 \subseteq (-1, 1)^4$ with $\left|\overline{\mathcal{X}}_i\right| = 1000$.

5.2.1 Friedman 1

The first goal for Friedman 1 is to identify that the variables x_6 to x_{10} have no contribution to the function. To this end we computed the approximation $S(\mathcal{X}_1, \mathcal{I}(U_2))f_1$ and considered the global sensitivity indices of the one-dimensional sets $\{i\}$, $i = 1, 2, \ldots, 10$. The result is depicted in Figure 3a. We observe that the variables x_1 to x_5 can be clearly separated from the rest. Therefore, we proceed with the active set

$$\tilde{U}_2 \coloneqq \{\boldsymbol{u} \subseteq \{1, 2, 3, 4, 5\} \colon |\boldsymbol{u}| \leq 2\}. \tag{29}$$

Our second goal is to find the active set of terms

$$U_1^* = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}\}.$$

To this end, we calculate the approximation $S(\mathcal{X}_1, \mathcal{I}(\tilde{U}_2))f_1$ with parameters $N_1 = N_2 = 4$. This yields a mean square error on the test nodes $\overline{\mathcal{X}}_1$ of 2.88. As depicted in Figure 3b we are able to recover U_1^* with a clear separation by choosing, e.g., $\boldsymbol{\varepsilon} = (0.03, 0.03)$.

We computed the approximation $S(\mathcal{X}, \mathcal{I}(U_1^*))f_1$ with parameters $N_1 = N_2 = 4$ on 100 randomly generated uniformly i.i.d. pairs of node sets $(\mathcal{X}, \mathcal{T}) \subseteq (-1, 1)^{10} \times (-1, 1)^{10}$ with $|\mathcal{X}| = 200$ and $|\mathcal{T}| = 1000$. The mean square error was calculated on the sets $\mathcal{T}$. The median of the 100 mses is 1.17.

5.2.2 Friedman 2

For the Friedman 2 function we want to identify the active set of terms from our scattered data $\mathcal{X}_2$ and $\boldsymbol{y}_2$. To this end, we computed the approximation $S(\mathcal{X}_2, \mathcal{I}(U_2))f_2$ with parameters $N_1 = N_2 = 2$. This yielded a mse on the data $\overline{\mathcal{X}}_2$ of $16.44 \cdot 10^3$. The resulting sensitivity indices are displayed in Figure 4. We deduce that the terms

$$U_2^* \coloneqq \{\emptyset, \{2\}, \{3\}, \{2, 3\}\} \tag{30}$$

are clearly more important than the rest. They can be obtained by choosing a threshold vector, e.g., $\boldsymbol{\varepsilon} = (0.03, 0.03)$.

We computed the approximation $S(\mathcal{X}, \mathcal{I}(U_2^*))f_2$ with parameters $N_1 = N_2 = 2$ on 100 randomly generated uniform i.i.d. pairs of node sets $(\mathcal{X}, \mathcal{T}) \subseteq (-1, 1)^{10} \times (-1, 1)^{10}$ with $|\mathcal{X}| = 200$ and $|\mathcal{T}| = 1000$. The mean square error was calculated on the sets $\mathcal{T}$. The median of the 100 mses is $16.09 \cdot 10^3$.

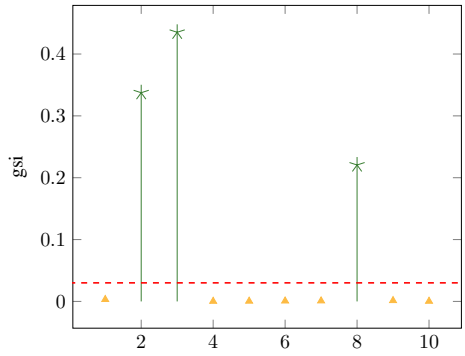


Figure 4: Global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}_2, \mathcal{I}(U_2))f_2)$, $\boldsymbol{u} \in U_2^*$ in green and $\boldsymbol{u} \in U_2 \setminus U_2^*$ in yellow with parameters $N_1 = N_2 = 2$.

5.2.3 Friedman 3

As before, we first aim to identify an active set of terms from the scattered data $\mathcal{X}_3$ and $\boldsymbol{y}_3$. The approximation $S(\mathcal{X}_3, \mathcal{I}(U_2))f_3$ with parameters $N_1 = 8$, $N_2 = 2$ yielded a mean square error of $1.8 \cdot 10^{-2}$ on the node set $\overline{X}_3$. The sensitivity indices are displayed in Figure 5. From this we identify the active set as

$$U_3^* := \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}\}, \tag{31}$$

e.g., with a choice of $\boldsymbol{\varepsilon} = (0.002, 0.002)$.

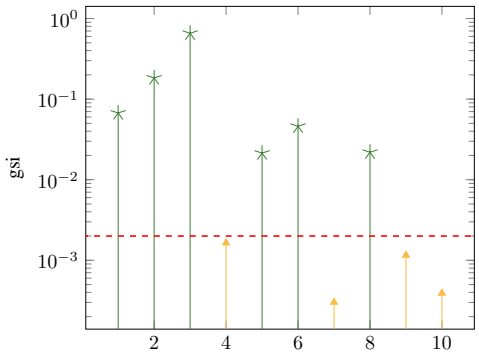


Figure 5: Global sensitivity indices $\varrho(\boldsymbol{u}, S(\mathcal{X}_3, \mathcal{I}(U_2))f_3)$, $\boldsymbol{u} \in U_3^*$ in green and $\boldsymbol{u} \in U_2 \setminus U_3^*$ in yellow with parameters $N_1 = 8$, $N_2 = 2$.

We performed the approximation $S(\mathcal{X}, \mathcal{I}(U_3^*))f_3$ with parameters $N_1 = 8$, $N_2 = 2$ on 100 randomly generated uniform i.i.d. pairs of node sets $(\mathcal{X}, \mathcal{T}) \subseteq (-1,1)^4 \times (-1,1)^4$ with $|\mathcal{X}| = 200$ and $|\mathcal{T}| = 1000$. The median of the mean square error on the sets $\mathcal{T}$ is $17.22 \cdot 10^{-3}$.

5.2.4 Comparison

Table 3 contains the benchmark data from [32] with a support vector machine (SVM), a linear model (lm), a neural network (mnet) and a random forest (rForst) as well as the results with our method (ANOVAapprox). We are able to achieve a more accurate approximation in the exact same setting for every one of the three functions. The value for ANOVAapprox was obtained by computing the model on 100 randomly generated node sets and computing the error on 100 randomly generated test sets.

6 Summary

In this paper we considered the classical ANOVA decomposition for functions f in weighted Lebesgue spaces $\mathrm{L}_2([-1,1]^d, \omega)$ with orthogonal polynomials as bases. Specifically, we proved relations between the basis coefficients

	svm	lm	mnet	rForst	ANOVAapprox
Friedman 1	4.36	7.71	9.21	6.02	**1.17**
Friedman 2 ($\cdot 10^3$)	18.13	36.15	19.61	21.50	**16.09**
Friedman 3 ($\cdot 10^{-3}$)	23.15	45.42	18.12	22.21	**17.22**

Table 3: Mean squared errors (MSE) for different methods when approximating Friedman functions in [32] compared to our method (ANOVAapprox). All values are the medians of the experiment MSEs and the best value for every function is highlighted.

of the projections $\mathrm{P}_{\boldsymbol{u}} f$, the ANOVA terms $f_{\boldsymbol{u}}$, and the function f. Furthermore, we considered sensitivity analysis and truncating the ANOVA decomposition to a certain subset of terms.

We introduced a method to determine important ANOVA terms, i.e., terms with a high global sensitivity index $\varrho(\boldsymbol{u}, f)$, by approximation with index sets with a low-dimensional structure related to the truncated ANOVA decomposition. Our scenario was scattered data approximation where only a node set $\mathcal{X}$ and possibly noisy function values $\boldsymbol{y} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$ are known. Properties of the corresponding normal equations were considered om the case of nodes distributed according to the Chebyshev density. We also introduced preconditioning for uniformly distributed nodes and considered properties of the resulting system as well.

The numerical experiments show that the method works using a specific test function consisting of sums of products of B-splines. The test function had a superposition dimension of 3 for an arbitrary accuracy, i.e., $\mathrm{T}_3 f = f$, and we were able to recover the active set of ANOVA terms with our approach. Experiments with the Friedman functions showed that we proposed a competitive method that yields better results on these functions as other well-known methods such as support vector machines.

Acknowledgments

We thank Tino Ullrich and Toni Volkmer for fruitful discussions on the contents of this paper. Daniel Potts acknowledges funding by Deutsche Forschungsgemeinschaft (German Research Foundation) – Project–ID 416228727 – SFB 1410. Michael Schmischke is supported by the BMBF grant 01|S20053A.

References

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

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

[3] I. H. Sloan, S. Joe, Lattice methods for multiple integration, Oxford Science Publications, The Clarendon Press Oxford University Press, New York, 1994.

[4] D. Potts, T. Volkmer, Multivariate sparse FFT based on rank-1 chebyshev lattice sampling, in: 2017 International Conference on Sampling Theory and Applications (SampTA), IEEE, 2017. doi:10.1109/sampta.2017.8024341.

[5] G. Plonka, D. Potts, G. Steidl, M. Tasche, Numerical Fourier Analysis, Applied and Numerical Harmonic Analysis, Birkhäuser, 2018. doi:10.1007/978-3-030-04306-3.

[6] C. F. J. Wu, M. S. Hamada, Experiments - Planning, Analysis, and Optimization, John Wiley & Sons, New York, 2011.

[7] R. Caflisch, W. Morokoff, A. Owen, Valuation of mortgage-backed securities using brownian bridges to reduce effective dimension, J. Comput. Finance 1 (1) (1997) 27–46. doi:10.21314/jcf.1997.005.

[8] H. Rabitz, O. F. Alis, General foundations of high dimensional model representations, J. Math. Chem. 25 (1999) 197–233. doi:10.1023/A:1019188517934.

[9] R. Liu, A. B. Owen, Estimating mean dimensionality of analysis of variance decompositions, J. Amer. Statist. Assoc. 101 (474) (2006) 712–721. doi:10.1198/016214505000001410.

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

[11] H.-J. Bungartz, M. Griebel, Sparse grids, Acta Numer. 13 (2004) 147–269. doi:10.1017/s0962492904000182.

[12] M. Griebel, M. Holtz, Dimension-wise integration of high-dimensional functions with applications to finance, J. Complexity 26 (5) (2010) 455–489. doi:10.1016/j.jco.2010.06.001.

[13] J. Baldeaux, M. Gnewuch, Optimal randomized multilevel algorithms for infinite-dimensional integration on function spaces with ANOVA-type decomposition, SIAM J. Numer. Anal. 52 (3) (2014) 1128–1155. doi:10.1137/120896001.

[14] M. Griebel, F. Y. Kuo, I. H. Sloan, The ANOVA decomposition of a non-smooth function of infinitely many variables can have every term smooth, Math. Comp. 86 (306) (2016) 1855–1876. doi:10.1090/mcom/3171.

[15] F. Y. Kuo, D. Nuyens, L. Plaskota, I. H. Sloan, G. W. Wasilkowski, Infinite-dimensional integration and the multivariate decomposition method, J. Comput. Appl. Math. 326 (2017) 217–234. doi:10.1016/j.cam.2017.05.031.

[16] F. Bartel, D. Potts, M. Schmischke, Grouped transformations in high-dimensional explainable ANOVA approximation, ArXiv e-prints 2010.10199 (2020).

[17] D. Potts, G. Steidl, M. Tasche, Fast algorithms for discrete polynomial transforms, Math. Comput. 67 (1998) 1577–1590.

[18] J. R. Driscoll, D. Healy, Computing Fourier transforms and convolutions on the 2–sphere, Adv. in Appl. Math. 15 (1994) 202–250.

[19] J. Keiner, S. Kunis, 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.

[20] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, H. Woźniakowski, On decompositions of multivariate functions, Math. Comp. 79 (270) (2009) 953–966. doi:10.1090/s0025-5718-09-02319-9.

[21] D. Potts, M. Schmischke, Approximation of high-dimensional periodic functions with Fourier-based methods, SIAM J. Numer. Anal. (to appear), arXiv: 1907.11412 [math.NA] (2019). arXiv:1907.11412.

[22] I. M. Sobol, On sensitivity estimation for nonlinear mathematical models, Keldysh AppliedMathematics Institute 1 (1990) 112–118.

[23] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulation 55 (1-3) (2001) 271–280. doi:10.1016/s0378-4754(00)00270-6.

[24] M. Griebel, F. Y. Kuo, I. H. Sloan, The smoothing effect of the ANOVA decomposition, J. Complexity 26 (5) (2010) 523–551. doi:10.1016/j.jco.2010.04.003.

[25] F. Y. Kuo, C. Schwab, I. H. Sloan, Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients, SIAM J. Numer. Anal. 50 (2012) 3351 – 3374. doi:10.1137/110845537.

[26] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, C. Schwab, I. H. Sloan, Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients, Numer. Math. 131 (2) (2014) 329–368. doi:10.1007/s00211-014-0689-y.

[27] F. Y. Kuo, D. Nuyens, Application of quasi-Monte Carlo methods to elliptic PDEs with random diffusion coefficients: A survey of analysis and implementation, Found. Comput. Math 16 (6) (2016) 1631–1696. doi:10.1007/s10208-016-9329-5.
URL https://doi.org/10.1007/s10208-016-9329-5

[28] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, I. H. Sloan, Circulant embedding with QMC: analysis for elliptic PDE with lognormal coefficients, Numer. Math. 140 (2) (2018) 479–511. doi:10.1007/s00211-018-0968-0.

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

[30] C. C. Paige, M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, ACM Trans. Math. Software 8 (1982) 43–71.

[31] A. Chkifa, A. Cohen, C. Schwab, Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDE, J. Math. Pures Appl. 103 (2015) 400 – 428. doi:10.1016/j.matpur.2014.04.009.

[32] D. Meyer, F. Leisch, K. Hornik, The support vector machine under test, Neurocomputing 55 (1) (2003) 169 – 186, support Vector Machines. doi:10.1016/S0925-2312(03)00431-4.

[33] G. Beylkin, J. Garcke, M. Mohlenkamp, Multivariate regression and machine learning with sums of separable functions, SIAM J. Scientific Computing 31 (2009) 1840–1857. doi:10.1137/070710524.

[34] P. Binev, W. Dahmen, P. Lamby, Fast high-dimensional approximation with sparse occupancy trees, J. Comput. Appl. Math. 235 (8) (2011) 2063 – 2076. doi:10.1016/j.cam.2010.10.005.

[35] D. Potts, M. Schmischke, Interpretable approximation of high-dimensional data, ArXiv e-prints 2103.13787 (2021).

[36] F. Bartel, M. Schmischke, ANOVAapprox Julia package, https://github.com/NFFT/ANOVAapprox/ (2020).