ANOVA approximation with mixed tensor product basis on scattered points

Daniel Potts* Pascal Schröter‡

June 15, 2023

Abstract

In this paper we consider an orthonormal basis, generated by a tensor product of Fourier basis functions, half period cosine basis functions, and the Chebyshev basis functions. We deal with the approximation problem in high dimensions related to this basis and design a fast algorithm to multiply with the underlying matrix, consisting of rows of the non-equidistant Fourier matrix, the non-equidistant cosine matrix and the non-equidistant Chebyshev matrix, and its transposed. This leads us to an ANOVA (analysis of variance) decomposition for functions with partially periodic boundary conditions through using the Fourier basis in some dimensions and the half period cosine basis or the Chebyshev basis in others. We consider sensitivity analysis in this setting, in order to find an adapted basis for the underlying approximation problem. More precisely, we find the underlying index set of the multidimensional series expansion. Additionally, we test this ANOVA approximation with mixed basis at numerical experiments, and refer to the advantage of interpretable results.

Key words

ANOVA, high-dimensional, approximation, Fourier approximation, fast Fourier methods, NFFT, Chebychev polynomials

AMS subject classifications

65T, 42B05

1 Introduction

The approximation of functions is a problem that arises in many scientific fields. As soon as data is recorded, questions how “Which correlations are in the data?”, “Which variables are dependent on one another” and “How data can be predicted at other points?” arises.
There are various algorithms as artificial neural networks or support vector machines for approximating functions in high dimensions, see e.g. [1, 8]. But these do not show which dependencies are hidden in the data. For this question of connections in data there is the ANOVA (analysis of variance) decomposition, cf. [5, 20, 12, 11, 9, 6]. The classical ANOVA is based on an integral projection operator. Based on this, we use the analytic global sensitivity indices [23, 24]. These tell us which variables are related and how big the influence of these relations are. We use consequently

*potts@math.tu-chemnitz.de, Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

‡pascal.schroeter@math.tu-chemnitz.de, Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

the series expansion in various basic functions to define the ANOVA decomposition, cf. [16, 17, 21]. A majority of real world systems are dominated by low-complexity interactions of their variables. This principle is known as sparsity-of-effects, see e.g. [25, Section 4.6],[7],[21, Section 4.2]. We use this principle to truncate the ANOVA decomposition. The resulting sums can be evaluated through algorithms like the non-equidistant fast Fourier transform (NFFT) [10],[14, Chapter 7] combined with grouped transforms, cf. [2]. Since we consider a finite truncation of the series expansion for approximation, the properties of the basis functions, such as the periodicity, are reflected in the approximation. It is therefore advantageous if the basis functions have similar properties as the data of the underlying process. Typical applications with such properties are functions with data on spheres $\mathbb{S}^d$ and balls $\mathbb{B}^d$, because their polar coordinates have periodic parts and non-periodic parts. Similar things happen by the approximation of a function with data on the rotation group SO(3). We will combine well known basis functions on $[0,1]$ like the Fourier basis functions $\phi_k^{\exp} := \exp(2\pi \mathrm{i} k \cdot)$, $k \in \mathbb{Z}$, the half period cosine basis functions $\phi_k^{\cos} := \sqrt{2}^{\,1-\delta_{k,0}} \cos(\pi k \cdot)$, $k \in \mathbb{N}_0$, and the Chebyshev basis functions $\phi_k^{\mathrm{alg}} := \sqrt{2}^{\,1-\delta_{k,0}} \cos(k \arccos(2 \cdot -1))$, $k \in \mathbb{N}_0$ in a tensor product structure to achieve more flexibility what properties are present in which dimensions. We denote this tensor product basis functions by $\phi_{\mathbf{k}}^{\mathbf{d}} := \prod_{j=1}^{d} \phi_{k_j}^{d_j}$. Here $\mathbf{d}$ is a vector containing the information which basis is used in which dimension. We assume that it is known from the application which base in which dimension should be used. To compute such approximations it is important to evaluate finite sums of basis functions

$$\sum_{\mathbf{k} \in \mathcal{I}} \hat{f}_{\mathbf{k}} \phi_{\mathbf{k}}(\mathbf{x})$$

with a known index set $\mathcal{I}$ at many nodes $\mathbf{x}_j \in [0,1]^d$ simultaneously. Since real world data is rarely equidistantly sampled, we have to evaluate these sums at arbitrary scattered nodes. For basis which consist of only one of this named one dimensional basis functions in every dimension, there are algorithms to evaluate these sums, see [14, Chapter 7]. In order to get good approximations, it is essential to use appropriate index sets $\mathcal{I}$ in the approximating sums. The choice of such an index set is always a trade-off between the number of indices, the number of training data available, and the needed computation time.
We will make use of the following identities. Firstly, we rewrite a sum of half period cosine basis functions to a sum of Fourier basis functions

$$\sum_{k=0}^{N-1} \hat{f}_k^{\cos} \underbrace{\sqrt{2}^{\,1-\delta_{k,0}} \cos(\pi k x)}_{=\phi_k^{\cos}(x)} = \sum_{k=-N+1}^{N-1} \underbrace{2^{\delta_{k,0}-1} \hat{f}_{|k|}^{\cos} \sqrt{2}^{1-\delta_{k,0}}}_{=:\hat{f}_k^{\exp}} \underbrace{\exp(\pi \mathrm{i} k x)}_{=\phi_k^{\exp}\left(\frac{x}{2}\right)}, \tag{1.1}$$

Secondly, we rewrite a sum of Chebyshev basis functions to a sum of Fourier basis functions

$$\sum_{k=0}^{N-1} \hat{f}_k^{\mathrm{alg}} \underbrace{\sqrt{2}^{\,1-\delta_{k,0}} \cos(k \arccos(2x-1))}_{=\phi_k^{\mathrm{alg}}(x)} = \sum_{k=-N+1}^{N-1} \underbrace{2^{\delta_{k,0}-1} \hat{f}_{|k|}^{\mathrm{alg}} \sqrt{2}^{1-\delta_{k,0}}}_{=:\hat{f}_k^{\exp}} \underbrace{\exp(\mathrm{i} k \arccos(2x-1))}_{=\phi_k^{\exp}\left(\frac{\arccos(2x-1)}{2\pi}\right)}. \tag{1.2}$$

These identities (1.1) and (1.2) can then be applied in every dimension with the half period cosine basis or the Chebyshev basis. This provides us a way to evaluate all these sums through an NFFT. We will use this to compute ANOVA approximations for the mixed tensor product basis for high-dimensional scattered data. In this way we get a model of the data with which we can predict further data. In addition, we can use this model to calculate approximated global sensitivity indices, which allow us to identify correlations in the data. Using these approximated global sensitivity indices, we can truncate our approximation even further to get better suitable index sets $\mathcal{I}$, which provides us even better approximations. We stress again that in many applications a fairly small

number of nodes is enough to get reasonably good approximations.
The paper is organized as follows. In Section 2 we set up notion and terminology. Firstly, in Subsection 2.1 we introduce needed function spaces and some of their relations. Subsection 2.2 introduces some well known orthonormal basis and finally the mixed basis with which we will deal with in the rest of the paper. In Section 3 we introduce the ANOVA approximation for the mixed basis based on an approach by Fourier series. Here, Subsection 3.1 deals with the definition of the ANOVA decomposition and Subsection 3.2 provides a way to compute an ANOVA approximation. We split this subsection in three parts, firstly we consider useful index sets then we describe a way to compute the ANOVA approximation for given index sets and to this end we describe how this index sets could be determined. In Section 4 we develop fast algorithms to evaluate sums of mixed basis functions. Subsection 4.1 provides through Theorem 4.1 a way to compute such sums through an NFFT which is summarized in Algorithm 1. In Subsection 4.2 we extend the grouped transform [2] to the mixed basis using the Algorithm 1. In Section 5 we show with some numerical experiments how this approximation procedure works. Subsection 5.1 deals with the approximation of a function. There are the steps shown to find good suitable index sets. In Subsection 5.2 we approximate a function where we only have access to uniformly sampled nodes. At this point we compare the approximation with a suitable mixed basis with the approximation with the half period cosine basis and the Fourier basis.

2 Preliminaries

This section presents basic definitions for the rest of this paper. In Subsection 2.1 various function spaces and some of their relations are introduced. Subsection 2.2 presents some orthonormal basis. In Definition 2.2 the mixed basis is defined.

2.1 Function Spaces

Let $\mathbb{D} \in \{\mathbb{T}^m \times [0,1]^n \mid m, n \in \mathbb{N}_0\}$ be a measurable set, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the torus which we identify with $[0,1)$. Moreover, let $\omega\colon \mathbb{D} \to (0,\infty)$ be a probability measure with $\int_{\mathbb{D}} \omega(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. We define the weighted Lebesgue spaces

$$\mathrm{L}_p(\mathbb{D},\omega) := \left\{ f\colon \mathbb{D} \to \mathbb{C} \;\middle|\; \int_{\mathbb{D}} |f(\mathbf{x})|^p\,\omega(\mathbf{x})\,\mathrm{d}\mathbf{x} < \infty \right\}$$

with the norm $\|f\|_{\mathrm{L}_p(\mathbb{D},\omega)} := (\int_{\mathbb{D}} |f(\mathbf{x})|^p\,\omega(\mathbf{x})\,\mathrm{d}\mathbf{x})^{\frac{1}{p}}$ for $p \in [1,\infty)$. Furthermore, we define

$$\mathrm{L}_\infty(\mathbb{D}) := \{f\colon \mathbb{D} \to \mathbb{C} \mid \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{D}} |f(\mathbf{x})| < \infty\}$$

with the norm $\|f\|_{\mathrm{L}_\infty(\mathbb{D})} := \operatorname{ess\,sup}_{\mathbf{x}\in\mathbb{D}} |f(\mathbf{x})|$. The Lebesgue space $\mathrm{L}_2(\mathbb{D},\omega)$ forms a Hilbert space with the scalar product $\langle f, g\rangle_{\mathrm{L}_2(\mathbb{D},\omega)} := \int_{\mathbb{D}} f(\mathbf{x})\overline{g(\mathbf{x})}\omega(\mathbf{x})\,\mathrm{d}\mathbf{x}$. We use the abbreviation $\mathrm{L}_p(\mathbb{D}) := \mathrm{L}_p(\mathbb{D},1)$. Let $\mathcal{B} = (\phi_k)_{k\in\mathcal{K}}$ be a basis with an index set $\mathcal{K}$ for the Hilbert space $\mathrm{L}_2(\mathbb{D},\omega)$. We define the Wiener space

$$\mathcal{A}(\mathbb{D},\omega) := \left\{ f \in \mathrm{L}_1(\mathbb{D},\omega) \;\middle|\; \sum_{k\in\mathcal{K}} \left|\langle f, \phi_k\rangle_{\mathrm{L}_2(\mathbb{D},\omega)}\right| < \infty \right\}$$

with the norm $\|f\|_{\mathcal{A}(\mathbb{D},\omega)} := \sum_{k\in\mathcal{K}} |\langle f, \phi_k\rangle_{\mathrm{L}_2(\mathbb{D},\omega)}|$.

Lemma 2.1. *Let $\mathrm{L}_2(\mathbb{D},\omega)$ be a weighted Lebesgue space with a basis $(\phi_k)_{k\in\mathcal{K}}$ with $\sup_{k\in\mathcal{K}}\|\phi_k\|_{\mathrm{L}_\infty(\mathbb{D})} < \infty$. Then every element of the corresponding Wiener space $\mathcal{A}(\mathbb{D},\omega)$ has a continuous representative.*

Proof. See [21, Lemma 2.5]. $\square$

Remark. Lemma 2.1 provides $\mathcal{A}(\mathbb{D},\omega) \subseteq C(\mathbb{D}) := \{f\colon \mathbb{D} \to \mathbb{R} \mid f \text{ continuous}\}$. Furthermore, we get

$$\begin{aligned}\|f\|_{\mathrm{L}_\infty(\mathbb{D})} &= \operatorname*{ess\,sup}_{x\in\mathbb{D}} \left|\sum_{k\in\mathcal{K}} \langle f,\phi_k\rangle_{\mathrm{L}_2(\mathbb{D},\omega)}\, \phi_k(x)\right| \\ &\leq \sup_{k\in\mathcal{K}} \|\phi_k\|_{\mathrm{L}_\infty(\mathbb{D})} \sum_{k\in\mathcal{K}} \left|\langle f,\phi_k\rangle_{\mathrm{L}_2(\mathbb{D},\omega)}\right| \\ &= \sup_{k\in\mathcal{K}} \|\phi_k\|_{\mathrm{L}_\infty(\mathbb{D})}\, \|f\|_{\mathcal{A}(\mathbb{D},\omega)} .\end{aligned}$$

It follows $\mathcal{A}(\mathbb{D},\omega) \subseteq \mathrm{L}_\infty(\mathbb{D})$ and $\mathcal{A}(\mathbb{D},\omega) \subseteq \mathrm{L}_2(\mathbb{D},\omega)$.

Due to this Lemma, we define the evaluation of a function $f \in \mathcal{A}(\mathbb{D},\omega)$ at a point $\mathbf{x} \in \mathbb{D}$ as evaluation of the continuous representative at the point $\mathbf{x}$. Next, we consider partial sums of the function f for finite subsets of the index set $\mathcal{I} \subset \mathcal{K}$, e.g. $S_{\mathcal{I}}(\mathcal{B})f := \sum_{k\in\mathcal{I}} \langle f,\phi_k\rangle_{\mathrm{L}_2(\mathbb{D},\omega)}\, \phi_k$. Furthermore, we define the set of polynomials related to the finite index set $\mathcal{I}$ as

$$\mathcal{T}_{\mathcal{I}}(\mathcal{B}) := \left\{ \sum_{k\in\mathcal{I}} c_k\phi_k \;\middle|\; c_k \in \mathbb{C} \right\}. \tag{2.1}$$

2.2 Orthonormal Basis

In this subsection we firstly introduce the one-dimensional basis functions which we use for the mixed basis. The functions $\phi_k^{\exp} = \exp(2\pi\mathrm{i}k\cdot)$ form the orthonormal Fourier basis of $\mathrm{L}_2(\mathbb{T})$. Additionally, the functions $\phi_k^{\cos} = \sqrt{2}\cos(\pi k\cdot)$ form together with the constant function with value one the orthonormal half period cosine basis of $\mathrm{L}_2([0,1])$. The functions $\phi_k^{\mathrm{alg}} = \sqrt{2}\cos(k\arccos(2\cdot -1))$ form together with the constant function with value one the orthonormal Chebyshev basis of $\mathrm{L}_2([0,1],\omega)$ with the weight $\omega\colon [0,1] \to (0,\infty)$, $\omega(x) := \frac{1}{\pi\sqrt{x-x^2}}$. In the following we are going to work with tensor products of these basis functions.

Definition 2.2. Let $\mathbf{d}$ be a d-dimensional tuple over the set $\{\exp,\cos,\mathrm{alg}\}$. We define the sets

$$\mathbb{D}^{\mathbf{d}} := \bigtimes_{j=1}^{d} \begin{cases} \mathbb{T}, & d_j = \exp \\ [0,1], & d_j \neq \exp \end{cases} \qquad \text{and} \qquad \mathbb{K}^{\mathbf{d}} := \bigtimes_{j=1}^{d} \begin{cases} \mathbb{Z}, & d_j = \exp \\ \mathbb{N}_0, & d_j \neq \exp \end{cases}$$

and the mixed functions

$$\phi_{\mathbf{k}}^{\mathbf{d}}\colon \mathbb{D}^{\mathbf{d}} \to \mathbb{C},\ \phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x}) := \prod_{j=1}^{d} \begin{cases} 1, & k_j = 0 \\ \exp(2\pi\mathrm{i}k_jx_j), & d_j = \exp,\ k_j \neq 0 \\ \sqrt{2}\cos(\pi k_jx_j), & d_j = \cos,\ k_j \neq 0 \\ \sqrt{2}\cos(k_j\arccos(2x_j-1)), & d_j = \mathrm{alg},\ k_j \neq 0 \end{cases}$$

for $\mathbf{k} \in \mathbb{K}^{\mathbf{d}}$. Furthermore, we define the weight function

$$\omega^{\mathbf{d}}\colon \mathbb{D}^{\mathbf{d}} \to (0,\infty),\ \omega^{\mathbf{d}}(\mathbf{x}) := \prod_{j=1}^{d} \begin{cases} 1, & d_j \neq \mathrm{alg} \\ \frac{1}{\pi\sqrt{x_j-x_j^2}}, & d_j = \mathrm{alg} \end{cases} .$$

The mixed functions $\phi_{\mathbf{k}}^{\mathbf{d}}$ form a basis of $\mathrm{L}_2(\mathbb{D}^{\mathbf{d}},\omega^{\mathbf{d}})$ because of their tensor product structure and because their factors are Fourier, half period cosine and Chebyshev basis functions. We name this basis $\mathcal{B}^{\mathbf{d}} := \{\phi_{\mathbf{k}}^{\mathbf{d}} \mid \mathbf{k} \in \mathbb{K}^{\mathbf{d}}\}$.

3 Interpretable ANOVA Approximation

In this section, we define the ANOVA approximation for the mixed basis. We follow the steps in [17]. We start in Subsection 3.1 with defining the ANOVA decomposition, see [5, 12, 11, 6], in the way like [16, 17] through a series expansion, by using the mixed basis. Furthermore, we define analytic global sensitivity indices [24]. In Subsection 3.2 we describe the procedure of ANOVA approximation [16, 17, 18], and we deduce a way to compute it numerically.

3.1 ANOVA Decomposition

Let f be an $\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})$ function. Since $\phi_{\mathbf{k}}^{\mathbf{d}}$ with $\mathbf{k} \in \mathbb{K}^{\mathbf{d}}$ form an orthonormal basis of $\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})$, f can be written as

$$f = \sum_{\mathbf{k} \in \mathbb{K}^{\mathbf{d}}} c_{\mathbf{k}}^{\mathbf{d}}(f) \phi_{\mathbf{k}}^{\mathbf{d}}$$

with coefficients $c_{\mathbf{k}}^{\mathbf{d}}(f) \coloneqq \langle f, \phi_{\mathbf{k}}^{\mathbf{d}} \rangle_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})}$. Furthermore, we get the Parseval equality

$$\|f\|^2_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})} = \sum_{\mathbf{k} \in \mathbb{K}^{\mathbf{d}}} |c_{\mathbf{k}}^{\mathbf{d}}(f)|^2,$$

from the fact that $\mathcal{B}^{\mathbf{d}}$ is a basis of $\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})$. Next, we decompose the function f into ANOVA terms. We denote subsets of coordinate indices with small boldface letters $\mathbf{u} \in \mathcal{P}([d])$. For every subset of indices $\mathbf{u}$ we define the ANOVA term

$$f_{\mathbf{u}}(\mathbf{x}) \coloneqq \sum_{\substack{\mathbf{k} \in \mathbb{K}^{\mathbf{d}} \\ \operatorname{supp} \mathbf{k} = \mathbf{u}}} c_{\mathbf{k}}^{\mathbf{d}}(f) \phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x}).$$

Note that such an ANOVA term $f_{\mathbf{u}}(\mathbf{x})$ is independent of x_j if $j \notin \mathbf{u}$. These ANOVA terms $f_{\mathbf{u}}$, $\mathbf{u} \subseteq [d]$ decompose the function f uniquely into

$$f = \sum_{\mathbf{k} \in \mathbb{K}^{\mathbf{d}}} c_{\mathbf{k}}^{\mathbf{d}}(f) \phi_{\mathbf{k}}^{\mathbf{d}} = \sum_{\mathbf{u} \in \mathcal{P}([d])} f_{\mathbf{u}}.$$

This follows since $\{\{\mathbf{k} \in \mathbb{K}^{\mathbf{d}} \mid \operatorname{supp} \mathbf{k} = \mathbf{u}\} \mid \mathbf{u} \in \mathcal{P}([d])\}$ is a partition of the set $\mathbb{K}^{\mathbf{d}}$. Additionally, we define the variance of a function f as $\sigma^2(f) = \int_{\mathbb{D}^{\mathbf{d}}} |f(\mathbf{x}) - c_0^{\mathbf{d}}(f)|^2 \omega^{\mathbf{d}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}$, which is equivalent to $\sigma^2(f) = \|f\|^2_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})} - |c_{\mathbf{0}}^{\mathbf{d}}(f)|^2$. The Parseval equality states

$$\sigma^2(f) = \sum_{\mathbf{k} \in \mathbb{K}^{\mathbf{d}} \setminus \{\mathbf{0}\}} |c_{\mathbf{k}}^{\mathbf{d}}(f)|^2.$$

Furthermore, we get the variance of ANOVA terms $f_{\mathbf{u}}$ through

$$\sigma^2(f_{\mathbf{u}}) = \sum_{\mathbf{k} \in \mathbb{K}^{\mathbf{d}} \setminus \{\mathbf{0}\}} |c_{\mathbf{k}}^{\mathbf{d}}(f_{\mathbf{u}})|^2 = \sum_{\substack{\mathbf{k} \in \mathbb{K}^{\mathbf{d}} \\ \operatorname{supp} \mathbf{k} = \mathbf{u}}} |c_{\mathbf{k}}^{\mathbf{d}}(f_{\mathbf{u}})|^2 = \|f_{\mathbf{u}}\|^2_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})}.$$

Finally, we define the analytic global sensitivity indices (GSI) like [24] through

$$\rho(\mathbf{u}, f) \coloneqq \frac{\sigma^2(f_{\mathbf{u}})}{\sigma^2(f)}.$$

We point out that the analytic GSIs depend on the weight $\omega^{\mathbf{d}}$. We use the analytic GSU as a tool to measure how important certain ANOVA terms $f_{\mathbf{u}}$ are for the reconstruction of the function f. We use this information for the construction of good suitable index sets. Next we truncate the ANOVA decomposition. We use a set of subsets of indices $U \subseteq \mathcal{P}([d])$ for this truncation. We define

$$\mathrm{T}_U f(\mathbf{x}) = \sum_{\mathbf{u} \in U} f_{\mathbf{u}}(\mathbf{x}).$$

To find this set U, we choose the ANOVA terms $f_{\mathbf{u}}$ with the highest GSI's to get $\mathrm{T}_U f \approx f$.

3.2 Numerical ANOVA Approximation

In this section our aim is to approximate a function $f \in \mathcal{A}(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})$. We are given a set of $M \in \mathbb{N}$ nodes $\mathcal{X} \subset \mathbb{D}^{\mathbf{d}}$, $|\mathcal{X}| = M$ and the corresponding function values $(f(\mathbf{x}))_{\mathbf{x}\in\mathcal{X}} \coloneqq \mathbf{f} \in \mathbb{C}^M$. We aim to find a mixed polynomial

$$f^{\mathbf{d}} \colon \mathbb{D}^{\mathbf{d}} \to \mathbb{C},\ f^{\mathbf{d}}(\mathbf{x}) \coloneqq \sum_{\mathbf{k}\in\mathcal{I}} \hat{f}^{\mathbf{d}}_{\mathbf{k}} \phi^{\mathbf{d}}_{\mathbf{k}}(\mathbf{x}), \quad \hat{f}^{\mathbf{d}}_{\mathbf{k}} \in \mathbb{C}$$

for which $f \approx f^{\mathbf{d}}$ holds, i.e. $\|f - f^{\mathbf{d}}\|_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}},\omega^{\mathbf{d}})}$ is small and $\mathcal{I}$ is a finite subset of $\mathbb{K}^{\mathbf{d}}$. In the next Subsection 3.2.1 we consider some ways to choose $\mathcal{I}$. The next Subsection 3.2.2 presents a way to find the mixed polynomial $f^{\mathbf{d}}$ which minimizes the $\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})$ norm of $f - f^{\mathbf{d}}$ for a given index set $\mathcal{I}$ and nodes $\mathcal{X}$. To this end we show in Subsection 3.2.3 how to choose and refine the truncation set U.

3.2.1 Grouped Index Sets

We present some index sets that are important for this paper. Since these index sets contain frequencies for the mixed basis we call them frequency sets. We begin with frequency sets which are full d-dimensional hypercubes, but since their size grows exponential in the dimension d we introduce better controllable frequency sets. We start with the frequency sets which are full d-dimensional hypercubes, i.e.

$$\mathcal{I}^{\mathbf{d}}_{\mathbf{N}} \coloneqq \bigtimes_{j=1}^{d} \begin{cases} \mathbb{Z} \cap \left[-\frac{N_j}{2}, \frac{N_j}{2}\right), & d_j = \exp \\ \mathbb{N}_0 \cap \left[0, \frac{N_j}{2}\right), & d_j \neq \exp \end{cases} \tag{3.1}$$

for a vector of bandwidths $\mathbf{N} = (N_j)_{j=1}^d \in (2\mathbb{N}_0)^d$. These sets have the cardinality

$$|\mathcal{I}^{\mathbf{d}}_{\mathbf{N}}| = \prod_{j=1}^{d} \begin{cases} N_j, & d_j = \exp \\ \frac{N_j}{2}, & d_j \neq \exp \end{cases}. \tag{3.2}$$

Next we define thinner frequency sets with less cardinality. In detail these frequency sets should have a high bandwidth along the coordinate axes, less bandwidth along the coordinate planes and so on. For this purpose we define the following frequency sets $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}}$ with bandwidths $\mathbf{N} = (N_j)_{j=1}^d \in (2\mathbb{N}_0)^d$. Here if N_j is zero the set $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}}$ should only contain the frequency zero, when it is projected onto the dimension j. If N_j is not zero the projection onto the dimension j should not contain the frequency zero, because it is contained in a lower dimensional set. This is being done by the frequency set

$$\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}} \coloneqq \bigtimes_{j=1}^{d} \begin{cases} \{0\}, & N_j = 0 \\ \mathbb{Z} \cap \left[-\frac{N_j}{2}, \frac{N_j}{2}\right) \setminus \{0\}, & d_j = \exp \text{ and } N_j \neq 0 \\ \mathbb{N}_0 \cap \left[0, \frac{N_j}{2}\right) \setminus \{0\}, & d_j \neq \exp \text{ and } N_j \neq 0 \end{cases}. \tag{3.3}$$

These frequency sets $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}}$ have the cardinality

$$|\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}}| = \prod_{j=1}^{d} \begin{cases} 1, & N_j = 0 \\ N_j - 1, & d_j = \exp \text{ and } N_j \neq 0 \\ \frac{N_j}{2} - 1, & d_j \neq \exp \text{ and } N_j \neq 0 \end{cases}. \tag{3.4}$$

Since these frequency sets are disjoint, if the bandwidths have different support, we can form the union of them to derive a new frequency set. We choose for every $\mathbf{u} \in U$ a bandwidth

$$\mathbf{N}^{\mathbf{u}} = (N^{\mathbf{u}}_j)_{j=1}^d \in (2\mathbb{N})^{\mathbf{u}} \coloneqq \bigtimes_{j=1}^{d} \begin{cases} (2\mathbb{N}), & j \in \mathbf{u} \\ \{0\}, & \text{else} \end{cases} \tag{3.5}$$

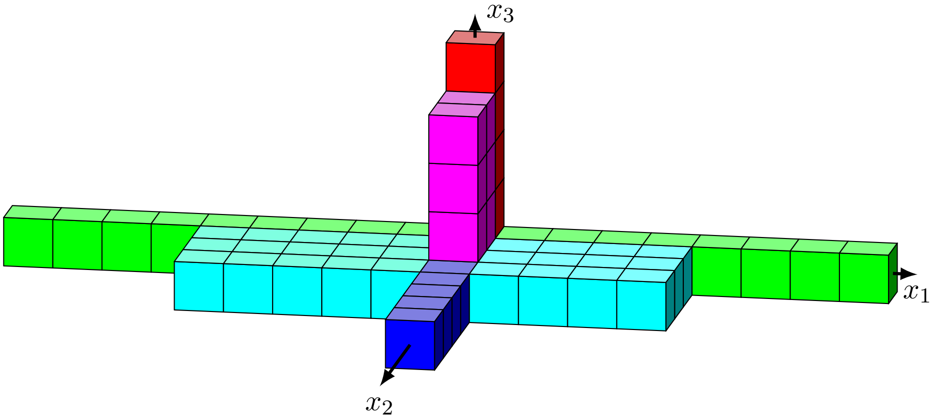

Figure 1: Frequency set $\mathcal{I}(U)$ for $\mathbf{d} = \begin{pmatrix} \text{exp} \\ \text{alg} \\ \text{cos} \end{pmatrix}$ and $U = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}\}$ with $\mathbf{N}^{\{1\}} = \begin{pmatrix} 18 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{N}^{\{2\}} = \begin{pmatrix} 0 \\ 16 \\ 0 \end{pmatrix}$, $\mathbf{N}^{\{3\}} = \begin{pmatrix} 0 \\ 0 \\ 10 \end{pmatrix}$, $\mathbf{N}^{\{1,2\}} = \begin{pmatrix} 10 \\ 8 \\ 0 \end{pmatrix}$, and $\mathbf{N}^{\{2,3\}} = \begin{pmatrix} 0 \\ 6 \\ 8 \end{pmatrix}$. The frequency set $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{1\}}}$ is shown in green, $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{2\}}}$ in blue, $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{3\}}}$ in red, $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{1,2\}}}$ in cyan, and $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{2,3\}}}$ in magenta.

and define the frequency set

$$\mathcal{I}(U) \coloneqq \bigcup_{\mathbf{u} \in U} \tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\mathbf{u}}}. \tag{3.6}$$

This frequency set $\mathcal{I}(U)$ hat the advantage that, if the truncation set U only contains subsets $\mathbf{u}$ of the size up to $|\mathbf{u}| \leq d_s \in \mathbb{N}$ it will only grow polynomial with the power d_s which is an improvement to the exponential growing full d-dimensional hypercubes $\mathcal{I}^{\mathbf{d}}_{\mathbf{N}}$.

Example. We show in Figure 1 a typical example for such a set $\mathcal{I}(U)$, where $\mathbf{d} = \begin{pmatrix} \text{exp} \\ \text{alg} \\ \text{cos} \end{pmatrix}$ and $U = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}\}$. Here we use the bandwidths $\mathbf{N}^{\{1\}} = \begin{pmatrix} 18 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{N}^{\{2\}} = \begin{pmatrix} 0 \\ 16 \\ 0 \end{pmatrix}$, $\mathbf{N}^{\{3\}} = \begin{pmatrix} 0 \\ 0 \\ 10 \end{pmatrix}$, $\mathbf{N}^{\{1,2\}} = \begin{pmatrix} 10 \\ 8 \\ 0 \end{pmatrix}$, and $\mathbf{N}^{\{2,3\}} = \begin{pmatrix} 0 \\ 6 \\ 8 \end{pmatrix}$. The parts $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\mathbf{u}}}$ of this frequency set $\mathcal{I}(U)$ are shown in different colors. The set $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{1\}}}$ is shown in green, $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{2\}}}$ in blue, $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{3\}}}$ in red, $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{1,2\}}}$ in cyan, and $\tilde{\mathcal{I}}^{\mathbf{d}}_{\mathbf{N}^{\{2,3\}}}$ in magenta.

3.2.2 Approximation

In this section we assume that the set U and the bandwidths $\mathbf{N}^{\mathbf{u}}$ are known. For a way to choose them we refer to Subsection 3.2.3. Now, we approximate the truncated function $\mathrm{T}_U f$ with a mixed polynomial $f^{\mathbf{d}} \in \mathcal{T}_{\mathcal{I}(U)}(\mathcal{B}^{\mathbf{d}})$, where the set of polynomials $\mathcal{T}_{\mathcal{I}(U)}(\mathcal{B}^{\mathbf{d}})$ is defined in (2.1). As result, we get $f \approx \mathrm{T}_U f \approx f^{\mathbf{d}}$.

The mixed polynomial is completely determined by finitely many mixed coefficients $(c^{\mathbf{d}}_{\mathbf{k}}(f^{\mathbf{d}}))_{\mathbf{k} \in \mathcal{I}(U)} \in \mathbb{C}^{|\mathcal{I}(U)|}$. Now it is our goal to find an approximation $\hat{\mathbf{f}}^{\mathbf{d}} \approx (c^{\mathbf{d}}_{\mathbf{k}}(f^{\mathbf{d}}))_{\mathbf{k} \in \mathcal{I}(U)}$ to this mixed coefficients. To achieve this, we compute the least squares solution

$$\|f - f^{\mathbf{d}}\|^2_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})} = \int_{\mathbb{D}^{\mathbf{d}}} |f(\mathbf{x}) - f^{\mathbf{d}}(\mathbf{x})|^2 \omega^{\mathbf{d}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}.$$

We approximate this integral by evaluating the function $|f(\mathbf{x}) - f^{\mathbf{d}}(\mathbf{x})|^2$ at the $M = |\mathcal{X}|$ given nodes $\mathcal{X}$, where we know the values $\mathbf{f} = (f(\mathbf{x}))_{\mathbf{x} \in \mathcal{X}}$. For this approximation we assume that the nodes in $\mathcal{X}$ are distributed in $\mathbb{D}^{\mathbf{d}}$ with the density $\omega^{\mathbf{d}}$. We get

$$\|f - f^{\mathbf{d}}\|^2_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}}, \omega^{\mathbf{d}})} \approx \frac{1}{M} \sum_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x}) - f^{\mathbf{d}}(\mathbf{x})|^2$$

$$= \frac{1}{M} \|\mathbf{f} - (f^{\mathbf{d}}(\mathbf{x}))_{\mathbf{x}\in\mathcal{X}}\|_2^2$$
$$= \frac{1}{M} \|\mathbf{f} - \mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))\hat{\mathbf{f}}^{\mathbf{d}}\|_2^2,$$

where $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))$ is the matrix $(\phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x}))_{\mathbf{x}\in\mathcal{X},\,\mathbf{k}\in\mathcal{I}(U)} \in \mathbb{C}^{M\times|\mathcal{I}(U)|}$. Now we choose $\hat{\mathbf{f}}^{\mathbf{d}}$ such that the distance between the function f and the approximation $f^{\mathbf{d}}$ is as small as possible, i.e.

$$\hat{\mathbf{f}}^{\mathbf{d}} := \underset{\hat{\mathbf{h}}^{\mathbf{d}}\in\mathbb{C}^{|\mathcal{I}(U)|}}{\arg\min} \|\mathbf{f} - \mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))\hat{\mathbf{h}}^{\mathbf{d}}\|_2^2. \tag{3.7}$$

We solve the minimization problem (3.7) with the LSQR algorithm [13]. This algorithm multiplies in every iteration step with the matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))$ and its adjoint $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))^*$. In Section 4 we develop fast algorithms to multiply with this kind of matrices $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))$ and $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))^*$. We point out that the number of iterations of the LSQR algorithm depends on the condition number of the underlying matrix. This condition number is in many applications much better than the worst case estimation, see e.g. [4, 3].
We obtain the approximation

$$f^{\mathbf{d}} \colon \mathbb{D}^{\mathbf{d}} \to \mathbb{C},\ f^{\mathbf{d}}(\mathbf{x}) := \sum_{\mathbf{k}\in\mathcal{I}(U)} \hat{f}_{\mathbf{k}}^{\mathbf{d}} \phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x})$$

for the function f. Let $\mathcal{X}_{\text{test}} \subset \mathbb{D}^{\mathbf{d}}$, $|\mathcal{X}_{\text{test}}| = M_{\text{test}}$ be a set with $M_{\text{test}} \in \mathbb{N}$ nodes, where we evaluate the approximation $f^{\mathbf{d}}$. Then $(f^{\mathbf{d}}(\mathbf{x}))_{\mathbf{x}\in\mathcal{X}_{\text{test}}} = \mathbf{\Phi}(\mathcal{X}_{\text{test}}, \mathcal{I}(U))\hat{\mathbf{f}}^{\mathbf{d}}$ holds.

3.2.3 How to choose the truncation set U?

We choose U in two steps, firstly we choose a large superposition set U_{d_s}. Then, we calculate the mixed coefficients $\hat{\mathbf{f}}^{\mathbf{d}}$ according to this set U_{d_s} for the approximating mixed polynomial $f^{\mathbf{d}}$. In this way we get approximated global sensitivity indices $\rho(\mathbf{u}, f^{\mathbf{d}})$ for every $\mathbf{u} \in U_{d_s}$. Using these approximated GSIs we can refine the set U_{d_s} to the final set U.
We choose U_{d_s} as set of all subsets of $[d]$ with cardinality smaller or equal to $d_s \in [d]$, i.e.

$$U_{d_s} := \{\mathbf{u} \subseteq [d] \mid |\mathbf{u}| \leq d_s\}. \tag{3.8}$$

Then we choose appropriate bandwidths $\mathbf{N}^{\mathbf{u}} \in (2\mathbf{N})^{\mathbf{u}}$. For the notion we refer to (3.5). It is good to choose the bandwidths in a way that we have an oversampling, i.e. $|\mathcal{I}(U_{d_s})| < M$ with $|\mathcal{I}(U_{d_s})| = \sum_{\mathbf{u}\in U_{d_s}} |\tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}}|$. For $|\tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}}|$ we refer to (3.4).
As the next step we calculate the mixed coefficients $\hat{\mathbf{f}}^{\mathbf{d}}$ for the approximating mixed polynomial $f^{\mathbf{d}}$ in the way we described it in the previous part. Using these mixed coefficients $\hat{\mathbf{f}}^{\mathbf{d}} = (\hat{f}_{\mathbf{k}}^{\mathbf{d}})_{\mathbf{k}\in\mathcal{I}(U_{d_s})}$ we calculate the approximated global sensitivity indices $\rho(\mathbf{u}, f^{\mathbf{d}})$ for the mixed polynomial $f^{\mathbf{d}}$ for all $\mathbf{u} \in U_{d_s}$ through

$$\rho(\mathbf{u}, f^{\mathbf{d}}) = \frac{\sigma^2(f_{\mathbf{u}}^{\mathbf{d}})}{\sigma^2(f^{\mathbf{d}})} = \frac{\sum\limits_{\substack{\mathbf{k}\in\mathbb{K}^{\mathbf{d}} \\ \operatorname{supp}\mathbf{k}=\mathbf{u}}} |\hat{f}_{\mathbf{k}}^{\mathbf{d}}|^2}{\sum\limits_{\mathbf{k}\in\mathbb{K}^{\mathbf{d}}\setminus\{\mathbf{0}\}} |\hat{f}_{\mathbf{k}}^{\mathbf{d}}|^2} = \frac{\sum\limits_{\mathbf{k}\in\tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}}} |\hat{f}_{\mathbf{k}}^{\mathbf{d}}|^2}{\sum\limits_{\mathbf{k}\in\mathcal{I}(U_{d_s})\setminus\{\mathbf{0}\}} |\hat{f}_{\mathbf{k}}^{\mathbf{d}}|^2}.$$

Since the mixed polynomial $f^{\mathbf{d}}$ approximates the function f, the approximated GSI $\rho(\mathbf{u}, f^{\mathbf{d}})$ should approximate the analytic GSI $\rho(\mathbf{u}, f)$.
To this end we choose a threshold $\theta > 0$ and define the set

$$U_\theta := \{\mathbf{u} \in U_{d_s} \mid \rho(\mathbf{u}, f^{\mathbf{d}}) > \theta\}. \tag{3.9}$$

With this set U_θ we do the approximating procedure again.

4 Fast Evaluation of Mixed Polynomials

In this section, we develop a fast algorithm for evaluating sums of the mixed basis functions $\phi_{\mathbf{k}}^{\mathbf{d}}$, i.e.

$$f^{\mathbf{d}} := \sum_{\mathbf{k}\in\mathcal{I}} \hat{f}_{\mathbf{k}}^{\mathbf{d}} \phi_{\mathbf{k}}^{\mathbf{d}} \tag{4.1}$$

with arbitrary coefficients $\hat{f}_{\mathbf{k}}^{\mathbf{d}} \in \mathbb{C}$ on a finite index sets $\mathcal{I} \subseteq \mathbb{K}^{\mathbf{d}}$ at $M \in \mathbb{N}$ nodes $\mathcal{X} \subset \mathbb{D}^{\mathbf{d}}$, $|\mathcal{X}| = M$ simultaneously. This evaluation is equivalent to the matrix-vector multiplication of the non-equidistant mixed matrix

$$\mathbf{\Phi}(\mathcal{X},\mathcal{I}) := (\phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x}))_{\mathbf{x}\in\mathcal{X},\mathbf{k}\in\mathcal{I}} \tag{4.2}$$

with the vector $\hat{\mathbf{f}}^{\mathbf{d}} = (\hat{f}_{\mathbf{k}}^{\mathbf{d}})_{\mathbf{k}\in\mathcal{I}}$, where $\mathcal{X}$ is the set of the nodes at which we are interested to evaluate the mixed polynomial $f^{\mathbf{d}}$. In the Subsection 4.1 we consider the index set $\mathcal{I} = \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$, defined in (3.1). Since we will use the thinner index sets $\mathcal{I} = \mathcal{I}(U)$, defined in (3.6), we will introduce in Subsection 4.2 an algorithm for this case which will rely on the algorithm in Subsection 4.1.

4.1 Non-Equidistant Fast Mixed Transform

In this subsection, we present a method with a computational cost of only $\mathcal{O}(|\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}|\log|\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}|+|\log\epsilon|^d M)$ for the evaluation of mixed polynomials $f^{\mathbf{d}}$ with the frequency set $\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$. This is faster than the straightforward matrix vector multiplication with the matrix $\mathbf{\Phi}(\mathcal{X},\mathcal{I}_{\mathbf{N}}^{\mathbf{d}})$ which takes $\mathcal{O}(|\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}|M)$ arithmetical operations.

We point out that the mixed polynomial $f^{\mathbf{d}}$ with $\mathbf{d} = (\exp,\dots,\exp) =: \mathbf{exp}$ is a trigonometric polynomial, which can be evaluated through the non-equidistant fast Fourier transform (NFFT) [14, Chapter 7] with a computational cost of $\mathcal{O}(|\mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}|\log|\mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}| + |\log\epsilon|^d M)$, where ϵ is the required precision, $|\mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}|$ is the cardinality of the frequency set given in (3.2), and M is the number of nodes where we evaluate the mixed polynomial $f^{\mathbf{exp}}$. Now, we make use of the NFFT in order to evaluate arbitrary mixed polynomials $f^{\mathbf{d}}$. The identities (1.1) and (1.2) provides us the possibility to transform polynomials $f^{\cos}$ and f^{alg} into trigonometric polynomials $f^{\exp}$. It follows, that one dimensional polynomials of the form $f^{\cos}$ and f^{alg} can be evaluated through an NFFT. Since our mixed basis functions $\phi_{\mathbf{k}}^{\mathbf{d}}$ have a tensor product structure, we use these identities (1.1) and (1.2) in every dimension where the half period cosine basis or the Chebyshev basis is used.

Theorem 4.1. *Let* $\hat{\mathbf{f}}^{\mathbf{d}} = (\hat{f}_{\mathbf{k}}^{\mathbf{d}})_{\mathbf{k}\in\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}} \in \mathbb{C}^{|\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}|}$ *be a coefficient vector for a mixed polynomial* $f^{\mathbf{d}}$ *defined in* (4.1) *and an arbitrary* $\mathbf{d} \in \{\exp,\cos,\mathrm{alg}\}^d$ *and* $d \in \mathbb{N}$. *We define the coefficient vector* $\hat{\mathbf{f}}^{\mathbf{exp}} = (\hat{f}_{\mathbf{k}}^{\mathbf{exp}})_{\mathbf{k}\in\mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}} \in \mathbb{C}^{|\mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}|}$ *through*

$$\hat{f}_{\mathbf{k}}^{\mathbf{exp}} := \hat{f}_{\mathbf{s}(\mathbf{k})}^{\mathbf{d}} \prod_{j=1}^{d} \begin{cases} 1, & d_j = \exp \ \textit{or}\ k_j = 0 \\ 0, & d_j \neq \exp \ \textit{and}\ k_j = -\frac{N_j}{2} \\ \frac{\sqrt{2}}{2}, & \textit{else} \end{cases} \tag{4.3}$$

for all $\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}$, *where* $\mathbf{s}$ *is the index transformation*

$$\mathbf{s}\colon \mathcal{I}_{\mathbf{N}}^{\mathbf{exp}} \to \mathcal{I}_{\mathbf{N}}^{\mathbf{d}},\ \mathbf{s}(\mathbf{k}) := \left(\begin{cases} k_j, & d_j = \exp \\ |k_j|, & d_j \neq \exp \ \textit{and}\ k_j \neq -\frac{N_j}{2} \\ 0, & d_j \neq \exp \ \textit{and}\ k_j = -\frac{N_j}{2} \end{cases} \right)_{j=1}^{d}. \tag{4.4}$$

Furthermore, we define the function

$$\mathbf{t}\colon \mathbb{D}^{\mathbf{d}} \to \mathbb{D}^{\mathbf{exp}},\ \mathbf{t}(\mathbf{x}) \coloneqq \left(\begin{cases} x_j, & d_j = \exp \\ \frac{x_j}{2}, & d_j = \cos \\ \frac{\arccos(2x_j-1)}{2\pi}, & d_j = \mathrm{alg} \end{cases} \right)_{j=1}^{d}. \tag{4.5}$$

Then the identity $f^{\mathbf{d}} = f^{\mathbf{exp}} \circ \mathbf{t}$ holds.

Proof. The assumptions follow from the identities (1.1) and (1.2) together with the fact, that $\phi_{\mathbf{k}}^{\mathbf{d}}$ have a tensor like structure. $\square$

Remark. The Theorem 4.1 provides us a decomposition of the non-equidistant mixed matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}}^{\mathbf{d}})$. For this purpose we define the diagonal matrix

$$\mathbf{D} \coloneqq \operatorname{diag} \left(\prod_{j=1}^{d} \begin{cases} 1, & d_j = \exp \text{ or } k_j = 0 \\ 0, & d_j \neq \exp \text{ and } k_j = -\frac{N_j}{2} \\ \frac{\sqrt{2}}{2}, & \text{else} \end{cases} \right)_{\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}},$$

the canonical map $\mathbf{\Pi} \coloneqq (\delta_{\mathbf{k},\mathbf{l}})_{\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}, \mathbf{l} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}}$, the projection $\mathbf{P} \coloneqq (\delta_{\mathbf{k},\mathbf{s}(\mathbf{k})})$, and the non-equidistant Fourier matrix $\mathbf{A} = (\exp(2\pi\mathrm{i}\langle \mathbf{k}, \mathbf{t}(\mathbf{x})\rangle))_{\mathbf{x} \in \mathcal{X}, \mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}}$. Then the matrix transformation

$$\mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}) = \mathbf{A}\mathbf{\Pi}^{\top}\mathbf{P}^{\top}\mathbf{D} \tag{4.6}$$

follows directly.

We summarize the procedure for the efficient evaluation of the mixed polynomials $f^{\mathbf{d}}$ at M arbitrary nodes as non-equidistant fast mixed transform (NFMT) in Algorithm 1.

Input: Vector $\mathbf{d} \in \{\exp, \cos, \mathrm{alg}\}^d$, bandwidths $\mathbf{N} \in (2\mathbb{N})^d$, coefficients $\hat{f}_{\mathbf{k}}^{\mathbf{d}} \in \mathbb{C}$ for $\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$, nodes $\mathcal{X} \subset \mathbb{D}^{\mathbf{d}}$, $|\mathcal{X}| = M$

1 Define the coefficients $\hat{f}_{\mathbf{k}}^{\mathbf{exp}}$ given in (4.3) for all $\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{exp}}$.

2 Compute

$$s(\mathbf{x}) = \sum_{\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}} \hat{f}_{\mathbf{k}}^{\mathbf{exp}} \exp\left(2\pi\mathrm{i}\left\langle \mathbf{k}, \tilde{\mathbf{x}} \right\rangle\right)$$

at the nodes $\tilde{\mathbf{x}} \in \{\mathbf{t}(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}$ with $\mathbf{t}\colon \mathbb{D}^{\mathbf{d}} \to \mathbb{D}^{\mathbf{exp}}$ given in (4.5) using a d-variate NFFT

Output: $s(\mathbf{x}) = f^{\mathbf{d}}(\mathbf{x})$ for $\mathbf{x} \in \mathcal{X}$

Computational cost: $\mathcal{O}(|\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}| \log |\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}| + |\log \epsilon|^d M)$

Algorithm 1: NFMT for the fast evaluation of mixed polynomials $f^{\mathbf{d}}$ for frequency sets $\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$ defined in (3.1).

In addition we evaluate the sum

$$h(\mathbf{k}) = \sum_{\mathbf{x} \in \mathcal{X}} g_{\mathbf{x}} \phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x}),\ g_{\mathbf{x}} \in \mathbb{C} \tag{4.7}$$

for all $\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$. This is equivalent to the matrix vector product of the transposed non-equidistant mixed matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}}^{\mathbf{d}})^{\top}$ with the vector $\mathbf{g} = (g_{\mathbf{x}})_{\mathbf{x} \in \mathcal{X}}$. We use the factorization (4.6) of the non-equidistant mixed matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}}^{\mathbf{d}})$ and get directly the Algorithm 2, which provides a method for the fast evaluation of the sum (4.7).

Input: Vector $\mathbf{d} \in \{\exp, \cos, \mathrm{alg}\}^d$, bandwidths $\mathbf{N} \in (2\mathbb{N})^d$, nodes $\mathcal{X} \subset \mathbb{D}^{\mathbf{d}}$, $|\mathcal{X}| = M$, coefficients $h_{\mathbf{x}} \in \mathbb{C}$ for $\mathbf{x} \in \mathcal{X}$

1 Compute

$$\tilde{h}^{\exp}(\mathbf{k}) = \sum_{\mathbf{x} \in \mathcal{X}} h_{\mathbf{x}} \exp\left(2\pi \mathrm{i} \left\langle \mathbf{k}, \tilde{\mathbf{x}} \right\rangle\right)$$

with $\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$ at the nodes $\tilde{\mathbf{x}} \in \{\mathbf{t}(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}$ with $\mathbf{t} \colon \mathbb{D}^{\mathbf{d}} \to \mathbb{D}^{\mathbf{exp}}$ defined in (4.5) using a d-variate $\mathsf{NFFT}^\top$

2 Compute

$$\tilde{h}(\mathbf{k}) = \sum_{\substack{\mathbf{l} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{exp}} \\ \mathbf{s}(\mathbf{l}) = \mathbf{k}}} \tilde{h}^{\exp}(\mathbf{l}) \prod_{j=1}^{d} \begin{cases} 1, & d_j = \exp \text{ or } k_j = 0 \\ 0, & d_j \neq \exp \text{ and } k_j = -\frac{N_j}{2} \\ \frac{\sqrt{2}}{2}, & \text{else} \end{cases}$$

with $\mathbf{s}$ defined in (4.4).

Output: $\tilde{h}(\mathbf{k}) = h(\mathbf{k})$, see (4.7), for $\mathbf{k} \in \mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$

Computational cost: $\mathcal{O}(|\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}| \log |\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}| + |\log \epsilon|^d M)$

Algorithm 2: $\mathsf{NFMT}^\top$ for the fast evaluation of sums of the form (4.7) for frequency sets $\mathcal{I}_{\mathbf{N}}^{\mathbf{d}}$ defined in (3.1).

Remark.

- We note that one can extend the non-equidistant fast mixed transformations to other orthogonal polynomials using [15]. The algorithm known as discrete polynomial transform provides a fast basis exchange for arbitrary orthogonal polynomials with satisfying a three-term recurrence into the Chebychev basis. These Chebychev polynomials can then be evaluated by the Algorithms 1 and 2.
- One can also use other transformations, such as the transformation of the unit interval [0,1] into the real numbers $\mathbb{R}$ from [19]. This allows us to handle normally distributed nodes.

It should be noted that these transformations can be performed in each dimension separately due to the tensor product structure of the basis and the flexibility of the mixed basis.

4.2 Grouped Mixed Transformations

In this section we derive a fast algorithm for the evaluation of mixed polynomials $f^{\mathbf{d}} = \sum_{\mathbf{k} \in \mathcal{I}(U)} \hat{f}_{\mathbf{k}}^{\mathbf{d}} \phi_{\mathbf{k}}^{\mathbf{d}}$ where $\mathcal{I}(U)$ is a frequency set defined in (3.6). We denote $\mathbf{x}_{\mathbf{u}} \coloneqq \mathbf{\Pi}_{\mathbf{u}} \mathbf{x}$, where $\mathbf{\Pi}_{\mathbf{u}}$ is the canonical map $\mathbf{\Pi}_{\mathbf{u}}$ onto the dimensions contained in $\mathbf{u}$.
The evaluation of this sum at the nodes $\mathbf{x} \in \mathcal{X}$ is equivalent to calculate the matrix vector product $\mathbf{f}^{\mathbf{d}} = \mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U)) \hat{\mathbf{f}}^{\mathbf{d}} \in \mathbb{C}^M$ with the matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))$, defined in (4.2) and the vector $\hat{\mathbf{f}}^{\mathbf{d}} \in \mathbb{C}^{|\mathcal{I}(U)|}$. We follow the steps in [2] and get

$$\sum_{\mathbf{k} \in \mathcal{I}(U)} \hat{f}_{\mathbf{k}}^{\mathbf{d}} \phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x}) = \sum_{\mathbf{u} \in U} \sum_{\mathbf{k} \in \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}}} \hat{f}_{\mathbf{k}}^{\mathbf{d}} \phi_{\mathbf{k}}^{\mathbf{d}}(\mathbf{x})$$

through the structure of the frequency set $\mathcal{I}(U)$. In other words, the matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))$ is a block matrix with horizontally arranged blocks $\mathbf{\Phi}(\mathcal{X}, \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}})$, $\mathbf{u} \in U$, i.e. $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U)) = (\mathbf{\Phi}(\mathcal{X}, \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}})^\top)_{\mathbf{u} \in U}^\top$. Thus, we divide the task. For every $\mathbf{u} \in U$ we multiply the vector $\hat{\mathbf{f}}^{\mathbf{d},\mathbf{u}} \coloneqq (\hat{f}_{\mathbf{k}}^{\mathbf{d}})_{\mathbf{k} \in \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}}}$ with the block $\mathbf{\Phi}(\mathcal{X}, \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}})$. We get for these blocks

$$\mathbf{\Phi}(\mathcal{X}, \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}}) = (\phi_{\mathbf{k}}^{\mathbf{d}_{\mathbf{u}}}(\mathbf{x}_{\mathbf{u}}))_{\mathbf{x} \in \mathcal{X},\, \mathbf{k} \in \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}_{\mathbf{u}}}} = \mathbf{\Phi}(\{\mathbf{x}_{\mathbf{u}} \mid \mathbf{x} \in \mathcal{X}\}, \tilde{\mathcal{I}}_{\mathbf{N}^{\mathbf{u}}}^{\mathbf{d}_{\mathbf{u}}}).$$

We define the vector $\hat{\mathbf{g}}^{\mathbf{d}_\mathbf{u}} = (\hat{g}_\mathbf{k}^{\mathbf{d}_\mathbf{u}})_{\mathbf{k}\in\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}} \in \mathbb{C}^{|\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}|}$, where now $\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}$ is a frequency set which we can use for an NFMT. We set each component of $\hat{\mathbf{g}}^{\mathbf{d}_\mathbf{u}}$ which is not contained in the set $\tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}$ to zero, e.g.

$$\hat{g}_\mathbf{k} = \begin{cases} \hat{f}_\mathbf{k}, & |\operatorname{supp}\mathbf{k}| = |\mathbf{u}| \\ 0, & \text{else} \end{cases}, \mathbf{k} \in \mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}.$$

Then we obtain

$$\sum_{\mathbf{k}\in\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}} \hat{g}_\mathbf{k}^{\mathbf{d}_\mathbf{u}} \phi_\mathbf{k}^{\mathbf{d}_\mathbf{u}}(\mathbf{x}) = \sum_{\mathbf{k}\in\tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}} \hat{f}_\mathbf{k}^\mathbf{d} \phi_\mathbf{k}^{\mathbf{d}_\mathbf{u}}(\mathbf{x}) + \sum_{\mathbf{k}\in\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}\setminus\tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}} 0\cdot\phi_\mathbf{k}^{\mathbf{d}_\mathbf{u}}(\mathbf{x}) = \sum_{\mathbf{k}\in\tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}} \hat{f}_\mathbf{k}^\mathbf{d} \phi_\mathbf{k}^{\mathbf{d}_\mathbf{u}}(\mathbf{x})$$

or in matrix vector form $\mathbf{\Phi}(\mathcal{X}, \tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})\hat{\mathbf{f}}^\mathbf{d} = \mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})\hat{\mathbf{g}}^{\mathbf{d}_\mathbf{u}}$. Which is equivalent to the matrix decomposition $\mathbf{\Phi}(\mathcal{X}, \tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}) = \mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})\tilde{\mathbf{\Pi}}^\top$ where $\tilde{\mathbf{\Pi}} = (\delta_{\mathbf{l},\mathbf{k}})_{\mathbf{l}\in\tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}, \mathbf{k}\in\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}} \in \mathbb{R}^{|\tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}|\times|\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}|}$ is a canonical map. To sum this up, we multiply the matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))$ with the vector $\hat{\mathbf{f}}^\mathbf{d} \in \mathbb{C}^{|\mathcal{I}(U)|}$ through calculating

$$\begin{aligned}
\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))\hat{\mathbf{f}}^\mathbf{d} &= \sum_{\mathbf{u}\in U} \mathbf{\Phi}(\mathcal{X}, \mathcal{I}_{\mathbf{N}^\mathbf{u}}^\mathbf{d})\hat{\mathbf{f}}^{\mathbf{d},\mathbf{u}} \\
&= \sum_{\mathbf{u}\in U} \mathbf{\Phi}(\{\mathbf{x}_\mathbf{u} \mid \mathbf{x}\in\mathcal{X}\}, \tilde{\mathcal{I}}_{\mathbf{N}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})\hat{\mathbf{f}}^{\mathbf{d},\mathbf{u}} \\
&= \sum_{\mathbf{u}\in U} \mathbf{\Phi}(\{\mathbf{x}_\mathbf{u} \mid \mathbf{x}\in\mathcal{X}\}, \mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})\tilde{\mathbf{\Pi}}^\top\hat{\mathbf{f}}^{\mathbf{d},\mathbf{u}}.
\end{aligned}$$

We calculate the last sum with $|U|$ many NFMT. This leads us to a computational cost of $\mathcal{O}(\sum_{\mathbf{u}\in U}(|\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}|\log|\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}| + m_{\mathsf{NFFT}}^{|\mathbf{u}|}M))$. Additionally, it can be easily parallelized, because every summand can be computed independently. We summarize the this procedure in Algorithm 3

Input: Vector $\mathbf{d}\in\{\exp,\cos,\operatorname{alg}\}^d$, truncation set U, bandwidths $\mathbf{N}^\mathbf{u}\in(2\mathbb{N})^\mathbf{u}$ for $\mathbf{u}\in U$, coefficients $\hat{f}_\mathbf{k}\in\mathbb{C}$ for all $\mathbf{k}\in\mathcal{I}(U)$, nodes $\mathcal{X}\subset\mathbb{D}^\mathbf{d}$, $|\mathcal{X}| = M$

```
f ← 0
foreach u ∈ U                                   // This loop can be parallelized
do
    X̃ ← {(x_j)_{j∈u}}_{x∈X}
    ĝ_k ← { f̂_k, |supp k| = |u| ; 0, else } , k ∈ I^{d_u}_{N^u_u}
    Compute f ← f + Φ(X̃, I^{d_u}_{N^u_u}) g using a |u|-variate NFMT
end
```

$$\tilde{\mathcal{X}} \leftarrow \{(x_j)_{j\in\mathbf{u}}\}_{\mathbf{x}\in\mathcal{X}}, \quad \hat{g}_\mathbf{k} \leftarrow \begin{cases} \hat{f}_\mathbf{k}, & |\operatorname{supp}\mathbf{k}| = |\mathbf{u}| \\ 0, & \text{else} \end{cases}, \ \mathbf{k}\in\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}, \quad \mathbf{f}\leftarrow\mathbf{f}+\mathbf{\Phi}(\tilde{\mathcal{X}}, \mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})\mathbf{g}$$

Output: $\mathbf{f} = \mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))\hat{\mathbf{f}}$

Computational cost: $\mathcal{O}\left(\sum_{\mathbf{u}\in U}\left(|\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}|\log|\mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}}| + m_{\mathsf{NFFT}}^{|\mathbf{u}|}M\right)\right)$

Algorithm 3: Grouped transform for the fast evaluation of mixed polynomials $f^\mathbf{d}$ with a frequency set $\mathcal{I}(U)$, see (3.6).

Furthermore, the identity

$$\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U)) = \left(\tilde{\mathbf{\Pi}}\mathbf{\Phi}(\{\mathbf{x}_\mathbf{u} \mid \mathbf{x}\in\mathcal{X}\}, \mathcal{I}_{\mathbf{N}_\mathbf{u}^\mathbf{u}}^{\mathbf{d}_\mathbf{u}})^\top\right)_{\mathbf{u}\in U}^\top \tag{4.8}$$

holds. This (4.8) leads us to an algorithm for multiplying with the adjoint matrix $\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))^*$, because

$$\mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))^* = \left(\tilde{\mathbf{\Pi}}\mathbf{\Phi}(\{\mathbf{x_u} \mid \mathbf{x} \in \mathcal{X}\}, \mathcal{I}^{\mathbf{d_u}}_{\mathbf{N^u_u}})^*\right)_{\mathbf{u}\in U}$$

holds. Thus, we have a fast algorithm for the evaluation of the sum

$$k(\mathbf{k}) = \sum_{\mathbf{x}\in\mathcal{X}} h_{\mathbf{x}} \phi^{m,n}_{\mathbf{k}}(\mathbf{x}) \tag{4.9}$$

for coefficients $\mathbf{h} = (h_{\mathbf{x}})_{\mathbf{x}\in\mathcal{X}} \in \mathbb{C}^M$ at the nodes $\mathbf{k} \in \mathcal{I}(U)$, which we summarize as 4.

Input: Vector $\mathbf{d} \in \{\exp, \cos, \mathrm{alg}\}^d$, truncation set U, bandwidths $\mathbf{N^u} \in (2\mathbb{N})^{\mathbf{u}}$ for $\mathbf{u} \in U$, nodes $\mathcal{X} \subset \mathbb{D}^{\mathbf{d}}$, $|\mathcal{X}| = M$, coefficients $h_{\mathbf{x}} \in \mathbb{C}$ for $\mathbf{x} \in \mathcal{X}$

foreach $\mathbf{u} \in U$ `// This loop can be parallelized`
do
  $\tilde{\mathcal{X}} \leftarrow \{(x_j)_{j\in\mathbf{u}}\}_{\mathbf{x}\in\mathcal{X}}$
  Compute $\mathbf{g^u} \leftarrow \mathbf{\Phi}(\tilde{\mathcal{X}}, \mathcal{I}^{\mathbf{d_u}}_{\mathbf{N^u_u}})^*(h_{\mathbf{x}})_{\mathbf{x}\in\mathcal{X}}$ using a $|\mathbf{u}|$-variate $\mathsf{NFMT}^\top$
  $\mathbf{f^u} \leftarrow (g^{\mathbf{u}}_j)_{j\in\mathcal{I}^{\mathbf{d_u}}_{\mathbf{N^u_u}}}$
end

Output: $(\mathbf{f^u})_{\mathbf{u}\in U} = \mathbf{\Phi}(\mathcal{X}, \mathcal{I}(U))^*(h_{\mathbf{x}})_{\mathbf{x}\in\mathcal{X}}$

Computational cost: $\mathcal{O}\left(\sum_{\mathbf{u}\in U}\left(|\mathcal{I}^{\mathbf{d_u}}_{\mathbf{N^u_u}}| \log |\mathcal{I}^{\mathbf{d_u}}_{\mathbf{N^u_u}}| + m^{|\mathbf{u}|}_{\mathsf{NFFT}^\top} M\right)\right)$

Algorithm 4: Adjoint grouped transformation for the fast evaluation of the sum (4.9) for frequency sets $\mathcal{I}(U)$ defined in (3.6).

5 Numerical Experiments

In this subsection, we test the ANOVA approximation with the mixed bases on synthetic data. In Subsection 5.1, we show how the approximation procedure works and how we determine the bandwidths in this case. In Subsection 5.2 we compare the ANOVA approximation with the mixed bases to the ANOVA approximation with a fully periodic and a fully non-periodic basis, respectively. Furthermore, we compare analytic global sensitivity indices in Appendix A to approximated ones. Furthermore, we investigate here the empirical convergence behaviour of the different approximation methods for this function.
We have extended the ANOVAapprox software [22] with the algorithms listed in Section 4 and run all the following tests in this framework.
To determine the quality of the ANOVA approximation $\tilde{f}$ for a function f, we consider the mean squared error (MSE),

$$\mathrm{MSE}(f, \tilde{f}, \mathcal{X}_{\mathrm{test}}) := \frac{1}{|\mathcal{X}_{\mathrm{test}}|} \sum_{\mathbf{x}\in\mathcal{X}_{\mathrm{test}}} \left| f(\mathbf{x}) - \tilde{f}(\mathbf{x}) \right|^2,$$

at the nodes $\mathcal{X}_{\mathrm{test}}$.

5.1 ANOVA approximation with a mixed basis

In this subsection we approximate a function using the ANOVA approximation with the mixed basis. A special focus lays in the question, how we determine the truncation set U and the according bandwidths numerically. The function we are approximating in this section is

$$f_1 \colon [0,1]^4 \to \mathbb{C},\ f_1(\mathbf{x}) := \exp(\sin(2\pi x_1)x_2) + \cos(\pi x_3)x_4^2 + \frac{1}{10}\sin^2(2\pi x_1) + 5\sqrt{x_2 x_4 + 1}.$$

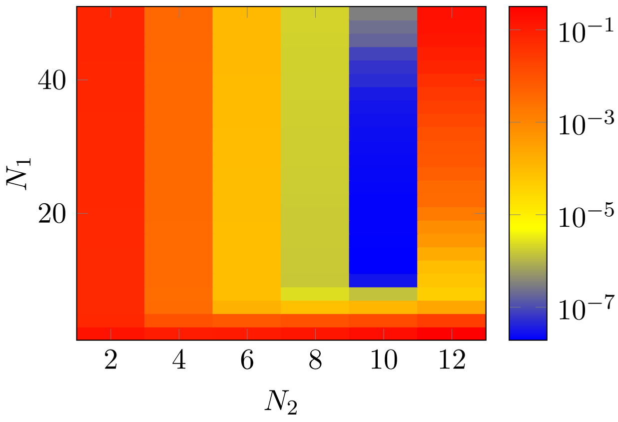

Figure 2: Mean squared errors for U_2 and bandwidths $\mathbf{N^u} = (|\{i\} \cap \mathbf{u}|N_{|\mathbf{u}|})_{i=1}^4$ and N_1 and N_2 for f_1.

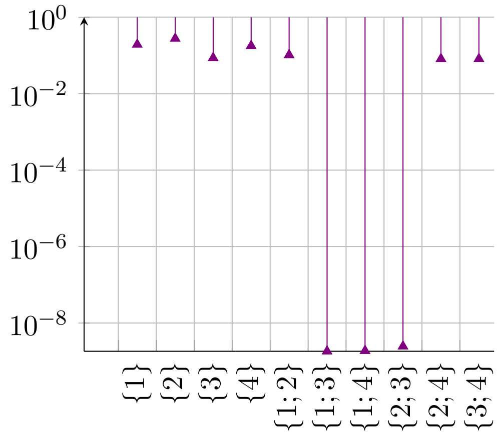

Figure 3: Approximated global sensitivity indices for U_2 and bandwidths $\mathbf{N^u} = (|\{i\} \cap \mathbf{u}|N_{|\mathbf{u}|})_{i=1}^4$ and $N_1 = 12$ and $N_2 = 10$ for f_1.

This function f_1 is smoothly periodizable in the first dimension, i.e. $f_1^{\text{per}}\colon \mathbb{T} \times [0,1]^3 \to \mathbb{C}$, $f_1^{\text{per}}(\mathbf{x}) := f_1(\mathbf{x})$ is infinitely differentiable. Furthermore, the function acts in the third dimension only as a cosine function. This leads us to use the mixed basis $\phi_{\mathbf{k}}^{\mathbf{d}_1}$ with $\mathbf{d}_1 := (\exp, \text{alg}, \cos, \text{alg})^\top$ for $\mathbf{k} \in \mathbb{K}^{\mathbf{d}_1}$ for approximating the function f_1.
For this approximation we restrict us to only 1000 nodes $\mathcal{X}$ in $\mathbb{D}^{\mathbf{d}_1}$ distributed with the density

$$\omega^{\mathbf{d}_1}(\mathbf{x}) = \frac{1}{\pi^2\sqrt{x_2 - x_2^2}\sqrt{x_4 - x_4^2}}.$$

Furthermore, we are given another 10000 nodes $\mathcal{X}_{\text{test}}$ in $\mathbb{D}^{\mathbf{d}_1}$ distributed with the density $\omega^{\mathbf{d}_1}$ for evaluating the mean squared error.
We follow the steps from Section 3.2.3. The function has only one dimensional and two-dimensional interactions between variables. Thus, we set $d_s = 2$ and consider the superposition set $U_2 = \{\{1\},\{2\},\{3\},\{4\},\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\}$ from (3.8). We choose one bandwidth parameter $N_1 \in (2\mathbb{N})$ for the one-dimensional frequency sets, e.g. every non-zero entry of $\mathbf{N}^{\{1\}}$, $\mathbf{N}^{\{2\}}$, $\mathbf{N}^{\{3\}}$ and $\mathbf{N}^{\{4\}}$ is set to N_1. Furthermore, we choose another bandwidth parameter $N_2 \in (2\mathbb{N})$ for the two-dimensional frequency sets, e.g. every non-zero entry of $\mathbf{N}^{\{1,2\}}$, $\mathbf{N}^{\{1,3\}}$, $\mathbf{N}^{\{1,4\}}$, $\mathbf{N}^{\{2,3\}}$, $\mathbf{N}^{\{2,4\}}$ and $\mathbf{N}^{\{3,4\}}$ is set to N_2. In short form we write this as $\mathbf{N^u} = (|\{i\} \cap \mathbf{u}|N_{|\mathbf{u}|})_{i=1}^4$ for $\mathbf{u} \in U_2$. We call the approximation of the function f_1 using the 1000 nodes $\mathcal{X}$ and the bandwidth parameters N_1 and N_2 $\tilde{f}_1^{N_1,\,N_2}$. We determine the optimal bandwidth parameters N_1 and N_2 numerically by minimizing the mean squared error $\text{MSE}(f_1, \tilde{f}_1^{N_1,\,N_2}, \mathcal{X}_{\text{test}})$, i.e.

$$(N_i)_{i=1}^2 = \operatorname*{arg\,min}_{(N_i)_{i=1}^2 \in (2\mathbb{N})^2} \text{MSE}(f_1, \tilde{f}_1^{N_1,N_2}, \mathcal{X}_{\text{test}}). \tag{5.1}$$

In other words, we use cross validation to determine the bandwidth parameters N_1 and N_2.

We see in Figure 2 the MSE for some choices of N_1 and N_2. The ANOVA approximation does the best approximation for $N_1 = 12$ and $N_2 = 10$. In Figure 3 the approximated GSIs for this approximation are shown. To this end we choose the threshold $\theta = 10^{-3}$ and find through (3.9) the truncation set $U_\theta = \{\{1\},\{2\},\{3\},\{4\},\{1,2\},\{2,4\},\{3,4\}\}$. Next, we find better bandwidths $\mathbf{N^u}$

Step	$N_{\{1\}}$	$N_{\{2\}}$	$N_{\{3\}}$	$N_{\{4\}}$	$N_{\{1,2\}}$	$N_{\{2,4\}}$	$N_{\{3,4\}}$	MSE
1	12	12	12	12	10	10	10	$1.31369 \cdot 10^{-8}$
2	12	12	12	12	10	10	12	$1.36285 \cdot 10^{-8}$
3	12	12	12	12	10	10	8	$1.25704 \cdot 10^{-8}$
4	12	12	12	12	10	10	6	$1.2035 \cdot 10^{-8}$
5	12	12	12	12	10	10	4	$3.97122 \cdot 10^{-8}$
6	12	12	12	12	10	12	6	$1.24734 \cdot 10^{-8}$
7	12	12	12	12	10	8	6	$1.15079 \cdot 10^{-8}$
8	12	12	12	12	10	6	6	$1.13142 \cdot 10^{-8}$
9	12	12	12	12	10	4	6	$1.10034 \cdot 10^{-8}$
10	12	12	12	12	10	2	6	$4.68726 \cdot 10^{-3}$
11	12	12	12	12	12	4	6	$1.20275 \cdot 10^{-10}$
12	12	12	12	12	14	4	6	$4.27822 \cdot 10^{-11}$
13	12	12	12	12	16	4	6	$5.42373 \cdot 10^{-11}$
14	12	12	12	14	14	4	6	$4.28631 \cdot 10^{-11}$
15	12	12	12	10	14	4	6	$4.2808 \cdot 10^{-11}$
16	12	12	14	12	14	4	6	$4.29279 \cdot 10^{-11}$
17	12	12	10	12	14	4	6	$4.26967 \cdot 10^{-11}$
18	12	12	8	12	14	4	6	$4.25943 \cdot 10^{-11}$
19	12	12	6	12	14	4	6	$4.24758 \cdot 10^{-11}$
20	12	12	4	12	14	4	6	$4.25071 \cdot 10^{-11}$
21	12	14	6	12	14	4	6	$4.24995 \cdot 10^{-11}$
22	12	10	6	12	14	4	6	$4.23524 \cdot 10^{-11}$
23	12	8	6	12	14	4	6	$4.17446 \cdot 10^{-11}$
24	12	6	6	12	14	4	6	$4.35334 \cdot 10^{-11}$
25	14	8	6	12	14	4	6	$6.80995 \cdot 10^{-12}$
26	16	8	6	12	14	4	6	$6.58355 \cdot 10^{-12}$
27	18	8	6	12	14	4	6	$6.5848 \cdot 10^{-12}$

Table 1: Mean squared errors for U_θ and bandwidths $\mathbf{N}^{\mathbf{u}} = (|\{i\} \cap \mathbf{u}|\, N_{|\mathbf{u}|})_{i=1}^4$ for some choices of the parameters N_1 and N_2.

for U_θ. To do this we introduce a new set of bandwidth parameters $N_{\mathbf{u}} \in (2\mathbb{N})$ for $\mathbf{u} \in U_\theta$, i.e. one parameter for every bandwidth. We get the bandwidths $\mathbf{N}^{\mathbf{u}}$ by setting every non-zero entry to $N_{\mathbf{u}}$, i.e. $\mathbf{N}^{\mathbf{u}} = (|\{i\} \cap \mathbf{u}|N_{\mathbf{u}})_{i=1}^4$. We optimise these bandwidth parameters one by one, starting with the parameters corresponding to the two-dimensional bandwidths. We do this through increasing the parameter firstly bigger until the MSE gets bigger. If the MSE gets bigger in the first step, we decrease the parameter until the MSE gets bigger. Then we use the parameter which has generated the minimal MSE. As a starting point we use the bandwidths $\mathbf{N}^{\mathbf{u}} = (|\{i\} \cap \mathbf{u}|\, N_{|\mathbf{u}|})_{i=1}^4$ generated by the optimal parameters $N_1 = 12$ and $N_2 = 10$ of the previous approximation step. In Table 1 we show the parameters we tried to find the optimal ones.

We get the bandwidths $\mathbf{N}^{\mathbf{u}} = (|\{i\} \cap \mathbf{u}|N_{\mathbf{u}})_{i=1}^4$ with the parameters $N_{\{1\}} = 16$, $N_{\{2\}} = 8$, $N_{\{3\}} = 6$, $N_{\{4\}} = 12$, $N_{\{1,2\}} = 14$, $N_{\{2,4\}} = 4$, and $N_{\{3,4\}} = 6$. Finally, we repeat the one by one optimizing procedure again with all parameters for the bandwidths, e.g. we consider every non-zero entry of each bandwidth as one parameter. As result, we get the bandwidths

$$\mathbf{N}^{\{1\}} = \begin{pmatrix} 16 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{N}^{\{2\}} = \begin{pmatrix} 0 \\ 8 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{N}^{\{3\}} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \qquad \mathbf{N}^{\{4\}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 10 \end{pmatrix},$$

$$\mathbf{N}^{\{1,2\}} = \begin{pmatrix} 16 \\ 8 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{N}^{\{2,3\}} = \begin{pmatrix} 0 \\ 2 \\ 4 \\ 0 \end{pmatrix}, \text{ and } \qquad \mathbf{N}^{\{2,4\}} = \begin{pmatrix} 0 \\ 8 \\ 0 \\ 8 \end{pmatrix}$$

with a mean squared error of $9.74704 \cdot 10^{-14}$. We point out that this is quite a good approximation since we use only 1000 nodes $\mathcal{X}$. All in all, we approximate the function f_1 with a sum of 365 basis functions combined with the same number of coefficients.

5.2 Comparison of analytic and approximated global sensitivity indices

In this subsection we approximate a function multiple times with different numbers of nodes M. This time we restrict ourselves to uniformly sampled nodes. This has the advantage that we can compare the ANOVA approximation with the mixed basis to the Fourier basis and with the half period cosine basis approximation. We also compare the approximated GSIs with analytically calculated ones. In order to do this, we consider the function

$$f_2 \colon [0,1]^4 \to \mathbb{C},\ f_2(x_1,x_2,x_3,x_4) \coloneqq (2x_1-1)^2 x_3 + 10\sin(2\pi x_1)\left(x_2 - \frac{1}{2}\right)^2 + \exp(x_3).$$

The function f_2 does not depend on the variable x_4. Furthermore, the function has the same values at the boundaries in dimension one and two, e.g.

$$\begin{aligned} f_2(0,x_2,x_3,x_4) &= f_2(1,x_2,x_3,x_4),\ \forall x_2,x_3,x_4 \in [0,1] \\ f_2(x_1,0,x_3,x_4) &= f_2(x_1,1,x_3,x_4),\ \forall x_1,x_3,x_4 \in [0,1]. \end{aligned}$$

Thus, we should use the Fourier basis for the first two coordinates and the half period cosine basis for the third, i.e. $\mathbf{d}_2 \coloneqq (\exp,\exp,\cos,\cos)^\top$. Furthermore, we test two more ANOVA approximations without mixed bases, namely one with a Fourier basis and one with a half period cosine basis. In the appendix A we calculate the analytic GSIs of this function f_2. The results are

$$\begin{aligned} \rho(\{1\}, f_2) &= \frac{133}{59 + 600\mathrm{e} - 180\mathrm{e}^2} &\approx 0.369507, \\ \rho(\{3\}, f_2) &= \frac{-530 + 1800\mathrm{e} - 540\mathrm{e}^2}{177 + 1800\mathrm{e} - 540\mathrm{e}^2} &\approx 0.345259, \\ \rho(\{1,2\}, f_2) &= \frac{100}{59 + 600\mathrm{e} - 180\mathrm{e}^2} &\approx 0.277825, \\ \rho(\{1,3\}, f_2) &= \frac{8}{177 + 1800\mathrm{e} - 540\mathrm{e}^2} &\approx 0.007409, \end{aligned}$$

and the other analytic GSIs are zero.

We now compare this with the approximated GSIs. For ANOVA approximation we use $M = 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000$ and 50000 uniformly distributed nodes $\mathcal{X}$. For this function f_2 we consider the superposition set U_{d_s} with $d_s = 2$, since the function f_2 has only one-dimensional and two dimensional interactions between variables. In this example we restrict ourselves to two bandwidth parameters, N_1 for one-dimensional bandwidths and N_2 for the two-dimensional bandwidths. For $M \leq 10000$ we determine the bandwidth parameters N_1 and N_2 numerically like in (5.1). The results are shown in the Table 2. We obtain the bandwidths for $M > 10000$ by extrapolating the previously determined optimal bandwidths.

In Figure 4 we plot the resulting mean squared errors. Here we notice that the error for the approximation with the Fourier basis decays with the rate M^{-1}. This decay rate M^{-1} is optimal, since the periodization $f_2^{\mathrm{per}} \colon \mathbb{T}^4 \to \mathbb{C},\ f_2^{\mathrm{per}}(\mathbf{x}) \coloneqq f_2(\mathbf{x})$ is not continuous, see [14]. The error of the approximation with the mixed basis and the half period cosine basis decays with the rate $M^{-\frac{3}{2}}$, while the approximation with the mixed basis gives a better constant. This rate $M^{-\frac{3}{2}}$ is optimal

M	**cos**		$\mathbf{d}_2$		**exp**	
	N_1	N_2	N_1	N_2	N_1	N_2
50	4	2	4	2	4	2
100	4	4	4	4	4	4
200	6	4	6	4	14	4
500	12	8	10	8	32	6
1000	18	10	14	10	76	6
2000	28	14	24	14	150	10
5000	56	22	40	22	300	14
10000	70	32	60	32	720	18
20000	170	46	110	46	1962	26
50000	382	76	224	76	6548	40

Table 2: Optimal bandwidths (for $M \leq 10000$) $\mathbf{N}^{\mathbf{u}} = (|\{i\} \cap \mathbf{u}|\, N_{|\mathbf{u}|})_{i=1}^{4}$ for U_2 for f_2 approximated at M training nodes and (for $M > 10000$) extrapolated bandwidths.

for the approximation with the cosine basis. Furthermore, we see that the approximation with the mixed basis decays with the same rate as the approximation with the half period cosine basis but provides a better constant factor.

Next, in Figure 5, we compare the approximated GSIs with the analytic GSIs and see that they converge. The approximated GSIs using the Fourier basis converge slower than the approximated GSIs using the mixed basis. In Figure 6 we consider the individual approximated GSIs for different numbers of training nodes. Here we notice that the approximated GSIs using the Fourier basis performs particularly poorly in the dimensions where the function f_2 is not continuously periodizable, e.g. for $\mathbf{u} = \{3\}$ we have particularly large deviations from the analytic GSI. Furthermore, we see for example at $\mathbf{u} = \{1, 2\}$ that the approximated GSIs using the half period cosine basis converge more slowly towards the analytic GSI than approximated GSIs using the Fourier basis. The approximated GSIs using the mixed basis combines the positive properties of the other two ANOVA approximations and therefore converges much faster.

References

[1] C. C. Aggarwal. *Data Classification: Algorithms and Applications*. CRC Press, Boca Ration, FL, US, 2015.

[2] F. Bartel, D. Potts, and M. Schmischke. Grouped transformations and regularization in high-dimensional explainable anova approximation. *SIAM J. Sci. Comput.*, 44(3):A1606–A1631, 2022. DOI: 10.1137/20M1374547.

[3] A. Böttcher, S. Kunis, and D. Potts. Probabilistic spherical Marcinkiewicz-Zygmund inequalities. *J. Approx. Theory*, 157:113–126, 2009.

[4] A. Böttcher and D. Potts. Probability against condition number and sampling of multivariate trigonometric random polynomials. *Electron. Trans. Numer. Anal.*, 26:178–189, 2007.

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

[6] C. Gu. *Smoothing Spline ANOVA Models*. Springer, New York, NY, US, 2013. DOI: 10.1007/978-1-4614-5369-7.

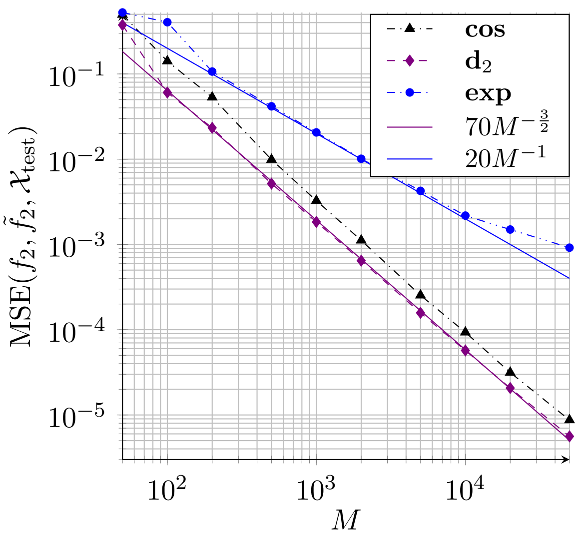

Figure 4: Mean squared errors of the ANOVA approximation $\tilde{f}_2$ to f_2 for the bandwidths from Table 2 and M training nodes, evaluated at 10000 nodes.

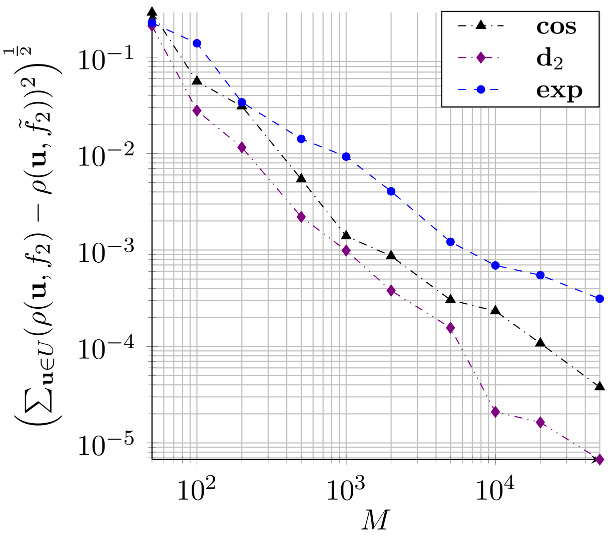

Figure 5: Deviation of approximated global sensitivity indices for bandwidths from Table 2 and M training nodes to the analytic GSIs.

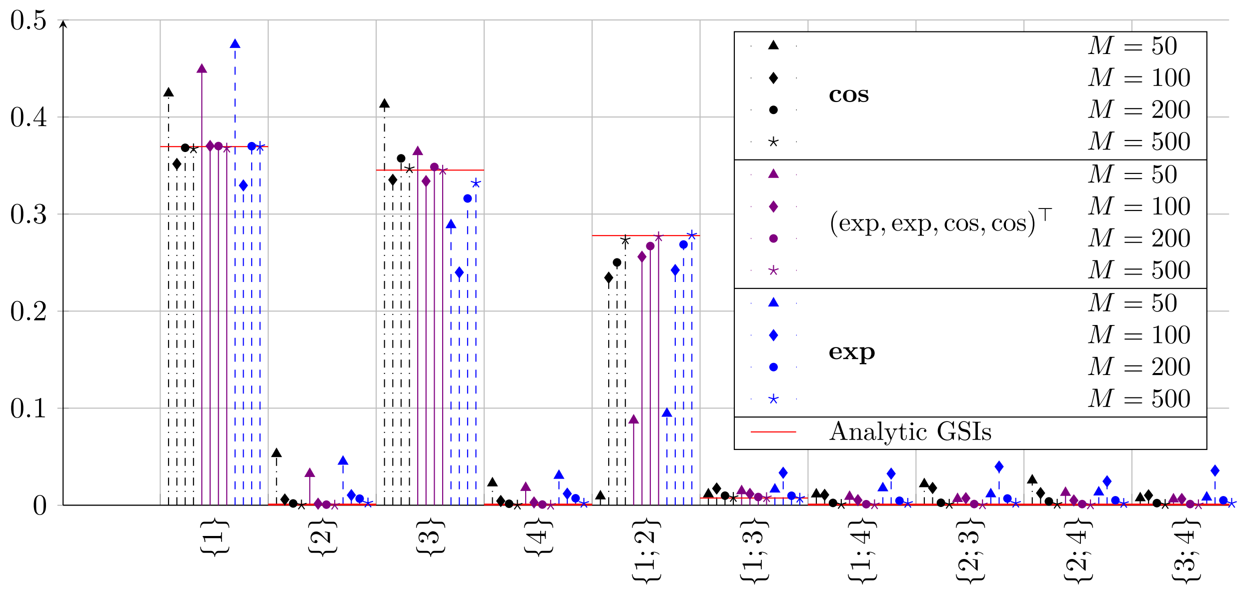

Figure 6: Global sensitivity indices for U_2 and bandwidths from Table 2 for f_2 trained on M randomly chosen nodes.

[7] A. Hashemi, H. Schaeffer, R. Shi, U. Topcu, G. Tran, and R. Ward. Generalization bounds for sparse random feature expansions. *Applied and Computational Harmonic Analysis*, 62:310–330, 2023. ISSN: 1063-5203. DOI: `https://doi.org/10.1016/j.acha.2022.08.003`.

[8] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning - Data Mining, Inference, and Prediction.* Springer, New York, NY, US, 2013. ISBN: 978-0-387-21606-5.

[9] M. Holtz. *Sparse grid quadrature in high dimensions with applications in finance and insurance*, volume 77 of *Lecture Notes in Computational Science and Engineering.* Springer-Verlag, Berlin, 2011, pages viii+182. DOI: `10.1007/978-3-642-16004-2`.

[10] J. Keiner, S. Kunis, and D. Potts. Using NFFT3 - a software library for various nonequispaced fast Fourier transforms. *ACM Trans. Math. Software*, 36:Article 19, 1–30, 2009.

[11] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski. On decompositions of multivariate functions. *Math. Comput.*, 79(270):953–966, 2010.

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

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

[14] G. Plonka, D. Potts, G. Steidl, and M. Tasche. *Numerical Fourier Analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser, 2018. ISBN: 978-3-030-04305-6. DOI: `10.1007/978-3-030-04306-3`.

[15] D. Potts. Fast algorithms for discrete polynomial transforms on arbitrary grids. *Linear Algebra Appl.*, 366:353–370, 2003.

[16] D. Potts and M. Schmischke. Approximation of high-dimensional periodic functions with Fourier-based methods. *SIAM J. Numer. Anal.*, 59(5):2393–2429, 2021. DOI: `10.1137/20M1354921`.

[17] D. Potts and M. Schmischke. Learning multivariate functions with low-dimensional structures using polynomial bases. *J. Comput. Appl. Math.*, 403:113821, 2022. ISSN: 0377-0427. DOI: `10.1016/j.cam.2021.113821`.

[18] D. Potts and M. Schmischke. Interpretable approximation of high-dimensional data. *SIAM J. Math. Data Sci.*, 3(4):1301–1323, 2021. DOI: `10.1137/21M1407707`. eprint: `2103.13787`.

[19] D. Potts and M. Schmischke. Interpretable transformed ANOVA approximation on the example of the prevention of forest fires. *Front. Appl. Math. Stat.*, 8:795250, 2022. DOI: `10.3389/fams.2022.795250`.

[20] H. Rabitz and O. F. Alis. General foundations of high dimensional model representations. *J. Math. Chem.*, 25:197–233, 1999. DOI: `10.1023/A:1019188517934`.

[21] M. Schmischke. *Interpretable Approximation of High-Dimensional Data based on the ANOVA Decomposition.* Dissertation. Universitätsverlag Chemnitz, 2022. ISBN: 978-3-96100-164-4.

[22] M. Schmischke. ANOVAapprox.jl Julia Package. DOI: `10.5281/zenodo.5657977`. Contributor: F. Bartel.

[23] I. M. Sobol. On sensitivity estimation for nonlinear mathematical models. *Keldysh Applied-Mathematics Institute*, 1:112–118, 1990.

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

[25] C.-F. Wu and M. Hamada. *Experiments: planning, analysis and parameter design optimization.* English. Wiley series in probability and statistics. Wiley, Hoboken, NJ, third edition edition, 2021. ISBN: 9781119470007. DOI: 10.1002/9781119470007.

A Analytic Calculation of Global Sensitivity Indices

In the following we calculate the analytic GSIs for the function

$$\begin{aligned} &f_2\colon \mathbb{D}^{\mathbf{d}_2} \to \mathbb{C}, \\ &f_2(x_1, x_2, x_3, x_4) \coloneqq (2x_1 - 1)^2 x_3 + 10\sin(2\pi x_1)\left(x_2 - \frac{1}{2}\right)^2 + \exp(x_3) \\ &\qquad = 4x_1^2 x_3 - 4x_1 x_3 + x_3 + 10\sin(2\pi x_1)x_2^2 - 10\sin(2\pi x_1)x_2 + \frac{5}{2}\sin(2\pi x_1) + \exp(x_3) \end{aligned}$$

with $\mathbf{d}_2 = (\exp, \exp, \cos, \cos)^\top$ using the basis $\mathcal{B}^{\mathbf{d}_2}$. First we calculate the mixed coefficients $c_{\mathbf{k}}^{\mathbf{d}_2}(f_2)$, exploiting linearity. For this purpose we define

$$\begin{aligned} &h_j\colon \mathbb{D}^{\mathbf{d}_2} \to \mathbb{C},\ j = 1, \dots, 7, \\ &h_1(x_1, x_2, x_3, x_4) = x_1^2 x_3, \\ &h_2(x_1, x_2, x_3, x_4) = x_1 x_3, \\ &h_3(x_1, x_2, x_3, x_4) = x_3, \\ &h_4(x_1, x_2, x_3, x_4) = \sin(2\pi x_1)x_2^2, \\ &h_5(x_1, x_2, x_3, x_4) = \sin(2\pi x_1)x_2, \\ &h_6(x_1, x_2, x_3, x_4) = \sin(2\pi x_1),\ \text{and} \\ &h_7(x_1, x_2, x_3, x_4) = \exp(x_3). \end{aligned}$$

and observe

$$c_{\mathbf{k}}^{\mathbf{d}_2}(f_2) = 4c_{\mathbf{k}}^{\mathbf{d}_2}(h_1) - 4c_{\mathbf{k}}^{\mathbf{d}_2}(h_2) + c_{\mathbf{k}}^{\mathbf{d}_2}(h_3) + 10c_{\mathbf{k}}^{\mathbf{d}_2}(h_4) - 10c_{\mathbf{k}}^{\mathbf{d}_2}(h_5) + \frac{5}{2}c_{\mathbf{k}}^{\mathbf{d}_2}(h_6) + c_{\mathbf{k}}^{\mathbf{d}_2}(h_7) \tag{A.1}$$

for all $\mathbf{k} \in \mathbb{K}^{\mathbf{d}_2}$. We know that, if $f \in \mathrm{L}_2(\mathbb{D}^{\mathbf{d}})$ is a function given as product $f(\mathbf{x}) = \prod_{j=1}^d f_j^{d_j}(x_j)$ of functions $f_j^{d_j} \in \mathrm{L}_2(\mathbb{D}^{d_j})$, $j = 1, \dots, d$, then for all $\mathbf{k} \in \mathbb{K}^{\mathbf{d}}$ we can decompose the mixed coefficients, i.e.

$$c_{\mathbf{k}}^{\mathbf{d}}(f) = \prod_{j=1}^{d} c_{k_j}^{d_j}(f_j^{d_j}(x_j)).$$

Thus, we decompose the functions h_i into

$$\begin{aligned} &g_j^{\exp}\colon \mathbb{D}^{\exp} \to \mathbb{C},\ j = 1, \dots, 4, && g_j^{\cos}\colon \mathbb{D}^{\cos} \to \mathbb{C},\ j = 1, \dots, 3, \\ &g_1^{\exp}(x) = 1, && g_1^{\cos}(x) = 1, \\ &g_2^{\exp}(x) = x, && g_2^{\cos}(x) = x, \\ &g_3^{\exp}(x) = x^2, && g_3^{\cos}(x) = \exp(x), \\ &g_4^{\exp}(x) = \sin(2\pi x). \end{aligned}$$

Next we calculate the Fourier coefficients and the cosine coefficients of these functions. We observe $c_k^{\exp}(g_1^{\exp}) = \delta_{k,0}$ and $c_k^{\cos}(g_1^{\cos}) = \delta_{k,0}$ because of the orthogonality of the basis functions and $\phi_0^{\exp} = \phi_0^{\cos} = 1$. We start with the zeroth Fourier coefficients and cosine coefficients and observe

$$c_0^{\exp}(g_2^{\exp}) = c_0^{\cos}(g_2^{\cos}) = \frac{1}{2}, \qquad c_0^{\exp}(g_3^{\exp}) = \frac{1}{3} \qquad \text{and} \qquad c_0^{\cos}(g_3^{\cos}) = \mathrm{e} - 1$$

through $c_0^{\exp} = c_0^{\cos} = \int_0^1 f(x)\,\mathrm{d}x$. For the case $k \neq 0$ we observe

$$c_k^{\exp}(g_2^{\exp}) = \frac{1}{2\pi k}\mathrm{i}, \qquad c_k^{\cos}(g_2^{\cos}) = \sqrt{2}\frac{(-1)^k - 1}{\pi^2 k^2},$$
$$c_k^{\exp}(g_3^{\exp}) = \frac{1}{2\pi^2 k^2} + \frac{1}{2\pi k}\mathrm{i} \qquad \text{and} \qquad c_k^{\cos}(g_3^{\cos}) = \sqrt{2}\frac{(-1^k)\mathrm{e} - 1}{\pi^2 k^2 + 1}.$$

To this end we use the identity $g_4^{\exp}(x) = \sin(2\pi x) = \frac{1}{2\mathrm{i}}\exp(2\pi\mathrm{i}x) - \frac{1}{2\mathrm{i}}\exp(-2\pi\mathrm{i}x)$ for the coefficients $c_k^{\exp}(g_4^{\exp})$ and get $c_1^{\exp}(g_4^{\exp}) = \frac{1}{2\mathrm{i}}$, $c_{-1}^{\exp}(g_4^{\exp}) = -\frac{1}{2\mathrm{i}}$, and the ohter coefficients are zero. Next, we consider the Fourier cosine coefficients $c_{\mathbf{k}}$ of the functions h_i, $i = 1, \ldots, 7$ and obtain

$$c_{\mathbf{k}}^{\mathbf{d}_2}(h_1) = \begin{cases} \frac{1}{6}, & k_1, k_2, k_3, k_4 = 0 \\ \frac{1}{4\pi^2 k_1^2} + \frac{1}{4\pi k_1}\mathrm{i}, & k_2, k_3, k_4 = 0, k_1 \neq 0 \\ \sqrt{2}\frac{(-1)^{k_3}-1}{3\pi^2 k_3^2}, & k_1, k_2, k_4 = 0, k_3 \neq 0 \\ \sqrt{2}\frac{(-1)^{k_3}-1}{2\pi^4 k_1^2 k_3^2} + \sqrt{2}\frac{(-1)^{k_3}-1}{2\pi^3 k_1 k_3^2}, & k_2, k_4 = 0, k_1, k_3 \neq 0 \\ 0, & \text{else} \end{cases},$$

$$c_{\mathbf{k}}^{\mathbf{d}_2}(h_2) = \begin{cases} \frac{1}{4}, & k_1, k_2, k_3, k_4 = 0 \\ \frac{1}{4\pi k_1}\mathrm{i}, & k_2, k_3, k_4 = 0, k_1 \neq 0 \\ \sqrt{2}\frac{(-1)^{k_3}-1}{2\pi^2 k_3^2}, & k_1, k_2, k_4 = 0, k_3 \neq 0 \\ 1\sqrt{2}\frac{(-1)^{k_3}-1}{2\pi^3 k_1 k_3^2}\mathrm{i}, & k_2, k_4 = 0, k_1, k_3 \neq 0 \\ 0, & \text{else} \end{cases},$$

$$c_{\mathbf{k}}^{\mathbf{d}_2}(h_3) = \begin{cases} \frac{1}{2}, & k_1, k_2, k_3, k_4 = 0 \\ \sqrt{2}\frac{(-1)^{k_3}-1}{\pi^2 k_3^2}, & k_1, k_2, k_4 = 0, k_3 \neq 0 \\ 0, & \text{else} \end{cases},$$

$$c_{\mathbf{k}}^{\mathbf{d}_2}(h_4) = \begin{cases} -\frac{\mathrm{i}}{6}, & k_2, k_3, k_4 = 0, k_1 = 1 \\ \frac{\mathrm{i}}{6}, & k_2, k_3, k_4 = 0, k_1 = -1 \\ \frac{1}{4\pi k_1} - \frac{1}{2\pi^2 k_1^2}\mathrm{i}, & k_3, k_4 = 0, k_1 = 1, k_2 \neq 0 \\ -\frac{1}{4\pi k_1} + \frac{1}{2\pi^2 k_1^2}\mathrm{i}, & k_3, k_4 = 0, k_1 = -1, k_2 \neq 0 \\ 0, & \text{else} \end{cases},$$

$$c^{\mathbf{d}_2}\mathbf{k}(h_5) = \begin{cases} -\frac{\mathrm{i}}{4}, & k_2, k_3, k_4 = 0, k_1 = 1 \\ \frac{\mathrm{i}}{4}, & k_2, k_3, k_4 = 0, k_1 = -1 \\ \frac{1}{4\pi k_2}, & k_3, k_4 = 0, k_1 = 1, k_2 \neq 0 \\ -\frac{1}{4\pi k_2}, & k_3, k_4 = 0, k_1 = -1, k_2 \neq 0 \\ 0, & \text{else} \end{cases},$$

$$c_{\mathbf{k}}^{\mathbf{d}_2}(h_6) = \begin{cases} -\frac{\mathrm{i}}{2}, & k_2, k_3, k_4 = 0, k_1 = 1 \\ \frac{\mathrm{i}}{2}, & k_2, k_3, k_4 = 0, k_1 = -1 \\ 0, & \text{else} \end{cases}, \text{ and}$$

$$c_{\mathbf{k}}^{\mathbf{d}_2}(h_7) = \begin{cases} \mathrm{e} - 1, & k_1, k_2, k_3, k_4 = 0 \\ \sqrt{2}\frac{(-1)^{k_3}\mathrm{e}-1}{\pi^2 k_3^2+1}, & k_1, k_2, k_4 = 0, k_3 \neq 0 \\ 0, & \text{else} \end{cases}.$$

Finally, using the (A.1) we calculate the Fourier cosine coefficients $c^{\mathbf{d}_2}_{\mathbf{k}}(f_2)$ for the function f_2,

$$c^{\mathbf{d}_2}_{\mathbf{k}}(f_2) = \begin{cases} \mathrm{e} - \frac{5}{6}, & k_1, k_2, k_3, k_4 = 0 \\ -\frac{1}{\pi^2} - \frac{5}{12}\mathrm{i}, & k_2, k_3, k_4 = 0, k_1 = 1 \\ -\frac{1}{\pi^2} + \frac{5}{12}\mathrm{i}, & k_2, k_3, k_4 = 0, k_1 = -1 \\ \frac{1}{\pi^2 k_1^2}, & k_2, k_3, k_4 = 0, |k_1| \geq 2 \\ \sqrt{2}\frac{(-1)^{k_3}-1}{\pi^2 k_3^2} + \sqrt{2}\frac{(-1)^k_3\mathrm{e}-1}{\pi^2 k_3^2+1}, & k_1, k_2, k_4 = 0, k_3 \neq 0 \\ -\frac{5}{2\pi^2 k_2^2}\mathrm{i}, & k_3, k_4 = 0, k_1 = 1, k_2 \neq 0 \\ \frac{5}{2\pi^2 k_2^2}\mathrm{i}, & k_3, k_4 = 0, k_1 = -1, k_2 \neq 0 \\ 2\sqrt{2}\frac{(-1)^{k_3}-1}{\pi^4 k_1^2 k_3^2}, & k_2, k_4 = 0, k_1, k_3 \neq 0 \\ 0, & \text{else} \end{cases}.$$

Furthermore, we consider the norm of the function f_2 and observe $\|f_2\|_{\mathrm{L}_2(\mathbb{D}^{\mathbf{d}_2})} = \frac{103}{120} + \frac{\mathrm{e}^2}{2}$. Using this we observe the variance $\sigma^2(f_2) = \frac{59}{360} + \frac{5}{3}\mathrm{e} - \frac{\mathrm{e}^2}{2}$. Next, we consider the ANOVA terms $\sigma^2(f_{\mathbf{u}})$ for subsets of indices $\mathbf{u} \subseteq [4]$ using their series representation $\sigma^2(f_{\mathbf{u}}) = \sum_{\substack{\mathbf{k}\in\mathbb{K}^{\mathbf{d}_2} \\ \operatorname{supp}\mathbf{k}=\mathbf{u}}} |c^{\mathbf{d}_2}_{\mathbf{k}}(f)|^2$. For the further calculation we need $\sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90}$ and $\sum_{k=1}^{\infty} \frac{1}{(2k-1)^4} = \frac{\pi^4}{96}$. Using this we get the variances $\sigma^2(f_{\mathbf{u}})$ of the ANOVA terms $f_{\mathbf{u}}$,

$$\begin{aligned} \sigma^2(f_{\{1\}}) &= \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} |c^{\mathbf{d}_2}_{k\mathbf{e}_1}|^2 = \frac{2}{\pi^4} + \frac{25}{72} + 2\sum_{k=2}^{\infty} \frac{1}{\pi^4 k^4} = \frac{133}{360} \\ \sigma^2(f_{\{1,2\}}) &= \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \sum_{\substack{j=-\infty \\ j\neq 0}}^{\infty} |c^{\mathbf{d}_2}_{k\mathbf{e}_1+j\mathbf{e}_2}|^2 = 4\sum_{k=1}^{\infty} \frac{25}{4\pi^4 k^4} = \frac{5}{18} \\ \sigma^2(f_{\{1,3\}}) &= \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \sum_{j=1}^{\infty} |c^{\mathbf{d}_2}_{k\mathbf{e}_1+j\mathbf{e}_3}|^2 = 2\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \frac{8\left((-1)^j - 1\right)^2}{\pi^8 k^4 j^4} = \frac{1}{135}. \end{aligned}$$

Since the other mixed coefficients except $c^{\mathbf{d}_2}_{k\mathbf{e}_3}(f_2)$ are zero, we get the variance $\sigma^2(f_{\{3\}})$ through the theorem of Parseval,

$$\sigma^2(f_{\{3\}}) = \sigma^2(f_2) - \sigma^2(f_{\{1\}}) - \sigma^2(f_{\{1,2\}}) - \sigma^2(f_{\{1,3\}}) = -\frac{53}{108} + \frac{5}{3}\mathrm{e} - \frac{\mathrm{e}^2}{2}.$$

Finally, we get the analytic global sensitivity indices

$$\begin{aligned} \rho(\{1\}, f_2) &= \frac{133}{59 + 600\mathrm{e} - 180\mathrm{e}^2} &&\approx 0.369507, \\ \rho(\{3\}, f_2) &= \frac{-530 + 1800\mathrm{e} - 540\mathrm{e}^2}{177 + 1800\mathrm{e} - 540\mathrm{e}^2} &&\approx 0.345259, \\ \rho(\{1,2\}, f_2) &= \frac{100}{59 + 600\mathrm{e} - 180\mathrm{e}^2} &&\approx 0.277825, \text{ and} \\ \rho(\{1,3\}, f_2) &= \frac{8}{177 + 1800\mathrm{e} - 540\mathrm{e}^2} &&\approx 0.007409. \end{aligned}$$