Hierarchical tensor approximation of high-dimensional functions of isotropic and anisotropic Sobolev smoothness

Emily Gajendran, Helmut Harbrecht, Remo von Rickenbach

Departement Mathematik und Informatik
Fachbereich Mathematik
Universität Basel
CH-4051 Basel

Preprint No. 2024-02
April 2024

dmi.unibas.ch

HIERARCHICAL TENSOR APPROXIMATION OF HIGH-DIMENSIONAL FUNCTIONS OF ISOTROPIC AND ANISOTROPIC SOBOLEV SMOOTHNESS

EMILY GAJENDRAN, HELMUT HARBRECHT, AND REMO VON RICKENBACH

Abstract. In this article, we study the hierarchical tensor decomposition of functions from both smoothness classes, isotropic Sobolev spaces and anisotropic Sobolev spaces. For this purpose, we consider the known rank estimates in case of bivariate approximation, which can be found in [14] and [16], and successively apply them to analyze the truncated hierarchical tensor decomposition. In comparison to the isotropic case, we obtain improved results with respect to anisotropic Sobolev spaces. Indeed, the associated ranks of the truncated hierarchical tensor decomposition stay essentially bounded which beats the curse of dimension that is observed in the isotropic case.

1. Introduction

In the realm of numerical analysis and scientific computing, the challenge of effectively representing and approximating functions in high dimensions has remained an important task. In this article, we therefore examine the approximation of functions on the d-fold product Ω^d of a bounded domain $\Omega \subset \mathbb{R}^n$ with itself, where $n \in \mathbb{N}$ and d is a power of 2. Specifically, we focus on the *low-rank approximation* of functions from isotropic and anisotropic Sobolev spaces $H^s(\Omega^d)$ and $H^s_{\text{mix}}(\Omega^d) := \bigotimes_{i=1}^d H^s(\Omega)$, respectively, achieved by means of tensor approximation.

Various tensor approximation schemes have been developed over the years such as tensor trains, the canonical tensor format, or hierarchical tensor decomposition, see e.g. [17, 18, 19] and the references therein. Note that a thorough introduction to tensor methods can be found in the textbook [17], while a comprehensive overview of current methods and their associated literature is available in [12].

A lot is known about the special situation when $d = 2$, which revolves around the use of the truncated singular value decomposition as found in [13, 14]. The case of higher dimensions, however, is still not well understood and has many open questions. For this general case, one shall study different tensor decompositions and corresponding truncations, searching for a balance between minimal error and minimal cost in the calculations. A common issue one encounters in this setting is the curse of dimension, which one observes in the computational cost of the approximations one constructs.

Date: April 21, 2024.

2020 *Mathematics Subject Classification.* Primary 41A46; Secondary 41A63, 46E35.

Key words and phrases. Low-rank approximation, hierarchical tensor decomposition, cost complexity.

This research has been supported by the SNSF Grant “Low-rank tensor approximation of high-dimensional functions” under grant number 200021_219782.

Limited progress has been made to combat this curse of dimension by low-rank tensor approximation so far, though in different areas of sciences tensor approximations have successfully been applied to high-dimensional problems in fields such as quantum mechanics and physics. However, it has been found in [16, 20] that the curse of dimension does not impact the computational cost as much, if functions from anisotropic Sobolev spaces are considered. The reason for this is that the cost for the truncated singular value decomposition of functions in such spaces is essentially only determined by the dimension $n \in \mathbb{N}$ of the domain $\Omega \subset \mathbb{R}^n$ under consideration. Here and in the following, *essentially* in the context of complexity bounds and error estimates means up to (poly-)logarithmic factors.

The main way we study tensor approximation in this article is by comparing tensor-ranks, see [9, 10, 17] for example. As the cost complexity is fully dependent on the tensor-ranks, we are especially interested in the anistropic Sobolev space due to the known estimate on the truncated singular value decomposition being essentially independent of the dimension d and only dependent on n, see [16, 21] for example. While the approximation of high-dimensional functions in the tensor train format has been studied in [15, 16], we focus here on the *hierarchical tensor decomposition*. Its construction by the higher order SVD is well understood, see, e.g. [11] or [20]. However, up to this point the tensor-ranks have not been studied in detail. Therefore, the content of this article is based on a combination of estimates of the singular value decomposition and the construction of the hierarchical tensor decomposition.

We like to emphasize that our results are in line with [20] for periodic functions on the unit d-cube, but generalize the findings therein. Moreover, the convergence of tensor approximation methods has also been studied by other authors. We refer the reader to [2, 3, 4, 5, 22] and the references therein. In contrast to our setting, the sparsity of the core tensors is considered in these articles, especially also for general networks. To this end, functions from Besov spaces instead of Sobolev spaces have been the objects of study.

The rest of the article is structured as follows. In Section 2, we consider notation and basic results which we will use later on. Then, in Section 3, we introduce the truncated hierarchical tensor decomposition for functions in the continuous setting. Section 4 is then dedicated to the error analysis of the truncated hierarchical tensor decomposition. The specific consequences for the ranks in case of isotropic or anisotropic smoothness are then considered in Section 5. Finally, we state concluding remarks in Section 6, while we prove a technical estimate in Appendix A.

Throughout this article, in order to avoid the repeated use of generic but unspecified constants, by $C \lesssim D$ we indicate that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Moreover, $C \gtrsim D$ is defined as $D \lesssim C$ and $C \sim D$ as $C \lesssim D$ and $D \lesssim C$.

2. Preliminaries

2.1. **Notation.** For some $n \in \mathbb{N}$, let $\Omega \subset \mathbb{R}^n$ be a sufficiently smooth domain. For $m \in \mathbb{N}$, we denote by $\mathbf{\Omega}_m$ the 2^m-fold product domain of Ω, i.e.

$$\mathbf{\Omega}_m = \underbrace{\Omega \times \cdots \times \Omega}_{2^m} = \Omega^{2^m}.$$

Furthermore, for sake of convenience, we set $d = 2^m$. The reason why we choose these values is because we shall consider the hierarchical tensor decomposition which is based on an arrangement of the product domain $\mathbf{\Omega}_m$ in form of a binary tree, where a particular vertex (ℓ, k) corresponds to a subdomain $\mathbf{\Omega}_{m-\ell}$. Here, $\ell = 1, \ldots, m$ denotes the level and $k = 1, \ldots, 2^\ell$ denotes the vertex in the respective level, counting from left to right, compare Figure 1 with nodes (ℓ, k).

We denote by

$$\boldsymbol{x}_{\ell,k} = (x_{2^{m-\ell}(k-1)+1}, \ldots, x_{2^{m-\ell}k}) \in \mathbf{\Omega}_{m-\ell}$$

the variables which are associated to the vertex (ℓ, k) of the binary tree while

$$\overline{\boldsymbol{x}}_{\ell,k} = (x_1, \ldots, x_{2^{m-\ell}(k-1)}, x_{2^{m-\ell}k+1}, \ldots, x_d) \in \overline{\mathbf{\Omega}}_{m-\ell}$$

denotes the remaining variables which are associated to the respective siblings. Here, for sake of simplicity in notation, we set $\overline{\mathbf{\Omega}}_{m-\ell} := \mathbf{\Omega}_{m-\ell}^{2^\ell-1}$ such that there holds $\mathbf{\Omega}_{m-\ell} \times \overline{\mathbf{\Omega}}_{m-\ell} = \mathbf{\Omega}_m$. Note that arbitrary $d \neq 2^m$ will also be possible with straightforward modifications by means of an unbalanced binary tree.

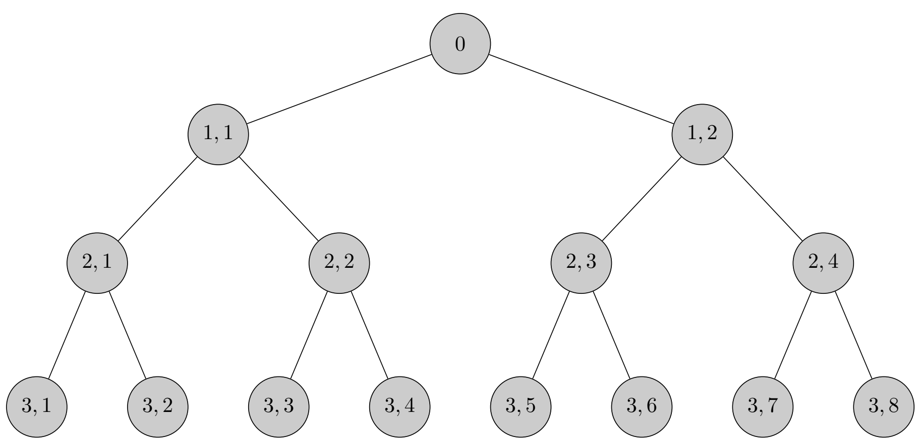

FIGURE 1. A (balanced) binary tree depicting the nodes of the hierarchical tensor format with three levels, i.e. for a function with 8 variables.

2.2. **Sobolev spaces.** Let $L^2(\mathbf{\Omega}_m)$ denote the space of squared integrable functions. Then, the canonical, standard (isotropic) Sobolev space $H^p(\mathbf{\Omega}_m)$ consists of all functions $f \in L^2(\mathbf{\Omega}_m)$ whose partial derivatives $\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} f$ of order $\|\boldsymbol{\alpha}\|_1 = \alpha_1 + \cdots + \alpha_d \leq p$ have finite L^2-norms. In contrast, the (anisotropic) Sobolev space $H^p_{\text{mix}}(\mathbf{\Omega}_m)$ of dominating mixed derivatives consists of all functions $f \in L^2(\mathbf{\Omega}_m)$ whose partial derivatives $\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} f$ of order $\|\boldsymbol{\alpha}\|_\infty = \max\{\alpha_1, \ldots, \alpha_d\} \leq p$ have finite L^2-norms. Especially, one has

$$H^p_{\text{mix}}(\mathbf{\Omega}_m) = \underbrace{H^p(\Omega) \otimes \cdots \otimes H^p(\Omega)}_{2^m}. \tag{2.1}$$

2.3. **Singular value decomposition.** We intend to compute low-rank approximations of functions $f \in L^2(\boldsymbol{\Omega}_m)$ by means of the singular value decomposition. In order to separate the variables $\boldsymbol{x}_{\ell,k} \in \boldsymbol{\Omega}_{m-\ell}$, associated with the vertex (ℓ,k), from $\overline{\boldsymbol{x}}_{\ell,k} \in \overline{\boldsymbol{\Omega}}_{m-\ell}$, belonging to the same generation $(\ell,1),\ldots,(\ell,k-1),(\ell,k+1),\ldots,(\ell,2^\ell)$, we define the kernel

$$k(\boldsymbol{x}_{\ell,k},\boldsymbol{x}'_{\ell,k}) := \int_{\overline{\boldsymbol{\Omega}}_{m-\ell}} f(\boldsymbol{x}_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k}) f(\boldsymbol{x}'_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k}) \,\mathrm{d}\overline{\boldsymbol{x}}_{\ell,k}$$

and compute the eigenpairs $\{(\lambda_{\ell,k}(\alpha_{\ell,k}),\varphi_{\ell,k}(\alpha_{\ell,k})\}_{\alpha_{\ell,k}=1}^\infty$ of the associated Hilbert-Schmidt operator

$$\begin{aligned}\lambda_{\ell,k}(\alpha_{\ell,k})\varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k}) &= \mathcal{K}\varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k}) \\ &= \int_{\boldsymbol{\Omega}_{m-\ell}} k(\boldsymbol{x}_{\ell,k},\boldsymbol{x}'_{\ell,k})\varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}'_{\ell,k}) \,\mathrm{d}\boldsymbol{x}'_{\ell,k}.\end{aligned}$$

Here, the sequence of eigenfunctions $\{\varphi_{\ell,k}(\alpha_{\ell,k})\}_{\alpha_{\ell,k}=1}^\infty$ constitutes an orthonormal basis in $L^2(\boldsymbol{\Omega}_{m-\ell})$ while the sequence of eigenvalues satisfies $\lambda_{\ell,k}(1) \geq \lambda_{\ell,k}(2) \geq \cdots \geq \lambda_{\ell,k}(\alpha_{\ell,k}) \to 0$. Setting

$$\psi_{\ell,k}(\overline{\boldsymbol{x}}_{\ell,k}) := \frac{1}{\sqrt{\lambda_{\ell,k}(\alpha_{\ell,k})}} \int_{\boldsymbol{\Omega}_{m-\ell}} f(\boldsymbol{x}_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k})\varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k}) \,\mathrm{d}\boldsymbol{x}_{\ell,k}, \tag{2.2}$$

the singular value decomposition of f is given by

$$f(\boldsymbol{x}_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k}) := \sum_{\alpha_{\ell,k}=1}^{\infty} \sqrt{\lambda(\alpha_{\ell,k})}\varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k})\psi_{\ell,k}(\alpha_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k}). \tag{2.3}$$

2.4. **Truncation of the singular value decomposition.** In order to compute a low-rank approximation of $f \in L^2(\boldsymbol{\Omega}_m)$, we shall truncate the singular value decomposition (2.3) appropriately. To this end, we recall the bounds on the truncation error proven [14, 16].

If $f \in H^p(\boldsymbol{\Omega}_m)$, then the truncated singular value decomposition

$$f_R(\boldsymbol{x}_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k}) := \sum_{\alpha_{\ell,k}=1}^{R} \sqrt{\lambda(\alpha_{\ell,k})}\varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k})\psi_{\ell,k}(\alpha_{\ell,k},\overline{\boldsymbol{x}}_{\ell,k})$$

satisfies the error estimate

$$\|f-f_R\|_{L^2(\boldsymbol{\Omega}_m)} = \sqrt{\sum_{\alpha_{\ell,k}=R+1}^{\infty} \lambda(\alpha_{\ell,k})} \lesssim R^{-2^{\ell-m}p/n}\|f\|_{H^p(\boldsymbol{\Omega}_m)} \tag{2.4}$$

if $f \in H^p(\boldsymbol{\Omega}_m)$, see [14] for the details. In contrast, in accordance with [16], we essentially obtain the truncation estimate

$$\|f-f_R\|_{L^2(\boldsymbol{\Omega}_m)} = \sqrt{\sum_{\alpha_{\ell,k}=R+1}^{\infty} \lambda(\alpha_{\ell,k})} \lesssim R^{-2p/n}\|f\|_{H^p_{\mathrm{mix}}(\boldsymbol{\Omega}_m)} \tag{2.5}$$

if $f \in H^p_{\mathrm{mix}}(\boldsymbol{\Omega}_m)$.

We emphasize that the decay in (2.4) suffers from the so-called curse of dimensionality while the decay in (2.5) is essentially dimension independent. Note that here and in the following "essentially" in the context of asymptotic estimates means up to (poly-)logarithmic factors.

2.5. **Projections.** Based on the truncated singular value decomposition (2.3) with respect to a given vertex (ℓ, k) of the binary tree under consideration, we shall define L^2-orthogonal projections $P_{\ell,k}^R : L^2(\boldsymbol{\Omega}_m) \to L^2(\boldsymbol{\Omega}_m)$ which will play an important role in the error analysis of the truncated hierarchical tensor decomposition. We define the projection $P_{\ell,k}^R$ by

$$P_{\ell,k}^R f(\boldsymbol{x}) := \sum_{\alpha_{\ell,k}=1}^{R} \int_{\boldsymbol{\Omega}_{m-\ell}} f(\boldsymbol{x}'_{\ell,k}, \overline{\boldsymbol{x}}_{\ell,k}) \varphi_{\ell,k}(\alpha_{\ell,k}, \boldsymbol{x}'_{\ell,k}) \, \mathrm{d}\boldsymbol{x}'_{\ell,k} \; \varphi_{\ell,k}(\alpha_{\ell,k}, \boldsymbol{x}_{\ell,k}). \tag{2.6}$$

In particular, in view of (2.2), there holds the identity

$$P_{\ell,k}^R f(\boldsymbol{x}) = \sum_{\alpha_{\ell,k}=1}^{R} \sqrt{\lambda(\alpha_{\ell,k})} \varphi_{\ell,k}(\alpha_{\ell,k}, \boldsymbol{x}_{\ell,k}) \psi_{\ell,k}(\alpha_{\ell,k}, \overline{\boldsymbol{x}}_{\ell,k}). \tag{2.7}$$

Hence, due to (2.4) and due to (2.5), we arrive at the error estimates

$$\|(I - P_{\ell,k}^R) f\|_{L^2(\boldsymbol{\Omega}_m)} \lesssim R^{-2^{\ell-m} p/n} \|f\|_{H^p(\boldsymbol{\Omega}_m)}, \quad \text{if } f \in H^p(\boldsymbol{\Omega}_m), \tag{2.8}$$

and (essentially)

$$\|(I - P_{\ell,k}^R) f\|_{L^2(\boldsymbol{\Omega}_m)} \lesssim R^{-2p/n} \|f\|_{H^p_{\text{mix}}(\boldsymbol{\Omega}_m)}, \quad \text{if } f \in H^p_{\text{mix}}(\boldsymbol{\Omega}_m), \tag{2.9}$$

respectively. It is important to note that the generic constants in these error estimates are dependent on m and p but independent of the rank R. In particular, estimate (2.8) depends strongly on the specific vertex (ℓ, k) of the binary tree, while estimate (2.9) depends only mildly on it as the rank decay depends only logarithmically on the dimension $2^{m-\ell}$ of the underlying domain $\boldsymbol{\Omega}_{m-\ell}$.

3. Hierarchical tensor decomposition

Consider a function $f(\boldsymbol{x}) = f(x_1, \ldots, x_d)$ with $d = 2^m$ for some $m \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^n$, where $n \in \mathbb{N}$ is arbitrary but fixed. In the hierarchical tensor decomposition, we apply the singular value decomposition to the function f to successively separate the desired variables from the rest. This leaves us with a tree-like pattern, where every branch provides a set of orthonormal functions for the hierarchical tensor decomposition.

Let $f \in L^2(\boldsymbol{\Omega}_m)$. We first consider the root vertex. Applying the singular value decomposition (2.3) to separate the variables $\boldsymbol{x}_{1,1} \in \boldsymbol{\Omega}_{m-1}$ and $\boldsymbol{x}_{1,2} \in \boldsymbol{\Omega}_{m-1}$ gives us the representation

$$f(\boldsymbol{x}) = \sum_{\alpha_1=1}^{\infty} \sqrt{\lambda_1(\alpha_1)} \varphi_{1,1}(\alpha_1, \boldsymbol{x}_{1,1}) \varphi_{1,2}(\alpha_1, \boldsymbol{x}_{1,2}),$$

where we set $\varphi_{1,2}(\alpha_1) := \psi_{1,2}(\alpha_1)$ for all $\alpha_1 \in \mathbb{N}$.

We next consider both children of the root vertex and apply the singular value decomposition directly to f again to separate the variable $\boldsymbol{x}_{2,k} \in \boldsymbol{\Omega}_{m-2}$ for each $k = 1, 2, 3, 4$ from the rest. We get the identity

$$f(\boldsymbol{x}) = \sum_{\alpha_{2,k}=1}^{\infty} \sqrt{\lambda_{2,k}(\alpha_{2,k})} \varphi_{2,k}(\alpha_{2,k}, \boldsymbol{x}_{2,k}) \psi_{2,k}(\alpha_{2,k}, \overline{\boldsymbol{x}}_{2,k}), \quad k = 1, 2, 3, 4.$$

The functions $\{\varphi_{2,k}(\alpha_{2,k})\}_{\alpha_{2,k}=1}^{\infty}$ constitute orthonormal bases in $L^2(\boldsymbol{\Omega}_{m-2})$ for each k. Thus, defining the core tensor in vertex $(1,1)$ by

$$\beta_{1,1}(\alpha_1,\alpha_{2,1},\alpha_{2,2}) := \int_{\boldsymbol{\Omega}_{m-1}} \varphi_{1,1}(\alpha_1,\boldsymbol{x}_{1,1})\varphi_{2,1}(\alpha_{2,1},\boldsymbol{x}_{2,1})\varphi_{2,2}(\alpha_{2,2},\boldsymbol{x}_{2,2})\,\mathrm{d}\boldsymbol{x}_{1,1}$$

and the core tensor in vertex $(1,2)$ by

$$\beta_{1,2}(\alpha_1,\alpha_{2,3},\alpha_{2,4}) := \int_{\boldsymbol{\Omega}_{m-1}} \varphi_{1,2}(\alpha_1,\boldsymbol{x}_{1,2})\varphi_{2,3}(\alpha_{2,3},\boldsymbol{x}_{2,3})\varphi_{2,4}(\alpha_{2,4},\boldsymbol{x}_{2,4})\,\mathrm{d}\boldsymbol{x}_{1,2}$$

gives us the representation

$$\begin{aligned} f(\boldsymbol{x}) = \sum_{\alpha_1=1}^{\infty} & \sqrt{\lambda_1(\alpha_1)} \\ & \cdot \left[\sum_{\alpha_{2,1}=1}^{\infty}\sum_{\alpha_{2,2}=1}^{\infty} \beta_{1,1}(\alpha_1,\alpha_{2,1},\alpha_{2,2})\varphi_{2,1}(\alpha_{2,1},\boldsymbol{x}_{2,1})\varphi_{2,2}(\alpha_{2,2},\boldsymbol{x}_{2,2})\right] \\ & \cdot \left[\sum_{\alpha_{2,3}=1}^{\infty}\sum_{\alpha_{2,4}=1}^{\infty} \beta_{1,2}(\alpha_1,\alpha_{2,3},\alpha_{2,4})\varphi_{2,3}(\alpha_{2,3},\boldsymbol{x}_{2,3})\varphi_{2,4}(\alpha_{2,4},\boldsymbol{x}_{2,4})\right]. \end{aligned}$$

In general, we employ the singular value decomposition of f in the vertex (ℓ,k) to separate the variable $\boldsymbol{x}_{\ell+1,2k-1} \in \boldsymbol{\Omega}_{m-\ell-1}$ from $\overline{\boldsymbol{x}}_{\ell+1,2k-1}$ and $\boldsymbol{x}_{\ell+1,2k} \in \boldsymbol{\Omega}_{m-\ell-1}$ from $\overline{\boldsymbol{x}}_{\ell+1,2k}$, respectively. This yields two sets of orthonormal bases

$$\{\varphi_{\ell+1,2k-1}(\alpha_{\ell+1,2k-1})\}_{\alpha_{\ell+1,2k-1}=1}^{\infty}, \quad \{\varphi_{\ell+1,2k}(\alpha_{\ell+1,2k})\}_{\alpha_{\ell+1,2k}=1}^{\infty}$$

of $L^2(\boldsymbol{\Omega}_{m-\ell-1})$. With the help of these bases, we expand the basis functions in the vertex (ℓ,k)

$$\begin{aligned} \varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k}) = \sum_{\alpha_{\ell+1,2k-1}=1}^{\infty}\sum_{\alpha_{\ell+1,2k}=1}^{\infty} & \beta_{\ell,k}(\alpha_{\ell,k},\alpha_{\ell+1,2k-1},\alpha_{\ell+1,2k}) \\ & \cdot \varphi_{\ell+1,2k-1}(\alpha_{\ell+1,2k-1},\boldsymbol{x}_{\ell+1,2k-1})\varphi_{\ell+1,2k}(\alpha_{\ell+1,2k},\boldsymbol{x}_{\ell+1,2k}), \end{aligned}$$

where the core tensor is given by

$$\begin{aligned} \beta_{\ell,k}(\alpha_{\ell,k},\alpha_{\ell+1,2k-1},\alpha_{\ell+1,2k}) = \int_{\boldsymbol{\Omega}_{m-\ell}} & \varphi_{\ell,k}(\alpha_{\ell,k},\boldsymbol{x}_{\ell,k}) \\ \cdot \varphi_{\ell+1,2k-1}(\alpha_{\ell+1,2k-1},\boldsymbol{x}_{\ell+1,2k-1}) & \varphi_{\ell+1,2k}(\alpha_{\ell+1,2k},\boldsymbol{x}_{\ell+1,2k})\,\mathrm{d}\boldsymbol{x}_{\ell,k}. \end{aligned}$$

Proceeding successively in this way in all vertices, which are different from the leaves of the binary tree, gives us the final hierarchical tensor decomposition in accordance with

$$\begin{aligned} f(\boldsymbol{x}) = \sum_{\alpha_1=1}^{\infty} \sqrt{\lambda_1(\alpha_1)} \sum_{\alpha_{2,1}=1}^{\infty} \cdots \sum_{\alpha_{2,4}=1}^{\infty} \beta_{1,1}(\alpha_1,\alpha_{2,1},\alpha_{2,2})\beta_{1,2}(\alpha_1,\alpha_{2,3},\alpha_{2,4}) \cdots \\ \sum_{\alpha_{m,1}=1}^{\infty} \cdots \sum_{\alpha_{m,2^m}=1}^{\infty} \beta_{m-1,1}(\alpha_{m-1,1},\alpha_{m,1},\alpha_{m,2}) \cdots \\ \beta_{m-1,2^{m-1}}(\alpha_{m-1,2^{m-1}},\alpha_{m,2^m-1},\alpha_{m,2^m})\varphi_{m,1}(\alpha_{m,1},\boldsymbol{x}_1)\cdots\varphi_{m,2^m}(\alpha_{m,2^m},\boldsymbol{x}_{2^m}). \end{aligned}$$

Of course, on a computer, we have to truncate the singular value decomposition in each vertex (ℓ, k), resulting in a finite rank R_ℓ. Therefore, we arrive at the low-rank approximation $f^{HT}_{R_1,\ldots,R_m} \approx f$ given by

$$f^{HT}_{R_1,\ldots,R_m}(\boldsymbol{x}) = \sum_{\alpha_1=1}^{R_1} \sqrt{\lambda_1(\alpha_1)} \sum_{\alpha_{2,1}=1}^{R_2} \cdots \sum_{\alpha_{2,4}=1}^{R_2} \beta_{1,1}(\alpha_1, \alpha_{2,1}, \alpha_{2,2}) \beta_{1,2}(\alpha_1, \alpha_{2,3}, \alpha_{2,4}) \cdots$$
$$\sum_{\alpha_{m,1}=1}^{R_m} \cdots \sum_{\alpha_{m,2^m}=1}^{R_m} \beta_{m-1,1}(\alpha_{m-1,1}, \alpha_{m,1}, \alpha_{m,2}) \cdots$$
$$\beta_{m-1,2^{m-1}}(\alpha_{m-1,2^{m-1}}, \alpha_{m,2^m-1}, \alpha_{m,2^m}) \varphi_{m,1}(\alpha_{m,1}, \boldsymbol{x}_1) \cdots \varphi_{m,2^m}(\alpha_{m,2^m}, \boldsymbol{x}_{2^m}).$$

Thus, in each vertex (ℓ, k) being different from a leaf, the truncated core tensor $\{\beta_{\ell,k}(\alpha_{\ell,k}, \alpha_{\ell+1,2k-1}, \alpha_{\ell+1,2k})\}$ needs to be stored. It is of size $R_\ell R^2_{\ell+1}$ except for the root vertex where we only have to store the R_1 singular values $\{\sqrt{\lambda_1}\}$. In addition, we also need to store the $d = 2^m$ basis sets $\{\varphi_{m,k}(\alpha_{m,k})\}_{\alpha_{m,k}=1}^{R_m}$.

4. Error analysis

The error analysis of the truncated hierarchical tensor decomposition follows essentially [11], but we translate the results therein to the continuous setting considered here. The analysis makes heavy use of the L^2-orthogonal projections $P^{R_\ell}_{\ell,k}$ from (2.6).

Lemma 4.1. *The truncated hierarchical tensor decomposition can be expressed by*

$$f^{HT}_{R_1,\ldots,R_m} = \left(\prod_{\ell=1}^{m} \prod_{k=1}^{2^\ell} P^{R_\ell}_{\ell,k} \right) f, \tag{4.1}$$

where the L^2-orthogonal projections $P^{R_\ell}_{\ell,k} : L^2(\boldsymbol{\Omega}_m) \to L^2(\boldsymbol{\Omega}_m)$ are given by (2.6).

Proof. We start by showing that the assertion holds true for the first vertices of the binary tree. We especially find

$$P_{1,1} f = P_{1,2} f = \sum_{\alpha_1=1}^{R_1} \sqrt{\lambda_1(\alpha_1)} \varphi_{1,1}(\alpha_1) \varphi_{1,2}(\alpha_1).$$

We proceed by continuing on level $\ell = 2$ of the tree and obtain

$$P_{2,1} P_{1,2} P_{1,1} f = \sum_{\alpha_1=1}^{R_1} \sqrt{\lambda_1(\alpha_1)}$$
$$\cdot \sum_{\alpha_{2,1}=1}^{R_2} \left[\int_{\boldsymbol{\Omega}_{m-2}} \varphi_{1,1}(\alpha_{1,2}, \boldsymbol{x}_{2,1}, \cdot) \varphi_{2,1}(\alpha_{2,1}, \boldsymbol{x}_{2,1}) \,\mathrm{d}\boldsymbol{x}_{2,1} \right] \varphi_{2,1}(\alpha_{2,1}) \varphi_{1,2}(\alpha_1)$$

and further

$$P_{2,2}P_{2,1}P_{1,2}P_{1,1}f = \sum_{\alpha_1=1}^{R_1} \sqrt{\lambda_1(\alpha_1)}$$
$$\cdot \sum_{\alpha_{2,1}=1}^{R_2} \sum_{\alpha_{2,2}=1}^{R_2} \left[\int_{\mathbf{\Omega}_{m-2}} \varphi_{1,1}(\alpha_{1,2}, \boldsymbol{x}_{2,1}, \boldsymbol{x}_{2,2}) \varphi_{2,1}(\alpha_{2,1}, \boldsymbol{x}_{2,1}) \varphi_{2,2}(\alpha_{2,2}, \boldsymbol{x}_{2,2}) \,\mathrm{d}\boldsymbol{x}_{2,1}\,\mathrm{d}\boldsymbol{x}_{2,2} \right]$$
$$\cdot \varphi_{2,1}(\alpha_{2,1})\varphi_{2,2}(\alpha_{2,2})\varphi_{1,2}(\alpha_1)$$
$$= \sum_{\alpha_1=1}^{R_1} \sqrt{\lambda_1(\alpha_1)} \sum_{\alpha_{2,1}=1}^{R_2} \sum_{\alpha_{2,2}=1}^{R_2} \beta_{1,1}(\alpha_1, \alpha_{2,1}, \alpha_{2,2}) \varphi_{2,1}(\alpha_{2,1}) \varphi_{2,2}(\alpha_{2,2}) \varphi_{1,2}(\alpha_1).$$

Likewise, we obtain

$$P_{2,4}P_{2,3}P_{2,2}P_{2,1}P_{1,2}P_{1,1}f = \sum_{\alpha_1=1}^{R_1} \sqrt{\lambda_1(\alpha_1)}$$
$$\sum_{\alpha_{2,1}=1}^{R_2} \cdots \sum_{\alpha_{2,4}=1}^{R_2} \beta_{1,1}(\alpha_1, \alpha_{2,1}, \alpha_{2,2}) \beta_{1,2}(\alpha_1, \alpha_{2,1}, \alpha_{2,2}) \varphi_{2,1}(\alpha_{2,1}) \cdots \varphi_{2,4}(\alpha_{2,4}).$$

In the general step, the basis set $\{\varphi_{\ell,k}(\alpha_{\ell,k})\}_{\alpha_{\ell,k}=1}^{R_\ell}$ with respect to the vertex (ℓ, k) is replaced by the application of the projections $P_{\ell+1,2k-1}P_{\ell+1,2k}$ in accordance with

$$\varphi_{\ell,k}(\alpha_{\ell,k}, \boldsymbol{x}_{\ell,k}) = \sum_{\alpha_{\ell+1,2k-1}=1}^{R_{\ell+1}} \sum_{\alpha_{\ell+1,2k}=1}^{R_{\ell+1}} \beta_{\ell,k}(\alpha_{\ell,k}, \alpha_{\ell+1,2k-1}, \alpha_{\ell+1,2k})$$
$$\cdot \varphi_{\ell+1,2k-1}(\alpha_{\ell+1,2k-1}, \boldsymbol{x}_{\ell+1,2k-1}) \varphi_{\ell+1,2k}(\alpha_{\ell+1,2k}, \boldsymbol{x}_{\ell+1,2k}).$$

By proceeding successively through the binary tree, one obtains the representation formula (4.1). □

The next lemma is the key ingredient to estimate the truncation error of the hierarchical tensor decomposition.

Lemma 4.2. *Let $f \in L^2(\mathbf{\Omega}_m)$ and let $P : L^2(\mathbf{\Omega}_m) \to L^2(\mathbf{\Omega}_m)$ and $Q : L^2(\mathbf{\Omega}_m) \to L^2(\mathbf{\Omega}_m)$ be arbitrary L^2-orthogonal projections. Then, there holds*

$$\|(I - PQ)f\|_{L^2(\mathbf{\Omega}_m)}^2 \leq \|(I - P)f\|_{L^2(\mathbf{\Omega}_m)}^2 + \|(I - Q)f\|_{L^2(\mathbf{\Omega}_m)}^2. \tag{4.2}$$

Proof. We have

$$\|(I - PQ)f\|_{L^2(\mathbf{\Omega}_m)}^2 = \|(I - P)f + P(I - Q)f\|_{L^2(\mathbf{\Omega}_m)}^2.$$

Due to the orthogonality of P and $I - P$, the claim follows by Pythagoras' theorem in accordance with

$$\begin{aligned}\|(I - PQ)f\|_{L^2(\mathbf{\Omega}_m)}^2 &= \|(I - P)f\|_{L^2(\mathbf{\Omega}_m)}^2 + \|P(I - Q)f\|_{L^2(\mathbf{\Omega}_m)}^2 \\ &\leq \|(I - P)f\|_{L^2(\mathbf{\Omega}_m)}^2 + \|(I - Q)f\|_{L^2(\mathbf{\Omega}_m)}^2.\end{aligned}$$

□

We shall use the above lemma and the representation formula (4.1) to find an upper bound of the over-all truncation error of the truncated hierarchical tensor decomposition.

Theorem 4.3. *There holds*

$$\big\|f - f^{HT}_{R_1,\ldots,R_m}\big\|^2_{L^2(\mathbf{\Omega}_m)} \le \sum_{\ell=1}^{m}\sum_{k=1}^{2^\ell} \big\|\big(I - P^{R_\ell}_{\ell,k}\big)f\big\|^2_{L^2(\mathbf{\Omega}_m)}.$$

Proof. In view of (4.1), successive application of (4.2) yields

$$\begin{aligned}
\left\|\left(I - \prod_{\ell=1}^{m}\prod_{k=1}^{2^\ell} P^{R_\ell}_{\ell,k}\right)f\right\|^2_{L^2(\mathbf{\Omega}_m)} &= \left\|\left(I - P^{R_m}_{m,2^m}\cdots P^{R_m}_{m,1}\prod_{\ell=1}^{m-1}\prod_{k=1}^{2^\ell} P^{R_\ell}_{\ell,k}\right)f\right\|^2_{L^2(\mathbf{\Omega}_m)}\\
&\le \big\|\big(I - P^{R_m}_{m,2^m}\big)f\big\|^2_{L^2(\mathbf{\Omega}_m)} + \left\|\left(I - P^{R_m}_{m,2^m-1}\cdots P^{R_m}_{m,1}\prod_{\ell=1}^{m-1}\prod_{k=1}^{2^\ell} P^{R_\ell}_{\ell,k}\right)f\right\|^2_{L^2(\mathbf{\Omega}_m)}\\
&\le \cdots \le \sum_{k=1}^{2^m}\big\|\big(I - P^{R_m}_{m,k}\big)f\big\|^2_{L^2(\mathbf{\Omega}_m)} + \left\|\left(I - \prod_{\ell=1}^{m-1}\prod_{k=1}^{2^\ell} P^{R_\ell}_{\ell,k}\right)f\right\|^2_{L^2(\mathbf{\Omega}_m)}.
\end{aligned}$$

By continuing this procedure also on the other levels, we conclude the claim. □

Given that the truncation ranks R_ℓ are chosen such that we always have

$$\big\|\big(I - P^{R_\ell}_{\ell,k}\big)f\big\|_{L^2(\mathbf{\Omega}_m)} \le \varepsilon$$

for all levels ℓ, we conclude from Theorem 4.3 that

$$\big\|f - f^{HT}_{R_1,\ldots,R_m}\big\|_{L^2(\mathbf{\Omega}_m)} \le \sqrt{2d-3}\,\varepsilon.$$

Here, we exploited that we apply only one projection in the root vertex while two projections are applied in all the other vertices which are not a leaf. This implies that $1 + \sum_{\ell=1}^{m-1} 2\cdot 2^\ell = 2d-3$ truncations are performed in total, each of which of order ε.

5. Cost complexity

5.1. **Ranks in the case of isotropic Sobolev smoothness.** We are now interested in determining the necessary ranks for the truncated hierarchical tensor decomposition $f^{HT}_{R_1,\ldots,R_m}$ for functions $f \in H^p(\mathbf{\Omega}_m)$. To this end, we exploit Theorem 4.3 which shows that all ranks should be chosen such that the truncation error in each singular value decomposition is of order ε. By (2.8), we have that the error in each vertex is given by

$$\big\|\big(I - P^{R_\ell}_{\ell,k}\big)f\big\|_{L^2(\mathbf{\Omega}_m)} \lesssim R_\ell^{-2^{\ell-m}p/n}\|f\|_{H^p(\mathbf{\Omega}_m)}.$$

Therefore, we conclude that we shall choose

$$R_\ell \sim \varepsilon^{-2^{m-\ell}n/p}$$

to ensure an error bound $\lesssim \varepsilon$.

We note that the ranks R_m associated with the leaves of the underlying binary tree are the smallest ones. By setting $R := R_m = \varepsilon^{-n/p}$, we conclude the exponentially increasing ranks illustrated in Figure 2. We like to emphasize that our analysis shows that the truncated hierarchical tensor decomposition $f^{HT}_{R,\ldots,R}$ with fixed rank R cannot be expected to approximate a function $f \in H^p(\mathbf{\Omega}_m)$ well in general.

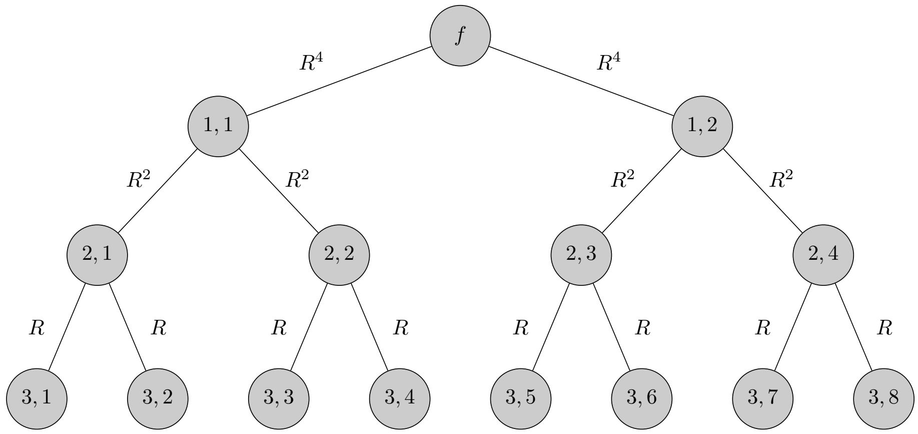

FIGURE 2. Rank requirements in case of a function f of isotropic Sobolev smoothness with $R \sim \varepsilon^{-n/p}$. The ranks grow exponentially with the level ℓ in accordance with $R^{2^{m-\ell}}$.

The total amount of coefficients in the core tensors required to represent $f^{HT}_{R_1,\ldots,R_m}$ is determined by the ranks in the truncated hierarchical tensor decomposition. We obtain

$$R_1 + \sum_{\ell=1}^{m-1} 2^\ell R_\ell R_{\ell+1}^2 = R^{2^{m-1}} + \sum_{\ell=1}^{m-1} 2^\ell R^{2^{m-\ell}} \big(R^{2^{m-\ell-1}}\big)^2$$
$$= R^{2^{m-1}} + \sum_{\ell=1}^{m-1} 2^\ell R^{2^{m-\ell+1}}.$$

Hence, by using the estimate

$$\sum_{\ell=1}^{m-1} 2^\ell R^{2^{m-\ell+1}} \lesssim R^{2^m}, \tag{5.1}$$

which is proven in the appendix, we conclude that the cost complexity for storing the core tensors is bounded by $\mathcal{O}(R^d)$. Note that the Tucker decomposition is of the same complexity, compare [15].

5.2. **Ranks in the case of anisotropic Sobolev smoothness.** We are next interested in studying the truncation ranks if we have $f \in H^p_{\text{mix}}(\mathbf{\Omega}_m)$ for arbitrary $d = 2^m \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^n$. Recall that the term *essentially* means up to powers of the logarithm.

We exploit again Theorem 4.3 and choose the ranks such that the truncation error in each singular value decomposition is of order ε. By (2.9), we have that the error in each vertex is essentially given by

$$\big\|\big(I - P_{\ell,k}^{R_\ell}\big)f\big\|_{L^2(\mathbf{\Omega}_m)} \lesssim R_\ell^{-2p/n} \|f\|_{H^p_{\text{mix}}(\mathbf{\Omega}_m)}.$$

Therefore, we conclude that we shall essentially choose

$$R_\ell \sim \varepsilon^{-2n/p}$$

to ensure an error bound $\lesssim \varepsilon$. We especially observe that the ranks are essentially independent of the level ℓ of the vertex, i.e., $R = \varepsilon^{-2n/p}$ everywhere in the binary tree. This behaviour is also illustrated in Figure 3.

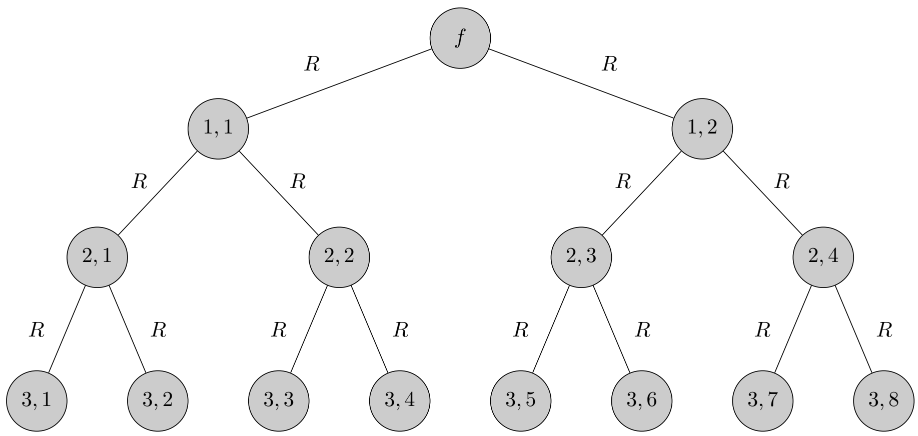

FIGURE 3. Rank requirements in case of a function f of anisotropic Sobolev smoothness with $R \sim \varepsilon^{-2n/p}$. The ranks stay essentially constant with the level ℓ.

The total amount of coefficients required to represent the core tensors of $f^{HT}_{R_1,\ldots,R_m}$ is computed analogously as in the previous subsection by

$$R + \sum_{\ell=1}^{m-1} 2^\ell R R^2 = R + R^3 \sum_{\ell=1}^{m-1} 2^\ell \lesssim dR^3.$$

Hence, the cost complexity for storing the core tensors is essentially of the order $\mathcal{O}(dR^3)$. In particular, it depends only mildly on the spatial dimension d.

5.3. **Approximation of the eigenfunctions.** We shall finally discuss the approximation of the basis sets $\{\varphi_{m,k}(\alpha_{m,k})\}_{\alpha_{m,k}=1}^{R_m}$. To this end, we decompose Ω into an admissible decomposition

$$\mathcal{T}_h := \{\tau_i \subset \mathbb{R}^n : i = 1, \ldots, N_h\}.$$

Here, we denote by h_i, $i = 1, \ldots, N_h$, the local mesh size and $h = \max_i h_i$ is the maximal mesh size.

To a shape-regular sequence $\{\mathcal{T}_h\}_h$, we relate the spatial finite element spaces

$$V_h = \left\{ v_h \in C(\overline{\Omega}) : v_h|_{\overline{\tau}} \in \Pi_r(\overline{\tau}) \text{ for all } \tau \in \mathcal{T}_h \right\}$$

of functions, which are globally continuous, piecewise polynomial functions. Here, $\Pi_r(A)$ denotes the space of polynomials of order r on a set $A \subset \mathbb{R}^n$. For $r = 2$, we obtain the usual piecewise linear hat functions. For $r > 2$, the present setting includes finite elements of higher order and B-splines, compare [6, 7, 8] for details.

We know from [13] that the basis functions $\varphi_{m,k}(\alpha_{m,k})$ are in $H^p(\Omega)$ if $f \in H^p(\mathbf{\Omega}_m)$. Hence, we obtain the error estimate

$$\inf_{v_h \in V_h} \|\varphi_{m,k}(\alpha_{m,k}) - v_h\|_{L^2(\Omega)} \lesssim N_h^{-\min\{p,r\}/n} \|\varphi_{m,k}(\alpha_{m,k})\|_{H^p(\Omega)}.$$

Although the H^p-norm of $\varphi_{m,k}(\alpha_{m,k})$ increases as $\alpha_{m,k}$ increases, it suffices to choose $N_h^{-\min\{p,r\}/n} \sim \varepsilon$. This choice ensures that the numerical realization of the projection $P_{m,k}^{R_m}$ produces an error of order ε, compare [14]. We therefore conclude that, in addition to the storage for the core tensors, we need a storage of $2^m R_m N_h = 2^m R N_h$ with $N_h := \varepsilon^{-n/\min\{p,r\}}$ to represent the required basis sets. We emphasize that these cost are independent of f being in the isotropic Sobolev space $H^p(\mathbf{\Omega}_m)$ or in the anisotropic Sobolev space $H^p_{\text{mix}}(\mathbf{\Omega}_m)$.

6. Conclusion

In this article, we explored the hierarchical tensor decomposition applied to functions from isotropic or anisotropic Sobolev spaces. By considering anisotropic Sobolev spaces alongside the typical isotropic Sobolev spaces, we could exploit the (essential) dimension independence of the ranks of the singular value decomposition in the multivariate case. With this, we could show that the curse of dimension is not present when applying the truncated hierarchical tensor decomposition to functions from anisotropic Sobolev spaces.

In contrast, we have proven that the curse of dimension is present when applying the truncated hierarchical tensor decomposition to functions from isotopic Sobolev spaces. This leads us to the conclusion that the required ranks to achieve a desired target accuracy for the truncated hierarchical tensor decomposition are significantly smaller when applied to functions from anisotropic Sobolev space, as opposed to the ranks necessary to achieve the same accuracy when considering functions from isotropic Sobolev spaces. In fact, we have seen that, in the isotropic case, the ranks grow exponentially as we move up in the binary tree, leading to infeasibly large ranks as we increase the dimension of the functions we wish to approximate.

Finally, we note that in this article we have restricted ourselves to the study of the ranks of the hierarchical tensor decomposition in the continuous setting. In addition, we have only looked at the case in which the hierarchical tree is in the form of a binary tree. Of course, should the scheme be performed, for example, for functions in an isotropic or anisotropic Sobolev space with a general product domain $\Omega_1 \times \cdots \times \Omega_d$, where each subdomain $\Omega_i \in \mathbb{R}^{n_i}$ might be different including different dimensions, the corresponding decomposition would have to be performed with the aim of splitting the variables into two sets of similar dimension in every step. This could lead to a tree with shorter and longer branches, with leaves, i.e. vertices containing the final desired functions, being on different levels of the tree. This would further complicate the construction of the approximation and, especially in the isotropic case, the determination of the resulting ranks.

Appendix A. Asymptic extimate

In this appendix, we shall prove the estimate (5.1). To this end, let us set $a := R^{2^{m+1}}$. We start to remark that the function $\ell \mapsto 2^\ell a^{2^{-\ell}}$ is monotonically decreasing for $\ell = 1, \ldots, m-1$. Indeed, if $g(x) := x a^{\frac{1}{x}}$, then we have $2^\ell a^{2^{-\ell}} = g(2^\ell)$. Since

$$g'(x) = a^{\frac{1}{x}} - \frac{1}{x} a^{\frac{1}{x}} \log a = a^{\frac{1}{x}} \left(1 - \frac{\log a}{x}\right),$$

we have $g'(x) \leq 0$ whenever $x \leq \log a = 2^{m+1} \log R$. Under the assumption that $R \geq 2$, this is the case whenever $x \leq 2^m$. Therefore, there holds

$$\sum_{\ell=1}^{m-1} 2^\ell a^{2^{-\ell}} = 2\sqrt{a} + 4\sqrt[4]{a} + \sum_{\ell=3}^{m-1} 2^\ell a^{2^{-\ell}} \leq 2R^{2^m} + 4R^{2^{m-1}} + \int_2^{m-1} 2^x a^{2^{-x}} \, \mathrm{d}x.$$

Thus, as $2R^{2^m} + 4R^{2^{m-1}} \lesssim R^d$, we just need to asymptotically bound the integral by R^d.

First, by the coordinate transform $x \mapsto \phi(x) := \log_2 x$, using that $\phi'(x) = \frac{1}{x \log 2}$, we obtain

$$\begin{aligned}
\int_2^{m-1} 2^x a^{2^{-x}} \, \mathrm{d}x &= \int_4^{2^{m-1}} x a^{\frac{1}{x}} \frac{1}{x \log 2} \, \mathrm{d}x \sim \int_4^{2^{m-1}} a^{\frac{1}{x}} \, \mathrm{d}x \\
&= \int_4^{2^{m-1}} \exp\left(\frac{\log a}{x}\right) \mathrm{d}x = \int_{4/\log a}^{2^{m-1}/\log a} \exp\left(\frac{1}{x}\right) \log a \, \mathrm{d}x \\
&= 2d \log R \int_{2/(d \log R)}^{1/(4 \log R)} \exp\left(\frac{1}{x}\right) \mathrm{d}x,
\end{aligned}$$

as $\log a = 2^{m+1} \log R = 2d \log R$. Hence, as

$$\begin{aligned}
\int \exp\left(\frac{1}{x}\right) \mathrm{d}x &= x \exp\left(\frac{1}{x}\right) - \mathrm{Ei}\left(\frac{1}{x}\right) + C \\
&= x \exp\left(\frac{1}{x}\right) + \int_{-\frac{1}{x}}^{\infty} \frac{\exp(-t)}{t} \, \mathrm{d}t + C,
\end{aligned}$$

see [1] for example, we have

$$\begin{aligned}
\int_2^{m-1} 2^x a^{2^{-x}} \, \mathrm{d}x &= 2d \log R \left[\frac{R^4}{4 \log R} - \frac{2R^{\frac{d}{2}}}{d \log R} - \int_{-d \log R/2}^{-4 \log R} \frac{\exp(-t)}{t} \, \mathrm{d}t \right] \\
&= \frac{d}{2} R^4 - 4R^{\frac{d}{2}} + 2d \log R \int_{4 \log R}^{d \log R/2} \frac{\exp(t)}{t} \, \mathrm{d}t.
\end{aligned}$$

As the integrand is monotonically increasing for $t \geq 1$, we can further estimate

$$\int_2^{m-1} 2^x a^{2^{-x}} \, \mathrm{d}x \leq \frac{d}{2} R^4 - 4R^{\frac{d}{2}} + 4d \log R \frac{\exp(\frac{d}{2} \log R)}{d \log R} \left(\frac{d}{2} \log R - 4 \log R \right)$$

provided that $4 \log R \geq 1$ which is satisfied for $R \geq 2$. Thus, we indeed conclude

$$\int_2^{m-1} 2^x a^{2^{-x}} \, \mathrm{d}x \leq \frac{d}{2} R^4 - 4R^{\frac{d}{2}} + 2R^{\frac{d}{2}} \log R \left(\frac{d}{2} - 4 \right) \lesssim R^d.$$

References

1. Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1972.
2. Mazen Ali and Antony Nouy, *Approximation with tensor networks. Part III: Multivariate approximation*, arXiv Preprints **arXiv:2101.11932** (2021).
3. ______, *Approximation theory of tree tensor networks: Tensorized univariate functions*, Constr. Approx. **58** (2023), 463–544.
4. Markus Bachmayr and Wolfgang Dahmen, *Adaptive low-rank methods for problems on Sobolev spaces with error control in L_2*, ESAIM Math. Model. Numer. Anal. **50** (2016), no. 4, 1107–1136.

5. Markus Bachmayr, Anthony Nouy, and Reinhold Schneider, *Approximation by tree tensor networks in high dimensions: Sobolev and compositional functions*, Pure Appl. Funct. Anal. **8** (2023), no. 2, 405–428.
6. Dietrich Braess, *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics*, 2nd ed., Cambridge University Press, Cambridge, 2001.
7. Susanne C. Brenner and L. Ridgway Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts in Applied Mathematics, Springer, New York, NY, 2008.
8. Carl de Boor, *A Practical Guide to Splines*, Applied Mathematical Sciences, vol. 27, Springer, New York, 1978.
9. Antonio Falcó and Wolfgang Hackbusch, *On minimal subspaces in tensor representations*, Found. Comput. Math. **12** (2012), 765–803.
10. Antonio Falcó, Wolfgang Hackbusch, and Anthony Nouy, *Tree-based tensor formats*, SeMA Journal **78** (2018), no. 2, 159–173.
11. Lars Grasedyck, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl. **31** (2010), no. 4, 2029–2054.
12. Lars Grasedyck, Daniel Kressner, and Christine Tobler, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitteilungen **1** (2013), no. 36, 53–78.
13. Michael Griebel and Helmut Harbrecht, *Approximation of bi-variate functions: Singular value decomposition versus sparse grids*, IMA J. Numer. Anal. **34** (2014), no. 1, 28–54.
14. ———, *Singular value decomposition versus sparse grids: Refined complexity estimates*, IMA J. Numer. Anal. **39** (2019), no. 4, 1652–1671.
15. ———, *Analysis of tensor approximation schemes for continuous functions*, Found. Comput. Math. **23** (2023), no. 1, 219–240.
16. Michael Griebel, Helmut Harbrecht, and Reinhold Schneider, *Low-rank approximation of continuous functions in Sobolev spaces with dominating mixed smoothness*, Math. Comput. **92** (2023), 1729–1746.
17. Wolfgang Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Series in Computational Mathematics, Springer, Berlin-Heidelberg, 2012.
18. Wolfgang Hackbusch and Stefan Kühn, *A new scheme for the tensor representation*, J. Fourier Anal. Appl. **15** (2009), 706–722.
19. Ivan V. Oseledets and Eugene E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM J. Sci. Comput. **31** (2009), no. 5, 3744–3759.
20. Reinhold Schneider and André Uschmajew, *Approximation rates for the hierarchical tensor format in periodic Sobolev spaces*, J. Complexity **30** (2014), no. 2, 56–71.
21. Vladimir N. Temlyakov, *Approximation of functions with bounded mixed derivative*, Proc. Steklov Inst. Math. **1** (1986), English translation from the original article in Proc. Steklov Inst. Math. 178.
22. Philipp Trunschke, Anthony Nouy, and Martin Eigel, *Weighted sparsity and sparse tensor networks for least squares approximation*, arXiv Preprints **arXiv:2310.08942** (2023).

Department of Mathematics and Computer Science, University of Basel, Spiegelgasse 1, 4051 Basel, Switzerland

Email address: {emily.gajendran,helmut.harbrecht,remo.vonrickenbach}@unibas.ch