LOW-RANK APPROXIMATION OF CONTINUOUS FUNCTIONS IN SOBOLEV SPACES WITH DOMINATING MIXED SMOOTHNESS

MICHAEL GRIEBEL, HELMUT HARBRECHT, AND REINHOLD SCHNEIDER

Abstract. Let $\Omega_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, be given domains. In this article, we study the low-rank approximation with respect to $L^2(\Omega_1 \times \cdots \times \Omega_m)$ of functions from Sobolev spaces with dominating mixed smoothness. To this end, we first estimate the rank of a bivariate approximation, i.e., the rank of the continuous singular value decomposition. In comparison to the case of functions from Sobolev spaces with isotropic smoothness, compare [13, 14], we obtain improved results due to the additional mixed smoothness. This convergence result is then used to study the tensor train decomposition as a method to construct multivariate low-rank approximations of functions from Sobolev spaces with dominating mixed smoothness. We show that this approach is able to beat the curse of dimension.

1. Introduction

Many problems in science and engineering lead to problems which are defined on the tensor product of two domains $\Omega_1 \subset \mathbb{R}^{n_1}$ and $\Omega_2 \subset \mathbb{R}^{n_2}$. Examples arise from the second moment analysis of partial differential domains with stochastic input parameters [20, 21, 30], two-scale homogenization [2, 7, 22], radiosity models and radiative transfer [35], or space-time discretizations of parabolic problems [16]. All these problems are directly given on the product of two domains. Furthermore, many problems are posed on higher-order product domains $\Omega_1 \times \cdots \times \Omega_m$ with $\Omega_i \subset \mathbb{R}^{n_i}$, $i = 1, \ldots, m$. Prominent examples are non-Newtonian flows. These can be modelled by a coupled system which consists of the Navier Stokes equation for the flow in a three-dimensional geometry described by Ω_1 and of the Fokker-Planck equation in a configuration space $\Omega_2 \times \cdots \times \Omega_m$ consisting of $m-1$ spheres. Here, m denotes the number of atoms in a chain-like molecule which constitutes the non-Newtonian behaviour of the flow, for details see [6, 23, 24]. Another class of examples arises from uncertainty quantification, where one has the product of a physical domain Ω_1 with a high-dimensional cube $\Omega_2 \times \cdots \times \Omega_m = [-1, 1]^{m-1}$ for the stochastic parameter,

2000 *Mathematics Subject Classification.* 41A17, 41A25, 41A30, 41A65.

Key words and phrases. Low-rank approximation, Sobolev spaces with dominating mixed smoothness, approximation error, rank complexity.

compare e.g. [10]. A third example stems from multiscale homogenization. Then, each scale i involves a corresponding physical domain Ω_i and we encounter a problem on the product domain $\Omega_1 \times \cdots \times \Omega_m$ for multiscale homogenization with m well-separated scales.

In this article, we therefore study the low-rank approximation for problems posed on product domains. We start our investigations first with the convergence of the bivariate approximation

$$f_R(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{r=1}^{R} g_r^{(1)}(\boldsymbol{x}_1) g_r^{(2)}(\boldsymbol{x}_2) \tag{1.1}$$

with respect to $L^2(\Omega_1 \times \Omega_2)$. While it is well known that the optimal rank R can be determined by the truncated singular value decomposition, the *convergence* with respect to the rank R is not so easy to determine. In [13, 14], the question of the optimal rank R has been answered in case of functions from isotropic Sobolev spaces. But the technique used there yields no gain if additional smoothness is provided by Sobolev spaces with dominant mixed smoothness. In order to exploit such extra regularity, we construct in this article specific low-rank approximations with known convergence properties by means of sparse tensor product approximations. These are known to exploit dominating mixed smoothness in an optimal way. As a consequence, we are able to improve the results from [13, 14] considerably. Indeed, the decay of the singular values is up to a factor two faster than in case of functions with isotropic Sobolev smoothness.

We then consider the situation of bivariate approximation if also mixed Sobolev smoothness is provided on each subdomain. This means that we study the low-rank approximation

$$f_R(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell, \boldsymbol{x}_{\ell+1}, \ldots, \boldsymbol{x}_m) = \sum_{r=1}^{R} g_r^{(1)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell) g_r^{(2)}(\boldsymbol{x}_{\ell+1}, \ldots, \boldsymbol{x}_m) \tag{1.2}$$

if mixed Sobolev smoothness is provided not only between the two subdomains $\Omega_1 \times \cdots \times \Omega_\ell$ and $\Omega_{\ell+1} \times \cdots \times \Omega_m$, but additionally *within* each of these subdomains as well. Thus, we have mixed smoothness for the full m-variate situation. We like to mention that our findings generalize the results from [31, 32, 33] for periodic functions on the m-cube to arbitrary product domains. We allow moreover arbitrary product domains with possibly different smoothness indices on each subdomain. Our results, however, coincide with [31, 32, 33] in the simple setting of $\Omega_1 = \cdots = \Omega_m = [0, 1]$ and periodic functions from Sobolev spaces of dominating mixed smoothness.

After studying the convergence of the approximation (1.2), we are ready to consider the tensor train approximation. The tensor train is a tensor format which can be used to efficiently approximate multivariate functions, compare [18, 27]. As we will see, this format is able to essentially beat the curse of dimension in case of functions with dominating mixed derivatives. Note that the tensor train (TT) format is a particular architecture of the hierarchical Tucker (HT) format [19]. Such tensor representations are known in computational quantum physics and quantum information theory as tensor networks or more precisely, as tree-based tensor networks [1]. All these names are used in the literature [3]. The present investigations can be extended to the hierarchical Tucker format by using the same or similar ideas, cf. [29]. This generalization is obvious, but requires an extended machinery of notations, definitions, etc. For the sake of simplicity of presentation, we refrain from a detailed consideration here and merely focus on the tensor train format.

The remainder of this article is organized as follows. In Section 2, we specify the requirements of multiscale hierarchies on each subdomain. They will be used to construct appropriate sparse tensor approximations in case of multivariate functions in Section 3. In Section 4, we compute bounds on the truncation error of the singular value decomposition (1.1) in the case of functions $f \in L^2(\Omega_1 \times \Omega_2)$. In Section 5, we then consider bounds of the truncated singular value decomposition (1.2) in the case of functions $f \in L^2(\Omega_1 \times \cdots \times \Omega_m)$. Then, in Section 6, we use the results of the previous sections to establish bounds for the tensor train format in the continuous setting. Finally, we state concluding remarks in Section 7.

Throughout this article, the notion "essential" in the context of complexity estimates means "up to logarithmic terms". Moreover, to avoid the repeated use of generic but unspecified constants, we denote by $C \lesssim D$ that C is bounded by a multiple of D independently of parameters on which C and D may depend. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \sim D$ as $C \lesssim D$ and $C \gtrsim D$.

2. Approximation on the subdomains

Let $\Omega \in \mathbb{R}^n$ be a sufficiently smooth, bounded domain.[1] We consider a nested sequence of finite dimensional subspaces

$$V_0 \subset V_1 \subset \cdots \subset V_j \subset \cdots \subset L^2(\Omega), \quad V_j = \operatorname{span}\{\Phi_j\}, \tag{2.3}$$

which consists of piecewise polynomial ansatz functions $\Phi_j := \{\varphi_{j,k} : k \in \Delta_j\}$, where Δ_j denotes a suitable index set, such that $\dim V_j \sim 2^{jn}$ and

$$L^2(\Omega) = \overline{\bigcup_{j \in \mathbb{N}_0} V_j}. \tag{2.4}$$

[1] An n-dimensional, smooth, compact, and orientable manifold in $\mathbb{R}^{n+1}$ can also be considered here and in the following.

Since we intend to approximate functions in these spaces V_j, we assume that the approximation property

$$\inf_{v_j \in V_j} \|u - v_j\|_{L^2(\Omega)} \lesssim h_j^s \|u\|_{H^s(\Omega)}, \quad u \in H^s(\Omega), \tag{2.5}$$

holds for $0 \leq s \leq r$ uniformly in j. Here we set $h_j := 2^{-j}$, i.e., h_j corresponds to the width of the mesh associated with the subspace V_j on Ω. The norm in $H^s(\Omega)$ is defined as usual, see [36] for example, while the integer r refers to the *polynomial exactness*, that is the maximal order of polynomials which are locally contained in the space V_j.

We now introduce a wavelet basis associated with the multiscale analysis (2.3) and (2.4) as follows: The wavelets $\Psi_j := \{\psi_{j,k} : k \in \nabla_j\}$, where $\nabla_j := \Delta_j \setminus \Delta_{j-1}$, are the bases of the complementary spaces W_j of V_{j-1} in V_j, i.e.,

$$V_j = V_{j-1} \oplus W_j, \quad V_{j-1} \cap W_j = \{0\}, \quad W_j = \operatorname{span}\{\Psi_j\}.$$

Recursively we obtain

$$V_J = \bigoplus_{j=0}^{J} W_j, \quad W_0 := V_0,$$

and thus, with

$$\Psi_J := \bigcup_{j=0}^{J} \Psi_j, \quad \Psi_0 := \Phi_0,$$

we get a wavelet basis in V_J. A final requirement is that the infinite collection $\Psi := \bigcup_{j\geq 0} \Psi_j$ forms a Riesz basis of $L^2(\Omega)$. Then, there exists also a biorthogonal, or dual, wavelet basis $\widetilde{\Psi} = \bigcup_{j\geq 0} \widetilde{\Psi}_j = \{\widetilde{\psi}_{j,k} : k \in \nabla_j,\ j \geq 0\}$ which defines a dual multiscale analysis, compare e.g. [8] for further details. In particular, each function $f \in L^2(\Omega)$ admits the unique representation

$$f = \sum_{j=0}^{\infty} \sum_{k \in \nabla_j} (f, \widetilde{\psi}_{j,k})_{L^2(\Omega)} \psi_{j,k}. \tag{2.6}$$

With the definition of the projections

$$Q_j : L^2(\Omega) \to W_j, \quad Q_j f = \sum_{k \in \nabla_j} (f, \widetilde{\psi}_{j,k})_{L^2(\Omega)} \psi_{j,k}$$

the atomic decomposition (2.6) gives rise to the multilevel decomposition

$$f = \sum_{j=0}^{\infty} Q_j f.$$

Then, for any $f \in H^s(\Omega)$, the approximation property (2.5) induces the estimate

$$\|Q_j f\|_{L^2(\Omega)} \lesssim 2^{-js}\|f\|_{H^s(\Omega)}, \quad 0 \le s \le r. \tag{2.7}$$

3. Sparse tensor product spaces

Consider now two domains $\Omega_1 \subset \mathbb{R}^{n_1}$ and $\Omega_2 \subset \mathbb{R}^{n_2}$ with $n_1, n_2 \in \mathbb{N}$. We aim at the approximation of functions in $L^2(\Omega_1 \times \Omega_2)$. To this end, we assume individually for each subdomain Ω_i, $i = 1, 2$, the multiscale analyses

$$V_0^{(i)} \subset V_1^{(i)} \subset V_2^{(i)} \subset \cdots \subset L^2(\Omega_i), \quad V_j^{(i)} = \operatorname{span}\{\Phi_j^{(i)}\}, \quad i = 1, 2,$$

with associated complementary spaces

$$V_j^{(i)} = V_{j-1}^{(i)} \oplus W_j^{(i)}, \quad V_{j-1}^{(i)} \cap W_j^{(i)} = \{0\}, \quad W_j^{(i)} = \operatorname{span}\{\Psi_j^{(i)}\}.$$

Furthermore, let us denote the polynomial exactnesses of the spaces $V_j^{(1)}$ and $V_j^{(2)}$ by r_1 and r_2, respectively.

In this article, we employ the special sparse tensor product space[2]

$$\widehat{V}_J^\sigma := \bigoplus_{j_1\sigma + \frac{j_2}{\sigma} \le J} W_{j_1}^{(1)} \otimes W_{j_2}^{(2)} = \bigoplus_{j_1\sigma + \frac{j_2}{\sigma} = J} V_{j_1}^{(1)} \otimes W_{j_2}^{(2)} \tag{3.8}$$

for an *arbitrary* parameter $\sigma > 0$. In particular, the index pairs $(j_1, j_2) \in \mathbb{N}_0 \times \mathbb{N}_0$ of the included tensor product spaces $W_{j_1}^{(1)} \otimes W_{j_2}^{(2)}$ satisfy the relations

$$0 \le j_1 \le \frac{1}{\sigma}J - \frac{1}{\sigma^2}j_2, \quad 0 \le j_2 \le \sigma J - \sigma^2 j_1.$$

Reasonable choices of the parameter σ could be as follows:

- We may equilibrate the degrees of freedom in all tensor product spaces $W_{j_1}^{(1)} \otimes W_{j_2}^{(2)}$, that is the dimension $\dim(W_{j_1}^{(1)} \otimes W_{j_2}^{(2)}) = \dim(W_{j_1}^{(1)}) \cdot \dim(W_{j_2}^{(2)})$, whose indices (j_1, j_2) satisfy $j_1\sigma + j_2/\sigma = J$. This choice leads to $\sigma = \sqrt{n_1/n_2}$.
- The sparse tensor product space $\widehat{V}_J^\sigma$ (3.8) can be rewritten as

$$\widehat{V}_J^\sigma = \sum_{j_1\sigma + j_2/\sigma = J} V_{j_1}^{(1)} \otimes V_{j_2}^{(2)}.$$

  Then, it can be seen easily that the choice $\sigma := \sqrt{r_1/r_2}$ equilibrates the approximation power of the contained tensor product spaces $V_{j_1}^{(1)} \otimes V_{j_2}^{(2)}$.

[2]Here and in the following, the summation limits are in general no natural numbers and must of course be rounded properly. We leave this to the reader to avoid cumbersome floor/ceil-notations.

- Following the idea of an *equilibrated cost-benefit rate* (see [5]), we get the condition
$$2^{j_1(n_1+r_1)} \cdot 2^{j_2(n_2+r_2)} \stackrel{!}{=} 2^{J\cdot const}.$$
Then, by choosing $const = \sqrt{(n_1+r_1)(n_2+r_2)}$, we arrive at $\sigma = \sqrt{\frac{n_1+r_1}{n_2+r_2}}$.

We now repeat the following results from [11] as they will be essential for our analysis of low-rank approximations in Sobolev spaces of dominating mixed smoothness.

Theorem 3.1 (see [11]). *The dimension of the sparse tensor product space*
$$\widehat{V}_J^\sigma = \bigoplus_{\sigma j_1 + j_2/\sigma \le J} W_{j_1}^{(1)} \otimes W_{j_2}^{(2)} \tag{3.9}$$
is essentially $\mathcal{O}(2^{J\max\{n_1/\sigma, n_2\sigma\}})$. *More precisely, it holds*
$$\dim \widehat{V}_J^\sigma \lesssim \begin{cases} 2^{J\max\{n_1/\sigma, n_2\sigma\}}, & \textit{if } n_1/\sigma \neq n_2\sigma, \\ 2^{Jn_2\sigma} J, & \textit{if } n_1/\sigma = n_2\sigma. \end{cases} \tag{3.10}$$

The constant in the estimate (3.10) depends on the particular choice of σ. Note that the sparse tensor product spaces $\widehat{V}_J^\sigma$ contains essentially less degrees of freedom than the full tensor product space $V_{J/\sigma}^{(1)} \otimes V_{J\sigma}^{(2)}$, which possesses, up to a constant, $2^{J(n_1/\sigma + n_2\sigma)}$ degrees of freedom.

In order to determine the best choice of σ later on, we need to know the rate of approximation in the sparse tensor spaces $\widehat{V}_J^\sigma$. To this end, for $s_1, s_2 \ge 0$, we introduce the anisotropic Sobolev spaces
$$H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2) := H^{s_1}(\Omega_1) \otimes H^{s_2}(\Omega_2),$$
which are defined as tensor product Hilbert spaces with respect to the usual cross norm. Obviously, the highest possible rate of convergence is attained in the space $H_{mix}^{r_1,r_2}(\Omega_1\times\Omega_2)$. Therefore, in the following, we restrict ourselves without loss of generality to $s_1 \le r_1$ and $s_2 \le r_2$.

Theorem 3.2 (see [11]). *Let* $0 < s_1 \le r_1$, $0 < s_2 \le r_2$ *and* $f \in H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)$. *Then, the approximation*
$$\widehat{f}_J = \sum_{j_1\sigma + \frac{j_2}{\sigma} \le J} \big(Q_{j_1}^{(1)} \otimes Q_{j_2}^{(2)}\big) f \in \widehat{V}_J^\sigma \tag{3.11}$$
satisfies
$$\|f - \widehat{f}_J\|_{L^2(\Omega_1\times\Omega_2)} \lesssim \begin{cases} 2^{-J\min\{s_1/\sigma, s_2\sigma\}} \|f\|_{H_{mix}^{s_1,s_2}(\Omega_1\times\Omega_2)}, & \textit{if } s_1/\sigma \neq s_2\sigma, \\ 2^{-Js_1/\sigma}\sqrt{J}\|f\|_{H_{mix}^{s_1,s_2}(\Omega_1\times\Omega_2)}, & \textit{if } s_1/\sigma = s_2\sigma. \end{cases} \tag{3.12}$$

The constant in estimate (3.12) depends again on the particular choice of σ. Moreover, if $s_1 < r_1$ and $s_2 < r_2$, the factor $\sqrt{J}$ for the case $s_1/\sigma = s_2\sigma$ in (3.12) can be removed by using more sophisticated estimates, compare [17, 28].

Note that the combination of Theorems 3.1 and 3.2 implies that, for any σ which satisfies the inequalities

$$\min\left\{\frac{r_1}{r_2}, \frac{n_1}{n_2}\right\} \le \sigma^2 \le \max\left\{\frac{r_1}{r_2}, \frac{n_1}{n_2}\right\},$$

the sparse tensor product spaces $\widehat{V}_J^\sigma$ offer essentially the same rate of convergence with respect to the degrees of freedom, compare [11]. In the following, we are looking for the low-rank approximation of functions. The key idea is to use the sparse tensor product space $\widehat{V}_J^\sigma$ as a tool to bound the rank properly.

4. Bivariate mixed Sobolev smoothness

We first like to estimate the rank which is required to represent functions in $\widehat{V}_J^\sigma$. To this end, we make use of the fact that the sparse tensor product space is given as a direct sum of tensor products of single-scale spaces $V_{j_1}^{(1)}$ and complement spaces $W_{j_2}^{(2)}$ in accordance with

$$\widehat{V}_J^\sigma = \bigoplus_{j_2=0}^{J\sigma} \bigoplus_{j_1=0}^{J/\sigma - j_2/\sigma^2} W_{j_1}^{(1)} \otimes W_{j_2}^{(2)} = \bigoplus_{\sigma j_1 + j_2/\sigma = J} V_{j_1}^{(1)} \otimes W_{j_2}^{(2)}, \tag{4.13}$$

compare (3.9). To this end, we make use of the bases $\Phi_j^{(1)}$ and $\Psi_j^{(2)}$ and exploit that any function

$$\widehat{f}_J(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\sigma j_1 + j_2/\sigma = J} \sum_{k_1 \in \Delta_{j_1}} \sum_{k_2 \in \nabla_{j_2}} c_{(j_1,k_1),(j_2,k_2)} \varphi_{j_1,k_1}^{(1)}(\boldsymbol{x}_1) \psi_{j_2,k_2}^{(2)}(\boldsymbol{x}_2) \in \widehat{V}_J^\sigma$$

can be rewritten as a tensor product function of finite rank in accordance with

$$\begin{aligned}\widehat{f}_J(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \sum_{\sigma j_1 + j_2/\sigma = J} \sum_{k_2 \in \nabla_{j_2}} g_{j_1,j_2,k_2}^{(1)}(\boldsymbol{x}_1) \psi_{j_2,k_2}^{(2)}(\boldsymbol{x}_2) \\ &= \sum_{\sigma j_1 + j_2/\sigma = J} \sum_{k_1 \in \Delta_{j_1}} \varphi_{j_1,k_1}^{(1)}(\boldsymbol{x}_1) g_{j_1,j_2,k_1}^{(2)}(\boldsymbol{x}_2),\end{aligned} \tag{4.14}$$

where

$$\begin{aligned} g_{j_1,j_2,k_2}^{(1)} &= \sum_{k_1 \in \Delta_{j_1}} c_{(j_1,k_1),(j_2,k_2)} \varphi_{j_1,k_1} \in V_{j_1}^{(1)}, \\ g_{j_1,j_2,k_1}^{(2)} &= \sum_{k_2 \in \nabla_{j_2}} c_{(j_1,k_1),(j_2,k_2)} \psi_{j_2,k_2} \in W_{j_2}^{(2)}. \end{aligned}$$

By switching between both representations in (4.14) for particular combinations of (j_1, j_2), we are able to derive the following result.

Theorem 4.1. *Any function in the sparse tensor product space $\widehat{V}_J^\sigma$ can be represented as a tensor product function*

$$\widehat{f}_J(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{r=0}^{R} g_r^{(1)}(\boldsymbol{x}_1) g_r^{(2)}(\boldsymbol{x}_2)$$

of rank R, where the rank is bounded by

$$R \lesssim R_J^\sigma := 2^{\frac{Jn_1n_2}{n_1/\sigma+\sigma n_2}}.$$

Proof. We start with the identity (4.13). Since the rank of functions in the space $V_{j_1}^{(1)} \otimes W_{j_2}^{(2)}$ is bounded by

$$\min\left\{\dim V_{j_1}^{(1)}, \dim W_{j_2}^{(2)}\right\} \sim \min\{2^{j_1n_1}, 2^{j_2n_2}\},$$

it holds

$$R \lesssim \sum_{j_1\sigma+j_2/\sigma=J} \min\{2^{j_1n_1}, 2^{j_2n_2}\}.$$

Assume that the equilibrium in the bracket is obtained for $j_1^\star n_1 = j_2^\star n_2$, then we can split the index set $\{(j_1, j_2) : j_1\sigma + j_2/\sigma = J\}$ into the two index sets $\mathcal{I}_1 := \{(j_1, j_2) : 0 \le j_1 \le j_1^\star,\ j_2 = J\sigma - j_1\sigma^2\}$ and $\mathcal{I}_2 := \{(j_1, j_2) : 0 \le j_2 \le j_2^\star,\ j_1 = J/\sigma - j_2/\sigma^2\}$. In view of $\min\{2^{j_1n_1}, 2^{j_2n_2}\} = 2^{j_1n_1}$ for all $(j_1, j_2) \in \mathcal{I}_1$ and $\min\{2^{j_1n_1}, 2^{j_2n_2}\} = 2^{j_2n_2}$ for all $(j_1, j_2) \in \mathcal{I}_2$, we obtain

$$R \lesssim \sum_{j_1=0}^{j_1^\star} 2^{j_1n_1} + \sum_{j_2=0}^{j_2^\star} 2^{j_2n_2}. \tag{4.15}$$

Since it always holds $j_1\sigma + j_2/\sigma = J$, we can determine $j_1^\star$ from the equation

$$j_1^\star n_1 = (J\sigma - j_1^\star\sigma^2)n_2,$$

i.e.,

$$j_1^\star = \frac{Jn_2}{n_1/\sigma + \sigma n_2} \quad \text{and} \quad j_2^\star = \frac{Jn_1}{n_1/\sigma + \sigma n_2}.$$

Inserting this into (4.15) implies the assertion

$$R \lesssim 2^{\frac{Jn_1n_2}{n_1/\sigma+\sigma n_2}}.$$

□

Now, by combining Theorem 3.2 with Theorem 4.1, we can express the convergence rate in terms of the rank R. This gives us an upper bound of the truncation error of the singular value decomposition for functions in spaces with dominating mixed smoothness.

Corollary 4.2. *Let $0 < s_1 \le r_1$, $0 < s_2 \le r_2$ and $f \in H^{s_1,s_2}_{mix}(\Omega_1 \times \Omega_2)$. Choose $\sigma > 0$ arbitrarily and set*

$$\beta := \frac{n_1/\sigma + \sigma n_2}{n_1 n_2} \min\{s_1/\sigma, s_2\sigma\} = \frac{1}{n_1 n_2} \min\left\{\frac{s_1 n_1}{\sigma^2} + s_1 n_2, s_2 n_1 + s_2 n_2 \sigma^2\right\}, \tag{4.16}$$

Then, there exists a rank-R approximation

$$f_R(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{r=1}^{R} g_r^{(1)}(\boldsymbol{x}_1) g_r^{(2)}(\boldsymbol{x}_2) \in L^2(\Omega_1 \times \Omega_2)$$

which approximates f as

$$\|f - f_R\|_{L^2(\Omega_1\times\Omega_2)} \lesssim \begin{cases} R^{-\beta}\|f\|_{H^{s_1,s_2}_{mix}(\Omega_1\times\Omega_2)}, & \text{if } s_1/\sigma \ne s_2\sigma, \\ R^{-\beta}\sqrt{\log R}\|f\|_{H^{s_1,s_2}_{mix}(\Omega_1\times\Omega_2)}, & \text{if } s_1/\sigma = s_2\sigma. \end{cases}$$

The constants in these estimates do depend on s_1, s_2, n_1, n_2, and σ, but not on the rank R.

Proof. Let $s_1/\sigma \ne s_2\sigma$ and observe that $R \lesssim R_J^\sigma = 2^{\frac{J n_1 n_2}{n_1/\sigma + \sigma n_2}}$ due to Theorem 4.1. In order to estimate the convergence with respect to the rank, we assume $R = R_J^\sigma$, which implies

$$R^{-\beta} = R^{-\frac{n_1/\sigma+\sigma n_2}{n_1 n_2}\min\{s_1/\sigma, s_2\sigma\}} \sim 2^{-J\min\{s_1/\sigma, s_2\sigma\}}.$$

This yields the first error estimate in view of (3.12).

In case of $s_1/\sigma = s_2\sigma$, the additional factor $\sqrt{J} \lesssim \sqrt{\log R}$ needs to be inserted as a multiplicative factor. This completes the proof. $\square$

The optimal rate of convergence with respect to the rank is given if the expressions in the minimum in (4.16) are balanced, i.e., if

$$\frac{s_1 n_1}{\sigma^2} + s_1 n_2 = s_2 n_1 + s_2 n_2 \sigma^2.$$

Straighforward calculation yields

$$\sigma = \sqrt{\frac{s_1}{s_2}},$$

which means we should equilibrate the approximation power in the underlying sparse tensor product space. This would yield the rank estimate

$$\|f - f_R\|_{L^2(\Omega_1\times\Omega_2)} \lesssim R^{-\beta}\sqrt{\log R}\|f\|_{H^{s_1,s_2}_{mix}(\Omega_1\times\Omega_2)} \tag{4.17}$$

with

$$\beta = \frac{s_1 n_2 + s_2 n_1}{n_1 n_2} = \frac{s_1}{n_1} + \frac{s_2}{n_2}. \tag{4.18}$$

In contrast, in [14], we were able to prove only the rate

$$\beta = \max\left\{\frac{s_1}{n_1}, \frac{s_2}{n_2}\right\}.$$

Especially, the latter rate is already achieved for functions in the related isotropic Sobolev space

$$H_{iso}^{s_1,s_2}(\Omega_1 \times \Omega_2) := H_{mix}^{s_1,0}(\Omega_1 \times \Omega_2) \cap H_{mix}^{0,s_2}(\Omega_1 \times \Omega_2).$$

As a consequence, the singular values of a function from $H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)$ converge by a factor up to two faster than the singular values of a function from $H_{iso}^{s_1,s_2}(\Omega_1 \times \Omega_2)$.

To conclude our study on the bivariate situation, we state the following result for the fully discrete singular value decomposition (where the eigenfunctions are approximated with respect to single-scale spaces of fixed level of resolution as introduced in Section 2):

Corollary 4.3. *The number of degrees of freedom, which is needed to approximate a function $f \in H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)$ by the truncated singular value decomposition to a prescribed accuracy ε, is*

$$\mathrm{dof}_{svd}(\varepsilon) \sim \varepsilon^{-\frac{n_1 n_2}{s_1 n_2 + s_2 n_1}} \varepsilon^{-\max\{\frac{n_1}{\min\{s_1,r_1\}}, \frac{n_2}{\min\{s_2,r_2\}}\}}. \tag{4.19}$$

In contrast, the sparse tensor product approximation requires essentially

$$\mathrm{dof}_{sg}(\varepsilon) \sim \varepsilon^{-\max\left\{\frac{n_1}{\min\{s_1,r_1\}}, \frac{n_2}{\min\{s_2,r_2\}}\right\}} \tag{4.20}$$

degrees of freedom to approximate functions from the anisotropic Sobolev space $H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)$, compare [11]. We emphasize, however, that the multiscale representation (4.14) of the singular value decomposition has the same complexity as the sparse tensor product approximation (4.20). This implies that the eigenfunctions can be stored in a compressed format to further improve the complexity (4.19) towards that of the respective sparse tensor product approximation.

Remark 4.4. *We should comment on the sharpness of our findings. At least in the simple bivariate situation with $\Omega := \Omega_1 = \Omega_2 \subset \mathbb{R}^n$, the rank estimate given by* (4.17), (4.18) *is sharp. This is seen by considering Matèrn kernels $k_\nu(\boldsymbol{x}_1, \boldsymbol{x}_2)$, also called Sobolev splines, where $\nu \geq 1/2$ is the smoothness parameter, compare* [9, 25]*. They are known to be the reproducing kernels of the Sobolev spaces $H^{\nu+n/2}(\Omega)$, hence they are themselves elements of $H^{2\nu+n}(\Omega \times \Omega)$ and consequently also elements of $H_{mix}^{s_1,s_2}(\Omega \times \Omega)$ for any $s_1 + s_2 = 2\nu + n$. The rank estimate given by* (4.17), (4.18) *coincides then essentially with Weyl's law* [34].

5. Multivariate mixed Sobolev smoothness

We shall generalize the results from the previous section to the bivariate approximation of functions which provide multivariate dominating mixed derivatives. This means that we consider m domains $\Omega_i \subset \mathbb{R}^{n_i}$ with $n_i \in \mathbb{N}$ for all $i = 1, 2, \ldots, m$ and aim at the bivariate approximation of functions in the anisotropic Sobolev spaces

$$\mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega}) := H^{s_1}(\Omega_1) \otimes H^{s_2}(\Omega_2) \otimes \cdots \otimes H^{s_m}(\Omega_m),$$

which are defined on the m-fold product domain $\boldsymbol{\Omega} := \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m$. To this end, we introduce the generalized m-fold sparse tensor product space

$$\widehat{V}_J^{\boldsymbol{\alpha}} := \bigoplus_{\boldsymbol{\alpha}^T \mathbf{j} \leq J} W_{j_1}^{(1)} \otimes W_{j_2}^{(2)} \otimes \cdots \otimes W_{j_m}^{(m)}$$

for an *arbitrary* vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m) > \mathbf{0}$ and $\mathbf{j} = (j_1, j_2, \ldots, j_m) \in \mathbb{N}_0^m$.

In accordance with [12], we have the following approximation results with respect to the sparse tensor product space $\widehat{V}_J^{\boldsymbol{\alpha}}$.

Theorem 5.1 (see [12]). *Let $\mathbf{0} \leq \mathbf{s} \leq \mathbf{r}$ and $f \in \mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})$. Then, the projector*

$$\widehat{Q}_J^{\boldsymbol{\alpha}} : \mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega}) \to \widehat{\mathbf{V}}_J^{\boldsymbol{\alpha}}, \quad \widehat{Q}_J^{\boldsymbol{\alpha}} f = \sum_{\boldsymbol{\alpha}^T \mathbf{j} \leq J} \big(Q_{j_1}^{(1)} \otimes Q_{j_2}^{(2)} \otimes \cdots \otimes Q_{j_m}^{(m)}\big) f \tag{5.21}$$

on the sparse tensor product space $\widehat{\mathbf{V}}_J^{\boldsymbol{\alpha}}$ satisfies

$$\|(I - \widehat{Q}_J^{\boldsymbol{\alpha}}) f\|_{L^2(\boldsymbol{\Omega})} \lesssim 2^{-J \min\{\frac{s_1}{\alpha_1}, \frac{s_2}{\alpha_2}, \ldots, \frac{s_m}{\alpha_m}\}} J^{(P-1)/2} \|f\|_{\mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})}. \tag{5.22}$$

Here, P counts how often the minimum is attained in the exponent.

After having identified the approximation power of the m-fold sparse tensor product space $\widehat{V}_J^{\boldsymbol{\alpha}}$ when representing a given function $f \in \mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})$, we shall next estimate the rank of this sparse tensor product approximation like in the bivariate situation. Therefore, we again make use of the bases $\Phi_j^{(1)}$ and $\Psi_j^{(i)}$, $i = 2, 3, \ldots, m$, and consider a given function

$$\widehat{f}_J(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \sum_{\boldsymbol{\alpha}^T \mathbf{j} = J} \sum_{k_1 \in \Delta_{j_1}} \sum_{k_2 \in \nabla_{j_2}} \cdots \sum_{k_m \in \nabla_{j_m}} c_{\mathbf{j},\mathbf{k}} \varphi_{j_1,k_1}^{(1)}(\boldsymbol{x}_1) \psi_{j_2,k_2}^{(2)}(\boldsymbol{x}_2) \cdots \psi_{j_m,k_m}^{(2)}(\boldsymbol{x}_m) \in \widehat{V}_J^{\boldsymbol{\alpha}}.$$

We intend to estimate the rank when separating the variables $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell)$ and $(\boldsymbol{x}_{\ell+1}, \ldots, \boldsymbol{x}_m)$, where $1 \leq \ell < m$ can be chosen arbitrary, but is fixed throughout this section. To this end, we have to bound the rank in each of the tensor product spaces $V_{j_1}^{(1)} \otimes W_{j_2}^{(2)} \otimes \cdots \otimes W_{j_m}^{(m)}$ when splitting the first ℓ variables from the last $m - \ell$ variables. The rank is given by the minimum of the dimensions of the particular subspaces, i.e.,

$$\operatorname{rank}(V_{j_1}^{(1)} \otimes W_{j_2}^{(2)} \otimes \cdots \otimes W_{j_m}^{(m)}) = \min\big\{2^{j_1 n_1 + \cdots + j_\ell n_\ell}, 2^{j_{\ell+1} n_{\ell+1} + \cdots + j_m n_m}\big\}.$$

We now have to sum up over all admissible $j_1, j_2, \dots, j_m$, i.e., we have to compute the sum

$$\sum_{\boldsymbol{\alpha}^T \mathbf{j} = J} \min\left\{2^{j_1 n_1 + \cdots + j_\ell n_\ell}, 2^{j_{\ell+1} n_{\ell+1} + \cdots + j_m n_m}\right\}.$$

In order to bound this sum, we abbreviate

$$\underline{m} := \max_{i \le \ell} \left\{\frac{n_i}{\alpha_i}\right\}, \quad \overline{m} := \max_{i > \ell} \left\{\frac{n_i}{\alpha_i}\right\} \tag{5.23}$$

and note that for given level $\tilde{J}$ with $0 \le \tilde{J} \le J$ there holds

$$\begin{aligned}
\max_{\sum_{i \le \ell} \alpha_i j_i = \tilde{J}} j_1 n_1 + \cdots + j_\ell n_\ell &\le \tilde{J} \underline{m}, \\
\max_{\sum_{i > \ell} \alpha_i j_i = \tilde{J}} j_{\ell+1} n_{\ell+1} + \cdots + j_m n_m &\le \tilde{J} \overline{m}.
\end{aligned}$$

We conclude

$$\sum_{\substack{\sum_{i \le \ell} \alpha_i j_i = \tilde{J} \\ \sum_{i > \ell} \alpha_i j_i = J - \tilde{J}}} \min\left\{2^{j_1 n_1 + \cdots + j_\ell n_\ell}, 2^{j_{\ell+1} n_{\ell+1} + \cdots + j_m n_m}\right\} \le J^{\max\{P_1, P_2\} - 1} \min\left\{2^{\tilde{J} \underline{m}}, 2^{(J - \tilde{J}) \overline{m}}\right\},$$

where $P_1 \le \ell$ and $P_2 \le m - \ell$ count how often the maxima $\underline{m}$ and $\overline{m}$, respectively, are attained in (5.23). Thus, we derive

$$\begin{aligned}
&\sum_{\boldsymbol{\alpha}^T \mathbf{j} = J} \min\left\{2^{j_1 n_1 + \cdots + j_\ell n_\ell}, 2^{j_{\ell+1} n_{\ell+1} + \cdots + j_m n_m}\right\} \\
&\quad = \sum_{\tilde{J}=0}^{J} \sum_{\substack{\sum_{i \le \ell} \alpha_i j_i = \tilde{J} \\ \sum_{i > \ell} \alpha_i j_i = J - \tilde{J}}} \min\left\{2^{j_1 n_1 + \cdots + j_\ell n_\ell}, 2^{j_{\ell+1} n_{\ell+1} + \cdots + j_m n_m}\right\} \\
&\quad \le J^{\max\{P_1, P_2\} - 1} \sum_{\tilde{J}=0}^{J} \min\left\{2^{\tilde{J} \underline{m}}, 2^{(J - \tilde{J}) \overline{m}}\right\}.
\end{aligned}$$

The minimum switches for the level $J\overline{m}/(\underline{m}+\overline{m})$, which leads to

$$\begin{aligned}
&\sum_{\boldsymbol{\alpha}^T\mathbf{j}=J} \min\left\{2^{j_1n_1+\cdots+j_\ell n_\ell}, 2^{j_{\ell+1}n_{\ell+1}+\cdots+j_mn_m}\right\} \\
&\qquad \le J^{\max\{P_1,P_2\}-1}\left[\sum_{\tilde{J}=0}^{J\overline{m}/(\underline{m}+\overline{m})} \min\left\{2^{\tilde{J}\underline{m}}, 2^{(J-\tilde{J})\overline{m}}\right\} + \sum_{\tilde{J}=J\overline{m}/(\underline{m}+\overline{m})}^{J} \min\left\{2^{\tilde{J}\underline{m}}, 2^{(J-\tilde{J})\overline{m}}\right\}\right] \\
&\qquad \le J^{\max\{P_1,P_2\}-1}\left[\sum_{\tilde{J}=0}^{J\overline{m}/(\underline{m}+\overline{m})} 2^{\tilde{J}\underline{m}} + \sum_{\tilde{J}=0}^{J\underline{m}/(\underline{m}+\overline{m})} 2^{\tilde{J}\overline{m}}\right] \\
&\qquad \lesssim J^{\max\{P_1,P_2\}-1}2^{J\underline{m}\overline{m}/(\underline{m}+\overline{m})}.
\end{aligned}$$

This determines the rank of the sparse tensor product approximation, which we exploit in the following theorem.

Theorem 5.2. *Let $0 \le \mathbf{s} \le \mathbf{r}$ and $f \in \mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})$. Then, there exists a rank-R approximation*

$$f_R(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m) = \sum_{r=0}^{R} g_r^{(1)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_\ell)g_r^{(2)}(\boldsymbol{x}_{\ell+1},\ldots,\boldsymbol{x}_m) \in L^2(\boldsymbol{\Omega})$$

which approximates f as

$$\|f - f_R\|_{L^2(\boldsymbol{\Omega})} \lesssim R^{-\beta}(\log R)^{(m-1)/2+\beta(\max\{P_1,P_2\}-1)}\|f\|_{\mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})}. \tag{5.24}$$

Here, the rate β is given by

$$\beta = \min_{i\le\ell}\left\{\frac{s_i}{n_i}\right\} + \min_{i>\ell}\left\{\frac{s_i}{n_i}\right\} \tag{5.25}$$

and P_1 and P_2 count how often the first and second minimum is attained. Moreover, the constants in this error estimate do depend on s_i and n_i, $i = 1, 2, \ldots, m$, but not on the rank R.

Proof. As in the bivariate situation, we shall equilibrate the approximation power of the extremal spaces in the sparse tensor product construction, which means that we choose $\alpha_i = s_i$. Thus, the rank R_J of

$$\widehat{Q}_J^{\mathbf{s}}f(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m) = \sum_{\mathbf{s}^T\mathbf{j}=J}\sum_{k_1\in\Delta_{j_1}}\sum_{k_2\in\nabla_{j_2}}\cdots\sum_{k_m\in\nabla_{j_m}} c_{\mathbf{j},\mathbf{k}}\varphi_{j_1,k_1}^{(1)}(\boldsymbol{x}_1)\psi_{j_2,k_2}^{(2)}(\mathbf{x_2})\cdots\psi_{j_m,k_m}^{(2)}(\boldsymbol{x}_m) \in \widehat{V}_J^{\mathbf{s}}$$

is essentially bounded by

$$R_J \lesssim 2^{J\underline{m}\overline{m}/(\underline{m}+\overline{m})}J^{\max\{P_1,P_2\}-1}, \quad \underline{m} := \max_{i\le\ell}\left\{\frac{n_i}{s_i}\right\}, \quad \overline{m} := \max_{i>\ell}\left\{\frac{n_i}{s_i}\right\}.$$

This is equivalent to $R_J \lesssim 2^{J/\beta} J^{\max\{P_1,P_2\}-1}$ with β as defined in (5.25). Hence, we have

$$2^{-J} \lesssim \left(\frac{R_J}{J^{\max\{P_1,P_2\}-1}}\right)^{-\beta} = R_J^{-\beta} J^{\beta(\max\{P_1,P_2\}-1)}. \tag{5.26}$$

As the convergence rate in (5.22) becomes essentially $2^{-J}\log J^{(m-1)/2}$ due to the specific choice of $\boldsymbol{\alpha} = \boldsymbol{s}$, we can insert (5.26) to get the convergence rate $R_J^{-\beta} J^{(m-1)/2+\beta\max\{P_1,P_2\}-1}$. By using finally $J \lesssim \log R_J$, we conclude the desired error estimate (5.24). $\square$

Remark 5.3. *(i) As we have seen, in case of multivariate mixed Sobolev smoothness, the decay of the singular values is essentially the same independent of the dimension of the two domains $\Omega_1 \times \cdots \times \Omega_\ell$ and $\Omega_{\ell+1} \times \cdots \times \Omega_m$ with which the bivariate approximation is performed. Nonetheless, for the approximation of the left and right eigenfunctions, this holds only true if they are represented in the respective sparse tensor product spaces instead of the full tensor product spaces.*
(ii) In order to optimally estimate the decay of the eigenvalues, one has to choose ansatz spaces which provide sufficient polynomial exactness, i.e., given $f \in \mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})$, one chooses ansatz spaces such that it holds $\mathbf{r} \geq \mathbf{s}$ in Theorem 5.2.

6. Tensor train format

We shall next apply the previous results for estimating the cost of the tensor train approximation in the continuous case. We basically proceed as in [15] and successively apply the singular value decomposition as studied in the previous section. This corresponds to the algorithmic procedure in practical computations. Related results, obtained by other proof techniques, can be found in [1, 3, 18, 29].

6.1. **Tensor train decomposition.** Given a function $f \in \mathbf{H}^{\mathbf{s}}(\boldsymbol{\Omega})$, we separate in the first step of the tensor train decomposition the variables $\boldsymbol{x}_1 \in \Omega_1$ and $(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_m) \in \Omega_2 \times \cdots \times \Omega_m$ by the singular value decomposition

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = \sum_{\alpha_1=1}^{\infty} \sqrt{\lambda_1(\alpha_1)}\varphi_1(\boldsymbol{x}_1, \alpha_1)\psi_1(\alpha_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m).$$

Herein, $\{\varphi_1(\alpha_1)\}_{\alpha_1\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Omega_1)$, $\{\psi_1(\alpha_1)\}_{\alpha_1\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Omega_2 \times \cdots \times \Omega_m)$, and $\left\{\sqrt{\lambda_1(\alpha_1)}\right\}_{\alpha_1\in\mathbb{N}}$ is a square summable sequence of singular values.

Next, since

$$\left[\sqrt{\lambda_1(\alpha_1)}\psi_1(\alpha_1)\right]_{\alpha_1=1}^{\infty} \in \ell^2(\mathbb{N}) \otimes L^2(\Omega_2 \times \cdots \times \Omega_m),$$

we can separate in the second step of the tensor train decomposition $(\alpha_1, \boldsymbol{x}_2) \in \mathbb{N} \times \Omega_2$ from $(\boldsymbol{x}_3, \ldots, \boldsymbol{x}_m) \in \Omega_3 \times \cdots \times \Omega_m$ by means of a second singular value decomposition. This leads to

$$\begin{aligned}\Big[\sqrt{\lambda_1(\alpha_1)}\psi_1(\alpha_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)\Big]_{\alpha_1=1}^{\infty} \\ = \sum_{\alpha_2=1}^{\infty} \sqrt{\lambda_2(\alpha_2)}\Big[\varphi_2(\alpha_1, \boldsymbol{x}_2, \alpha_2)\Big]_{\alpha_1=1}^{\infty} \psi_2(\alpha_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_m).\end{aligned} \tag{6.27}$$

Herein, $\big\{[\varphi_2(\alpha_1, \alpha_2)]_{\alpha_1\in\mathbb{N}}\big\}_{\alpha_2\in\mathbb{N}}$ is an orthonormal basis of $\ell^2(\mathbb{N}) \otimes L^2(\Omega_2)$, $\{\psi_2(\alpha_2)\}_{\alpha_1\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Omega_3 \times \cdots \times \Omega_m)$, and $\big\{\sqrt{\lambda_2(\alpha_2)}\big\}_{\alpha_2\in\mathbb{N}}$ is a square summable sequence of singular values.

By repeating the second step and successively separating $(\alpha_{j-1}, \boldsymbol{x}_j) \in \mathbb{N} \times \Omega_j$ from $(\boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_m) \in \Omega_{j+1} \times \cdots \times \Omega_m$ for $j = 3, \ldots, m-1$, we finally arrive at the representation

$$\begin{aligned}f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \sum_{\alpha_1=1}^{\infty} \cdots \sum_{\alpha_{m-1}=1}^{\infty} \varphi_1(\alpha_1, \boldsymbol{x}_1)\varphi_2(\alpha_1, \boldsymbol{x}_2, \alpha_2) \\ \cdots \varphi_{m-1}(\alpha_{m-2}, \boldsymbol{x}_{m-1}, \alpha_{m-1})\varphi_m(\alpha_{m-1}, \boldsymbol{x}_m),\end{aligned}$$

where

$$\varphi_m(\alpha_{m-1}, \boldsymbol{x}_m) = \sqrt{\lambda_{m-1}(\alpha_{m-1})}\psi_{m-1}(\alpha_{m-1}, \boldsymbol{x}_m).$$

In practice, we truncate the singular value decomposition in step j after r_j terms, thus arriving at the finite dimensional representation

$$\begin{aligned}f^{\mathrm{TT}}_{r_1,\ldots,r_{m-1}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{m-1}=1}^{r_{m-1}} \varphi_1(\alpha_1, \boldsymbol{x}_1)\varphi_2(\alpha_1, \boldsymbol{x}_2, \alpha_2) \\ \cdots \varphi_{m-1}(\alpha_{m-2}, \boldsymbol{x}_{m-1}, \alpha_{m-1})\varphi_m(\alpha_{m-1}, \boldsymbol{x}_m).\end{aligned}$$

One readily infers by using Pythagoras' theorem that the truncation error is then given by

$$\|f - f^{\mathrm{TT}}_{r_1,\ldots,r_{m-1}}\|_{L^2(\Omega_1\times\cdots\times\Omega_m)} \leq \sqrt{\sum_{j=1}^{m-1} \sum_{\alpha_j=r_j+1}^{\infty} \lambda_j(\alpha_j)},$$

compare [4, Proposition 9]. Note that, for $j \geq 2$, the singular values $\{\lambda_j(\alpha)\}_{\alpha\in\mathbb{N}}$ in this estimate do not coincide with the singular values from the untruncated tensor train decomposition due to the truncation after the ranks r_j.

6.2. **Regularity.** We shall next give bounds on the input data for the singular value decomposition and the truncation rank in each step of the tensor train decomposition. We mention that our analysis covers the computational practice: We compute successively the truncated singular value decomposition with prescribed accuracy $\varepsilon > 0$ as it is done in practice.

In the first step of the tensor train decomposition, we compute the function

$$\boldsymbol{g}_1(\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m) := \left[\sqrt{\lambda_1(\alpha_1)}\psi_1(\alpha_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m)\right]_{\alpha_1=1}^{r_1}.$$

It satisfies

$$\begin{aligned}
\|\boldsymbol{g}_1\|^2_{[\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)]^{r_1}} &= \sum_{\alpha_1=1}^{r_1} \lambda_1(\alpha_1)\|\psi_1(\alpha_1)\|^2_{\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)}\\
&= \sum_{\alpha_1=1}^{r_1} \lambda_1(\alpha_1)\|\varphi_1(\alpha_1)\otimes\psi_1(\alpha_1)\|^2_{L^2(\Omega_1)\otimes\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)}\\
&= \left\|\sum_{\alpha_1=1}^{r_1}\sqrt{\lambda_1(\alpha_1)}\varphi_1(\alpha_1)\otimes\psi_1(\alpha_1)\right\|^2_{L^2(\Omega_1)\otimes\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)},
\end{aligned}$$

where we used the notation $\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1) := H^{s_2}(\Omega_2)\otimes\cdots\otimes H^{s_m}(\Omega_m)$. Vice versa, we have by using Pythagoras' theorem

$$\begin{aligned}
\|f\|^2_{L^2(\Omega_1)\otimes\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)} &= \left\|\sum_{\alpha_1=1}^{\infty}\sqrt{\lambda_1(\alpha_1)}\varphi_1(\alpha_1)\otimes\psi_1(\alpha_1)\right\|^2_{L^2(\Omega_1)\otimes\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)}\\
&= \left\|\sum_{\alpha_1=1}^{r_1}\sqrt{\lambda_1(\alpha_1)}\varphi_1(\alpha_1)\otimes\psi_1(\alpha_1)\right\|^2_{L^2(\Omega_1)\otimes\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)}\\
&\qquad + \left\|\sum_{\alpha_1=r_1+1}^{\infty}\sqrt{\lambda_1(\alpha_1)}\varphi_1(\alpha_1)\otimes\psi_1(\alpha_1)\right\|^2_{L^2(\Omega_1)\otimes\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)}.
\end{aligned}$$

Putting both estimates together yields

$$\|\boldsymbol{g}_1\|_{[\mathbf{H}^{\boldsymbol{t}_1}(\boldsymbol{\Upsilon}_1)]^{r_1}} \le \|f\|_{\mathbf{H}^{\boldsymbol{s}}(\boldsymbol{\Omega})}.$$

In the j-th step of the tensor train decomposition, $j = 2,3,\ldots,m-1$, one computes the singular value decomposition for the vector-valued function

$$\boldsymbol{g}_{j-1}(\boldsymbol{x}_j,\ldots,\boldsymbol{x}_m) := \left[\sqrt{\lambda_{j-1}(\alpha_{j-1})}\psi_{j-1}(\alpha_{j-1},\boldsymbol{x}_j,\ldots,\boldsymbol{x}_m)\right]_{\alpha_{j-1}=1}^{r_{j-1}}.$$

This leads to the representation

$$\boldsymbol{g}_{j-1}(\boldsymbol{x}_j,\ldots,\boldsymbol{x}_m) = \left[\sum_{\alpha_j=1}^{\infty}\sqrt{\lambda_j(\alpha_j)}\varphi_j(\alpha_{j-1},\boldsymbol{x}_j,\alpha_j)\psi_j(\alpha_j,\boldsymbol{x}_{j+1},\ldots,\boldsymbol{x}_m)\right]_{\alpha_{j-1}=1}^{r_{j-1}} \tag{6.28}$$

which separates $(\alpha_{j-1},\boldsymbol{x}_j)$ from $(\boldsymbol{x}_{j+1},\ldots,\boldsymbol{x}_m)$, coupled only by α_j.

We truncate (6.28) after r_j terms and derive the new function

$$\boldsymbol{g}_j(\boldsymbol{x}_{j+1},\ldots,\boldsymbol{x}_m) := \left[\sqrt{\lambda_j(\alpha_j)}\psi_j(\alpha_j,\boldsymbol{x}_{j+1},\ldots,\boldsymbol{x}_m)\right]_{\alpha_j=1}^{r_j}.$$

For $\boldsymbol{t}_j := (s_{j+1},\ldots,s_m)$, we find that

$$\begin{aligned}
\|\boldsymbol{g}_j\|^2_{[\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)]^{r_j}} &= \sum_{\alpha_j=1}^{r_j}\lambda_j(\alpha_j)\|\psi_j(\alpha_j)\|^2_{\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)}\\
&= \sum_{\alpha_{j-1}=1}^{r_{j-1}}\sum_{\alpha_j=1}^{r_j}\lambda_j(\alpha_j)\|\varphi_j(\alpha_{j-1},\alpha_j)\otimes\psi_j(\alpha_j)\|^2_{L^2(\Omega_j)\otimes\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)}\\
&= \left\|\sum_{\alpha_{j-1}=1}^{r_{j-1}}\sum_{\alpha_j=1}^{r_j}\sqrt{\lambda_j(\alpha_j)}\varphi_j(\alpha_{j-1},\alpha_j)\otimes\psi_j(\alpha_j)\right\|^2_{L^2(\Omega_j)\otimes\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)}.
\end{aligned}$$

by exploiting the orthonormality of the vector-valued functions $[\varphi_j(\alpha_{j-1},\alpha_j)]_{\alpha_{j-1}=1}^{r_{j-1}}$, $\alpha_j = 1,\ldots,r_j$. Analogously to above, we infer that

$$\|\boldsymbol{g}_j\|_{[\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)]^{r_j}} \le \|\boldsymbol{g}_{j-1}\|_{[L^2(\Omega_j)\otimes\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)]^{r_{j-1}}},$$

hence, we conclude

$$\|\boldsymbol{g}_j\|_{[\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)]^{r_j}} \le \|f\|_{\mathbf{H}^{\boldsymbol{s}}(\boldsymbol{\Omega})} \quad \text{for all } j = 1,2,\ldots,m-1. \tag{6.29}$$

6.3. **Truncation ranks.** After having proven the regularity of the functions $\mathbf{g}_j$ for all $j = 1,\ldots,m-1$, we shall next determine the truncation ranks. To this end, consider a prescribed approximation accuracy $\varepsilon > 0$.

In the first step of the tensor train decomposition, i.e., for $j = 1$, we can immediately apply Theorem 5.2 to get essentially the decay $r_1^{-\beta_1}$ in the truncation error, where the rate β_1 is given by

$$\beta_1 = \left\{\frac{s_1}{n_1}\right\} + \min_{i=2}^{m}\left\{\frac{s_i}{n_i}\right\}. \tag{6.30}$$

Hence, we have to choose $r_1 := \varepsilon^{-1/\beta_1}$ to essentially get the truncation error ε in the first step.

In the j-th step of the tensor train decomposition, we set $\boldsymbol{y}_j := (\boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_m) \in \boldsymbol{\Upsilon}_j := \Omega_{j+1} \times \cdots \times \Omega_m$ and compute the kernel function $\kappa_j \in L^2(\boldsymbol{\Upsilon}_j, \boldsymbol{\Upsilon}_j)$ given by

$$\kappa_j(\boldsymbol{y}_j, \boldsymbol{y}'_j) := \sum_{\alpha_{j-1}=1}^{r_{j-1}} \lambda_{j-1}(\alpha_{j-1}) \int_{\Omega_j} \psi_{j-1}(\alpha_{j-1}, \boldsymbol{x}_j, \boldsymbol{y}_j)\psi_{j-1}(\alpha_{j-1}, \boldsymbol{x}_j, \boldsymbol{y}'_j) \,\mathrm{d}\boldsymbol{x}_j.$$

This gives rise to the spectral decomposition

$$\kappa_j(\boldsymbol{y}_j, \boldsymbol{y}'_j) = \sum_{\alpha_j=1}^{\infty} \lambda_j(\alpha_j)\psi_j(\alpha_j, \boldsymbol{y}_j)\psi_j(\alpha_j, \boldsymbol{y}'_j). \tag{6.31}$$

By setting

$$\varphi_j(\alpha_{j-1}, \boldsymbol{x}_j, \alpha_j) := \frac{\sqrt{\lambda_{j-1}(\alpha_{j-1})}}{\sqrt{\lambda_j(\alpha_j)}} \int_{\boldsymbol{\Upsilon}_j} \psi_{j-1}(\alpha_{j-1}, \boldsymbol{x}_j, \boldsymbol{y}_j)\psi_j(\alpha_j, \boldsymbol{y}_j) \,\mathrm{d}\boldsymbol{y}_j,$$

we arrive at the desired singular value decomposition (6.28).

From

$$\partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \partial_{\boldsymbol{y}'}^{\boldsymbol{\beta}} \kappa_j(\boldsymbol{y}_j, \boldsymbol{y}'_j) = \sum_{\alpha_{j-1}=1}^{r_{j-1}} \lambda_{j-1}(\alpha_{j-1}) \int_{\Omega_j} \partial_{\boldsymbol{y}}^{\boldsymbol{\alpha}} \psi_{j-1}(\alpha_{j-1}, \boldsymbol{x}_j, \boldsymbol{y}_j) \partial_{\boldsymbol{y}'}^{\boldsymbol{\beta}} \psi_{j-1}(\alpha_{j-1}, \boldsymbol{x}_j, \boldsymbol{y}'_j) \,\mathrm{d}\boldsymbol{x}_j$$

for any pair of multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_0^{m-j}$, we conclude

$$\begin{aligned} \|\kappa_j\|_{\mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j) \otimes \mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)} &\le \sum_{\alpha_{j-1}=1}^{r_{j-1}} \lambda_{j-1}(\alpha_{j-1}) \|\psi_{j-1}(\alpha_{j-1})\|^2_{L^2(\Omega_j) \otimes \mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)} \\ &= \|\boldsymbol{g}_{j-1}\|^2_{[L^2(\Omega_j) \otimes \mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)]^{r_{j-1}}}. \end{aligned}$$

Therefore, it holds $\kappa_j \in \mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j) \otimes \mathbf{H}^{\boldsymbol{t}_j}(\boldsymbol{\Upsilon}_j)$ which essentially implies the rate of convergence $r_j^{-2\beta_j - 1/2}$ with

$$\beta_j = \min_{i=j+1}^{m} \left\{ \frac{s_i}{n_i} \right\} - \frac{1}{4} \tag{6.32}$$

in the spectral decomposition (6.31). This in turn leads to the choice $r_j := \varepsilon^{-1/\beta_j}$ to essentially derive[3] the truncation error ε.

[3]The estimate $\sqrt{\sum_{j>r_j} \lambda_j^2} \lesssim r_j^{-2\beta_j - 1/2}$ implies $\lambda_j^2 \lesssim r_j^{-4\beta_j - 2}$ and hence $\lambda_j \lesssim r_j^{-2\beta_j - 1}$. This in turn gives $\sum_{j>r_j} \lambda_j \lesssim r_j^{-2\beta_j}$ and $\sqrt{\sum_{j>r_j} \lambda_j} \lesssim r_j^{-\beta_j}$, respectively.

6.4. **Complexity.** We summarize our findings for the continuous tensor train decomposition in the following theorem, in which we estimate the ranks required.

Theorem 6.1. *Let* $f \in \mathbf{H}^{\boldsymbol{s}}(\boldsymbol{\Omega})$ *for some fixed* $\boldsymbol{s} > \mathbf{0}$ *and* $0 < \varepsilon < 1$. *Choose the truncation ranks* $r_j := \varepsilon^{-1/\beta_j}$ *for* $j = 1, \ldots, m-1$, *where* β_j *is given in* (6.30) *for* $j = 1$ *and in* (6.32) *for* $j > 1$. *Then, the overall truncation error of the tensor train decomposition is essentially bounded by*

$$\|f - f^{\mathrm{TT}}_{r_1,\ldots,r_{m-1}}\|_{L^2(\boldsymbol{\Omega})} \lesssim \sqrt{m}\varepsilon.$$

The storage cost for $f^{\mathrm{TT}}_{r_1,\ldots,r_{m-1}}$ *are given by*

$$r_1 + \sum_{j=2}^{m-1} r_{j-1} r_j = \varepsilon^{-1/\beta_1} + \varepsilon^{-1/\beta_1 - 1/\beta_2} + \cdots + \varepsilon^{-1/\beta_{m-2} - 1/\beta_{m-1}} \tag{6.33}$$

and hence are bounded by $\mathcal{O}\big(m\varepsilon^{-2/\min_{j=1}^{m-1}\{\beta_j\}}\big)$.

Remark 6.2. *In case of* $n := n_1 = \cdots = n_m$ *and* $s := s_1 = \cdots = s_m$, *the cost of the tensor train decomposition is* $\mathcal{O}(m\varepsilon^{-2/\beta})$ *with* $\beta = \frac{s}{n} - \frac{1}{4}$. *Thus, the cost is essentially independent of the number* m *of subdomains.*

7. Concluding remarks

In this article, we have studied the convergence of the truncated singular value decomposition for functions from Sobolev spaces with dominating mixed smoothness. With these convergence results at hand, we proved that the particular ranks of the tensor train approximation are essentially independent of the overall dimension of the product domain. This is in contrast to the situation of approximating functions from the isotropic Sobolev spaces, where the maximum rank grows exponentially with the overall dimension, compare [15].

We also like also to comment on the relationship of the tensor train representation and deep neural networks (DNN). Indeed, it has been shown that one-dimensional basis functions, e.g. polynomials or wavelets, can be approximated up to an error ε by DNNs with ReLU activation functions at cost $\sim |\log \varepsilon|$. Moreover, the multiplication

$$(x, y) \mapsto x \cdot y = \frac{1}{4}\big\{(x+y)^2 - (x-y)^2\big\}$$

can easily be expressed by an additional layer and the $x \mapsto x^2$ activation function, which in turn can be approximated by a ReLU network with $\mathcal{O}(|\log \varepsilon|)$ layers, compare [37]. Vice versa, the tensor train representation corresponds to a network with m layers and the bivariate activation function $(x, y) \mapsto x \cdot y$, which itself can be represented by a multilayer

ReLU network as described above. This implies that upper bounds for the DNN complexity can easily be derived from our presented results.

We finally want to highlight that the major progress of DNNs relies on the use of compositions of nonlinear functions as a tool for approximation. In [26], the authors introduced a class of (m-variate) functions which can be expressed by the composition of bivariate functions. Furthermore, it was shown that the overall complexity of a DNN for such functions is bounded by the number of compositions (i.e. layers) times the approximation complexity of the bivariate functions. The tensor train representation is obviously contained in this class and the nonlinearity is simply the bilinear map. In the recent work [3], the authors have made the remarkable observation that the converse holds also true to a certain extent: The approximation rate of a corresponding tree based tensor network provides the same convergence rate as DNNs in the case of typical activation functions, as for example the ReLU activation function.

Acknowledgment. Michael Griebel was supported by the *Hausdorff Center for Mathematics* in Bonn, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy -EXC-2047/1-390685813 and the Sonderforschungsbereich 1060 *The Mathematics of Emergent Effects*.

References

[1] M. Ali and A. Nouy. Approximation with tensor networks. Part III: Multivariate approximation. arXiv:2101.11932 (2021).

[2] G. Allaire and M. Briane. Multiscale convergence and reiterated homogenization. *Proc. Roy. Soc. Edin.* **126A** (1996), 297–342.

[3] M. Bachmayr, A. Nouy, and R. Schneider. Approximation by tree tensor networks in high dimensions: Sobolev and compositional functions. arXiv:2112.01474 (2021).

[4] D. Bigoni, A. Engsig-Karup, and Y. Marzouk. Spectral tensor-train decomposition. *SIAM J. Sci. Comput.* **48**(4) (2016) A2405–A2439.

[5] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica* **13** (2004), 147–269.

[6] J. Barrett, D. Knezevic, and E. Süli. Kinetic models of dilute polymers. Analysis, approximation and computation. 11th School on Mathematical Theory in Fluid Mechanics 22–29 May 2009, Kacov, Czech Republic, Necas Center for Mathematical Modeling, Prague, 2009.

[7] D. Cioranescu, A. Damlamian, and G. Griso. The periodic unfolding method in homogenization. *SIAM J. Appl. Math.* **40** (2008), 1585–1620.

[8] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numerica* **6** (1997), 55–228.

[9] G.E. Fasshauer and Q. Ye. Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators. *Numer. Math.*, 119:585–611 (2011).

[10] R. Ghanem and P. Spanos. *Stochastic Finite Elements. A Spectral Approach.* Springer, New York, 1991.

[11] M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Math. Comput.* **82**(282) (2013), 975–994.

[12] M. Griebel and H. Harbrecht. A note on the construction of L-fold sparse tensor product spaces. *Constr. Approx.* **38**(2) (2013), 235–251.

[13] M. Griebel and H. Harbrecht. Approximation of bi-variate functions. Singular value decomposition versus sparse grids. *IMA J. Numer. Anal.* **34** (2014), 28–54.

[14] M. Griebel and H. Harbrecht. Singular value decomposition versus sparse grids. Refined complexity estimates. *IMA J. Numer. Anal.* **39** (2019), 1652–1671.

[15] M. Griebel and H. Harbrecht. Analysis of tensor approximation schemes for continuous functions. *Found. Comput. Math.* **23** (2023), 219–240.

[16] M. Griebel and D. Oeltz. A sparse grid space-time discretization scheme for parabolic problems. *Computing* **81** (2007), 1–34.

[17] M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numer. Math.* **83** (1999), 279–312.

[18] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus.* 2nd ed., Springer, Berlin-Heidelberg, 2019.

[19] W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *J. Fourier Anal. Appl.* **15**(5) (2009), 706–722.

[20] H. Harbrecht. A finite element method for elliptic problems with stochastic input data. *Appl. Numer. Math.* **60** (2010), 227–244.

[21] H. Harbrecht, R. Schneider, and C. Schwab. Sparse second moment analysis for elliptic problems in stochastic domains. *Numer. Math.* **109** (2008), 167–188.

[22] V. Hoang and C. Schwab. High-dimensional finite elements for elliptic problems with multiple scales. *SIAM Multiscale Model. Simul.* **3** (2005), 168–194.

[23] C. Le Bris and T. Lelièvre. Multiscale modelling of complex fluids: A mathematical initiation. In *Multiscale Modeling and Simulation in Science*, Lecture Notes in Computational Science and Engineering, vol. 66, Springer, Berlin, 2009, pp. 49–138.

[24] A. Lozinski, R. Owens, and T. Phillips. The Langevin and Fokker-Planck equations in polymer rheology. In R. Glowinski (ed.), *Handbook of Numerical Analysis XVI/XVII*, Elsevier North-Holland, 2011, pp. 211–303.

[25] B. Matérn. *Spatial variation*, volume 36 of *Lecture Notes in Statistics.* Springer, Berlin, second edition, 1986.

[26] H. Mhaskar and T. Poggio. Deep vs. shallow networks: An approximation theory perspective. *Anal. Appl.* **14**(6) (2016), 829–848.

[27] I. Oseledets. Tensor-train decomposition. *SIAM J. Sci. Comput.* **33**(5) (2011), 2295–2317.

[28] T. von Petersdorff and C. Schwab. Numerical solution of parabolic equations in high dimensions. *M2AN Math. Model. Numer. Anal.* **38**(1) (2004), 93–127.

[29] R. Schneider and A. Uschmajew. Approximation rates for the hierarchical tensor format in periodic Sobolev spaces. *J. Complexity* **30**(2) (2014), 56–71.

[30] C. Schwab and R.-A. Todor. Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.* **95** (2003), 707–734.

[31] V. Temlyakov. Approximation of functions with bounded mixed derivative. *Tr. Mat. Inst. Steklova* **178** (1986) (in Russian); *Proc. Steklov Inst. Math.* **1** (English translation).

[32] V. Temlyakov. Estimates of best bilinear approximations of periodic functions. *Tr. Mat. Inst. Steklova* **181** (1988), 250–267 (in Russian); *Proc. Steklov Inst. Math.* **4**, 181–293 (English translation).

[33] V. Temlyakov. Bilinear approximation and related questions, *Tr. Mat. Inst. Steklova* **194** (1992), 229–248 (in Russian); *Proc. Steklov Inst. Math.* **4**, 245–265 (English translation).

[34] H. Weyl. *Über die asymptotische Verteilung der Eigenwerte.* Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen (1911), 110–117.

[35] G. Widmer, R. Hiptmair and C. Schwab. Sparse adaptive finite elements for radiative transfer. *J. Comp. Phys.* **227** (2008), 6071–6105.

[36] J. Wloka. *Partial differential equations.* Cambridge University Press, Cambridge, 1987.

[37] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Netw.* **94** (2017), 103–114.

Michael Griebel, Institut für Numerische Simulation, Universität Bonn, Friedrich-Hirzebruch-Allee 7, 53115 Bonn, Germany and Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany

Email address: `griebel@ins.uni-bonn.de`

Helmut Harbrecht, Departement Mathematik und Informatik, Universität Basel, Spiegelgasse 1, 4051 Basel, Switzerland

Email address: `helmut.harbrecht@unibas.ch`

Reinhold Schneider, Institute of Mathematics, Technical University of Berlin, Strasse des 17. Juni 136, 10623 Berlin, Germany

Email address: `schneidr@math.tu-berlin.de`