Analysis of tensor approximation schemes for continuous functions

Michael Griebel · Helmut Harbrecht

Communicated by Wolfgang Dahmen

Abstract In this article, we analyze tensor approximation schemes for continuous functions. We assume that the function to be approximated lies in an isotropic Sobolev space and discuss the cost when approximating this function in the continuous analogue of the Tucker tensor format or of the tensor train format. We especially show that the cost of both approximations are dimension-robust when the Sobolev space under consideration provides appropriate dimension weights.

Keywords Tensor format, approximation error, rank complexity, Sobolev space with dimension weights

Mathematics Subject Classification (2020) 41A46, 41A63, 46E35

1 Introduction

The efficient approximate representation of multivariate functions is an important task in numerical analysis and scientific computing. In this article, we hence consider the approximation of functions which live on the product of m bounded domains $\Omega_1 \times \cdots \times \Omega_m$, each of which satisfies $\Omega_j \subset \mathbb{R}^{n_j}$. Besides a sparse grid approximation of the function under consideration, being discussed in, e.g., [8, 17, 18, 50], one can also apply a low-rank approximation by means of a tensor approximation scheme, see, e.g., [15, 21, 22, 34, 35] and the references therein.

The low-rank approximation in the situation of the product of $m = 2$ domains is well understood. It is related to the *singular value decomposition* and has been studied for arbitrary product domains in, e.g., [19, 20], see also [46–48] for the

Michael Griebel
Institut für Numerische Simulation, Universität Bonn, Endenicher Allee 19b, 53115 Bonn, and Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany
E-mail: griebel@ins.uni-bonn.de

Helmut Harbrecht
Departement Mathematik und Informatik, Universität Basel, Spiegelgasse 1, 4051 Basel, Switzerland
E-mail: helmut.harbrecht@unibas.ch

periodic case. However, the situation is not that clear for the product of $m > 2$ domains, where one ends up with *tensor decompositions*. Such tensor decompositions are generalizations of the well known singular value decomposition and the corresponding low-rank matrix approximation methods of two dimensions to the higher-dimensional setting. There, besides the curse of dimension, we encounter – due to the non-existence of an Eckart-Young-Mirsky theorem – that the concepts of singular value decomposition and low-rank approximation can be generalized to higher dimensions in more than one way. Consequently, there exist many generalizations of the singular value decomposition of a function and of low-rank approximations to tensors. To this end, various schemes have been developed over the years in different areas of the sciences and have successfully been applied to various high-dimensional problems ranging from quantum mechanics and physics via biology and econometrics, computer graphics and signal processing to numerical analysis. Recently, tensor methods have even been recognized as special deep neural networks in machine learning and big data analysis [11,28]. As tensor approximation schemes, we have, for example, matrix product states, DMRG, MERA, PEPS, CP, CANDECOMP, PARAFAC, Tucker, tensor train, tree tensor networks and hierarchical Tucker, to name a few. A mathematical introduction into tensor methods is given in the seminal book [21], while a survey on existing methods and their literature can be found in [16]. Also various software packages have been developed for an algebra of operators dealing with tensors.

Tensor methods are usually analyzed as low-rank approximations to a full *discrete tensor* of data with respect to the ℓ_2-norm or Frobenius-norm. In this respective, they can be seen as compression methods which may avoid the curse of dimensionality, which is inherent in the full tensor representation. The main tool for studying tensor compression schemes is the so-called *tensor-rank*, compare [12, 13,21]. Thus, instead of $\mathcal{O}(N^n)$ storage, as less as $\mathcal{O}(nNr^3)$ or even only $\mathcal{O}(nNr^2)$ storage is needed, where N denotes the number of data points in one coordinate direction, n denotes the dimension of the tensor under consideration and r denotes the respective tensor rank of the data. The cost complexity of the various algorithms working with sparse tensor representations is correspondingly reduced and working in a sparse tensor format allows to alleviate or to completely break the curse of dimension for suitable tensor data classes, i.e., for sufficiently small r.

However, the question where the tensor data stem from and the issue of the accuracy of the full tensor approximation, i.e., the discretization error of the full tensor itself and its relation to the error of a subsequent low-rank tensor approximation, is usually not adequately addressed.[1] Instead, only the approximation property of a low-rank tensor scheme with respect to the full tensor data is considered. But the former question is important since it clearly makes no sense to derive a tensor approximation with an error that is substantially smaller than the error which is already inherent in the full tensor data due to some discretization process for a continuous high-dimensional function which stems from some certain function class.

The approximation rates to *continuous functions* can be determined by a recursive use of the singular value decomposition, which is successively applied to convert the function into a specific continuous tensor format. We studied the singular value decomposition for arbitrary domains in [19,20] and we now can apply

---

[1] We are only aware of [2,3,5,39], where this question has been considered so far.

these results to discuss approximation rates of continuous tensor formats. In the present article, given a function $f \in H^k(\Omega_1 \times \cdots \times \Omega_m)$, we study the continuous analogues of the Tucker tensor decomposition and of the tensor train decomposition. We give bounds on the ranks required to ensure that the tensor decomposition admits a prescribed target accuracy. Especially, our analysis takes into account the influence of errors induced by truncating infinite expansions to finite ones. We therefore study an *algorithm* that computes the desired tensor expansion which is in contrast to the question of the smallest tensor-rank. We finally show that (isotropic) Sobolev spaces *with dimension weights* help to beat the curse of dimension when the number m of product domains tends to infinity.

Besides the simple situation of $\Omega_1 = \cdots = \Omega_m = [0,1]$, which is usually considered in case of tensor decompositions, there are many more applications of our general setting. For example, non-Newtonian flow can be modeled 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 $3(\ell-1)$-dimensional configuration space $\Omega_2 \times \cdots \times \Omega_\ell$, consisting of $\ell-1$ spheres. Here, ℓ denotes the number of atoms in a chain-like molecule which constitutes the non-Newtonian behavior of the flow, for details see [4,29,31,37]. Another example is homogenization. After unfolding [10], a two-scale homogenization problem gives raise to the product of the macroscopic physical domain and the periodic microscopic domain of the cell problem, see [32]. For multiple scales, several periodic microscopic domains appear which reflect the different separable scales, see e.g. [27]. Also the m-th moment of linear elliptic boundary value problems with random source terms, i.e. $Au(\omega) = f(\omega)$ in Ω, are known to satisfy a deterministic partial differential equation on the m-fold product domain $\Omega \times \cdots \times \Omega$. There, the solution's m-th moment $\mathcal{M}_u$ is given by the equation

$$(A \otimes \cdots \otimes A)\mathcal{M}_u = \mathcal{M}_f \text{ in } \Omega \times \cdots \times \Omega,$$

see [40,41]. This approach extends to boundary value problems with random diffusion and to random domains as well [9,25]. Moreover, we find the product of several domains in quantum mechanics for e.g. the Schrödinger equation or the Langevin equation, where each domain is three-dimensional and corresponds to a single particle. Finally, we encounter it in uncertainty quantification, where one has the product of the physical domain Ω_1 and of in general infinitely many intervals $\Omega_2 = \Omega_3 = \Omega_4 = \ldots$ for the random input parameter, which reflects its series expansion by the Karhunen-Lòeve decomposition or the Lévy-Ciesielski decomposition.

The remainder of this article is organized as follows: In Section 2, we give a short introduction to our results on the singular value decomposition, which are needed to derive the estimates for the continuous tensor decompositions. Then, in Section 3, we study the continuous Tucker tensor format, computed by means of the higher-oder singular value decomposition. Next, we study the continuous tensor train decomposition in Section 4, computed by means of a repeated use of a vector-valued singular value decomposition. Finally, Section 5 concludes with some final remarks.

Throughout this article, 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 which C and D may depend on. Obviously, $C \gtrsim D$ is defined as

$D \lesssim C$, and $C \sim D$ as $C \lesssim D$ and $C \gtrsim D$. Moreover, given a Lipschitz-smooth domain $\Omega \subset \mathbb{R}^n$, $L^2(\Omega)$ means the space of square integrable functions on Ω. For real numbers $k \geq 0$, the associated Sobolev space is denoted by $H^k(\Omega)$, where its norm $\|\cdot\|_{H^k(\Omega)}$ is defined in the standard way, compare [33,45]. As usual, we have $H^0(\Omega) = L^2(\Omega)$. The seminorm in $H^k(\Omega)$ is denoted by $|\cdot|_{H^k(\Omega_1)}$. Although not explicitly written, our subsequent analysis covers also the situation of Ω being not a domain but a (smooth) manifold.

2 Singular value decomposition

2.1 Definition and calculation

Let $\Omega_1 \subset \mathbb{R}^{n_1}$ and Ω_2 be Lipschitz-smooth domains. To represent functions $f \in L^2(\Omega_1 \times \Omega_2)$ on the tensor product domain $\Omega_1 \times \Omega_2$ in an efficient way, we will consider low-rank approximations which separate the variables $\boldsymbol{x} \in \Omega_1$ and $\boldsymbol{y} \in \Omega_2$ in accordance with

$$f(\boldsymbol{x}, \boldsymbol{y}) \approx f_r(\boldsymbol{x}, \boldsymbol{y}) := \sum_{\alpha=1}^r \sqrt{\lambda(\alpha)} \varphi(\boldsymbol{x}, \alpha) \psi(\alpha, \boldsymbol{y}). \tag{2.1}$$

It is well known (see e.g. [30,38,42]) that the best possible representation (2.1) in the L^2-sense is given by the singular value decomposition, also called Karhunen-Lòeve expansion.[2] Then, the coefficients $\sqrt{\lambda(\alpha)} \in \mathbb{R}$ are the singular values and the $\varphi(\alpha) \in L^2(\Omega_1)$ and $\psi(\alpha) \in L^2(\Omega_2)$ are the left and right (L^2-normalized) eigenfunctions of the integral operator

$$\mathcal{S}_f : L^2(\Omega_1) \to L^2(\Omega_2), \quad u \mapsto (\mathcal{S}_f u)(\boldsymbol{y}) := \int_{\Omega_1} f(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{x}) \, d\boldsymbol{x}.$$

This means that

$$\sqrt{\lambda(\alpha)} \psi(\alpha, \boldsymbol{y}) = \big(\mathcal{S}_f \varphi(\alpha)\big)(\boldsymbol{y}) \quad \text{and} \quad \sqrt{\lambda(\alpha)} \varphi(\boldsymbol{x}, \alpha) = \big(\mathcal{S}_f^\star \psi(\alpha)\big)(\boldsymbol{x}), \tag{2.2}$$

where

$$\mathcal{S}_f^\star : L^2(\Omega_2) \to L^2(\Omega_1), \quad v \mapsto (\mathcal{S}_f^\star v)(\boldsymbol{x}) := \int_{\Omega_2} f(\boldsymbol{x}, \boldsymbol{y}) v(\boldsymbol{y}) \, d\boldsymbol{y}.$$

is the adjoint of $\mathcal{S}_f$. Especially, the left and right eigenfunctions $\{\varphi(\alpha)\}_{\alpha=1}^\infty$ and $\{\psi(\alpha)\}_{\alpha=1}^\infty$ form orthonormal bases in $L^2(\Omega_1)$ and $L^2(\Omega_2)$, respectively.

In order to compute the singular value decomposition, we need to solve the eigenvalue problem

$$\mathcal{K}_f \varphi(\alpha) = \lambda(\alpha) \varphi(\alpha)$$

for the integral operator

$$\mathcal{K}_f = \mathcal{S}_f^\star \mathcal{S}_f : L^2(\Omega_1) \to L^2(\Omega_1), \quad u \mapsto (\mathcal{K}_f u)(\boldsymbol{x}) := \int_{\Omega_1} k_f(\boldsymbol{x}, \boldsymbol{x}') u(\boldsymbol{x}') \, d\boldsymbol{x}'. \tag{2.3}$$

---

[2] We refer the reader to [44] for a comprehensive historical overview on the singular value decomposition.

Since $f \in L^2(\Omega_1 \times \Omega_2)$, the kernel

$$k_f(\boldsymbol{x}, \boldsymbol{x}') = \int_{\Omega_2} f(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{x}', \boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \in L^2(\Omega_1 \times \Omega_1) \tag{2.4}$$

is a symmetric Hilbert-Schmidt kernel. Hence, there exist countably many eigenvalues

$$\lambda(1) \geq \lambda(2) \geq \cdots \geq \lambda(m) \stackrel{m \to \infty}{\longrightarrow} 0$$

and the associated eigenfunctions $\{\varphi(\alpha)\}_{\alpha \in \mathbb{N}}$ constitute an orthonormal basis in $L^2(\Omega_1)$.

Likewise, to obtain an orthonormal basis of $L^2(\Omega_2)$, we can solve the eigenvalue problem

$$\widetilde{\mathcal{K}}_f \psi(\alpha) = \widetilde{\lambda}(\alpha) \psi(\alpha)$$

for the integral operator

$$\widetilde{\mathcal{K}}_f = \mathcal{S}_f \mathcal{S}_f^\star : L^2(\Omega_2) \to L^2(\Omega_2), \quad u \mapsto (\widetilde{\mathcal{K}}_f u)(\boldsymbol{y}) := \int_{\Omega_2} \widetilde{k}_f(\boldsymbol{y}, \boldsymbol{y}') u(\boldsymbol{y}') \, \mathrm{d}\boldsymbol{y}'$$

with symmetric Hilbert-Schmidt kernel

$$\widetilde{k}_f(\boldsymbol{y}, \boldsymbol{y}') = \int_{\Omega_1} f(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{x}, \boldsymbol{y}') \, \mathrm{d}\boldsymbol{x} \in L^2(\Omega_2 \times \Omega_2). \tag{2.5}$$

It holds $\lambda(\alpha) = \widetilde{\lambda}(\alpha)$ and the sequences $\{\varphi(\alpha)\}$ and $\{\psi(\alpha)\}$ are related by (2.2).

2.2 Regularity of the eigenfunctions

Now, we consider functions $f \in H^k(\Omega_1 \times \Omega_2)$. In the following, we collect results from [19,20] concerning the singular value decomposition of such functions. We repeat the proof whenever needed for having explicit constants. To this end, we define the mixed Sobolev space $H_{\mathrm{mix}}^{k,\ell}(\Omega_1 \times \Omega_2)$ as a tensor product of Hilbert spaces

$$H_{\mathrm{mix}}^{k,\ell}(\Omega_1 \times \Omega_2) := H^k(\Omega_1) \otimes H^\ell(\Omega_2),$$

which we equip with the usual cross norm

$$\|f\|_{H_{\mathrm{mix}}^{k,\ell}(\Omega_1 \times \Omega_2)} := \sqrt{\sum_{|\boldsymbol{\alpha}| \leq k} \sum_{|\boldsymbol{\beta}| \leq \ell} \left\| \frac{\partial^{|\boldsymbol{\alpha}|} \partial^{|\boldsymbol{\beta}|}}{\partial \mathbf{x}^{\boldsymbol{\alpha}} \partial \mathbf{y}^{\boldsymbol{\beta}}} f \right\|_{L^2(\Omega_1 \times \Omega_2)}^2}.$$

Note that

$$H^k(\Omega_1 \times \Omega_2) \subset H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2), \quad H^k(\Omega_1 \times \Omega_2) \subset H_{\mathrm{mix}}^{0,k}(\Omega_1 \times \Omega_2).$$

Lemma 1 *Assume that $f \in H^k(\Omega_1 \times \Omega_2)$ for some fixed $k \geq 0$. Then, the operators*

$$\mathcal{S}_f : L^2(\Omega_1) \to H^k(\Omega_2), \quad \mathcal{S}_f^\star : L^2(\Omega_2) \to H^k(\Omega_1)$$

are continuous with

$$\left\| \mathcal{S}_f \right\|_{L^2(\Omega_1) \to H^k(\Omega_2)} \leq \|f\|_{H_{\mathrm{mix}}^{0,k}(\Omega_1 \times \Omega_2)}, \quad \left\| \mathcal{S}_f^\star \right\|_{L^2(\Omega_2) \to H^k(\Omega_1)} \leq \|f\|_{H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)}.$$

Proof From $H^k(\Omega_1 \times \Omega_2) \subset H^{0,k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)$ it follows for $f \in H^k(\Omega_1 \times \Omega_2)$ that $f \in H^{0,k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)$. Therefore, the operator $\mathcal{S}_f : L^2(\Omega_1) \to H^k(\Omega_2)$ is continuous since

$$
\begin{aligned}
\|\mathcal{S}_f u\|_{H^k(\Omega_2)} &= \sup_{\|v\|_{H^{-k}(\Omega_2)}=1} (\mathcal{S}_f u, v)_{L_2(\Omega_2)} \\
&= \sup_{\|v\|_{H^{-k}(\Omega_2)}=1} (f, u \otimes v)_{L^2(\Omega_1 \times \Omega_2)} \\
&\leq \sup_{\|v\|_{H^{-k}(\Omega_2)}=1} \|f\|_{H^{0,k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)} \|u \otimes v\|_{H^{0,-k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)} \\
&\leq \|f\|_{H^{0,k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)} \|u\|_{L^2(\Omega_1)}.
\end{aligned}
$$

Note that we have used here $H^{0,-k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2) = L^2(\Omega_1) \otimes H^{-k}(\Omega_2)$. Proceeding likewise for $\mathcal{S}_f^\star : L^2(\Omega_2) \to H^k(\Omega_1)$ completes the proof.

Lemma 2 *Assume that* $f \in H^k(\Omega_1 \times \Omega_2)$ *for some fixed* $k \geq 0$. *Then, it holds* $\mathcal{S}_f \varphi(\alpha) \in H^k(\Omega_2)$ *and* $\mathcal{S}_f^\star \psi(\alpha) \in H^k(\Omega_1)$ *for all* $\alpha \in \mathbb{N}$ *with*

$$
\|\varphi(\alpha)\|_{H^k(\Omega_1)} \leq \frac{1}{\sqrt{\lambda(\alpha)}} \|f\|_{H^{k,0}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)},
$$

$$
\|\psi(\alpha)\|_{H^k(\Omega_2)} \leq \frac{1}{\sqrt{\lambda(\alpha)}} \|f\|_{H^{0,k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)}.
$$

Proof According to (2.2) and Lemma 1, we have

$$
\|\varphi(\alpha)\|_{H^k(\Omega_1)} = \frac{1}{\sqrt{\lambda(\alpha)}} \|\mathcal{S}_f^\star \psi(\alpha)\|_{H^k(\Omega_1)} \leq \frac{1}{\sqrt{\lambda(\alpha)}} \|f\|_{H^{k,0}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)} \|\psi(\alpha)\|_{L^2(\Omega_2)}.
$$

This proves the first assertion. The second assertion follows by duality.

As an immediate consequence of Lemma 2, we obtain

$$
\sum_{\alpha=1}^{r} \lambda(\alpha) \|\varphi(\alpha)\|^2_{H^k(\Omega_1)} \leq r \|f\|^2_{H^{k,0}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)}
$$

and

$$
\sum_{\alpha=1}^{r} \lambda(\alpha) \|\psi(\alpha)\|^2_{H^k(\Omega_2)} \leq r \|f\|^2_{H^{0,k}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)}.
$$

We will show later in Lemma 4 how to improve this estimate by sacrificing some regularity.

2.3 Truncation error

We next give estimates on the decay rate of the eigenvalues of the integral operator $\mathcal{K}_f = \mathcal{S}_f^\star \mathcal{S}_f$ with kernel (2.4). To this end, we exploit the smoothness in the function's first variable and assume hence $f \in H^{k,0}_{\mathrm{mix}}(\Omega_1 \times \Omega_2)$. We introduce finite element spaces $U_r \subset L^2(\Omega_1)$, which consist of r discontinuous, piecewise polynomial functions of total degree $\lceil k \rceil$ on a quasi-uniform triangulation of Ω_1 with

mesh width $h_r \sim r^{-1/n_1}$. Then, given a function $w \in H^k(\Omega_1)$, the L^2-orthogonal projection $P_r : L^2(\Omega_1) \to U_r$ satisfies

$$\|(I - P_r)w\|_{L^2(\Omega_1)} \leq c_k r^{-k/n_1} |w|_{H^k(\Omega_1)} \tag{2.6}$$

uniformly in r due to the Bramble-Hilbert lemma, see e.g., [6,7].

For the approximation of $f(\boldsymbol{x}, \boldsymbol{y})$ in the first variable, i.e. $\big((P_r \otimes I)f\big)(\boldsymbol{x}, \boldsymbol{y})$, we obtain the following approximation result for the present choice of U_r, see [20] for the proof.

Lemma 3 *Assume that $f \in H^k(\Omega_1 \times \Omega_2)$ for some fixed $k \geq 0$. Let $\lambda(1) \geq \lambda(2) \geq \ldots \geq 0$ be the eigenvalues of the operator $\mathcal{K}_f = \mathcal{S}_f^\star \mathcal{S}_f$ and $\lambda_r(1) \geq \lambda_r(2) \geq \ldots \geq \lambda_r(r) \geq 0$ those of $\mathcal{K}_f^r := P_r \mathcal{K}_f P_r$. Then, it holds*

$$\|f - (P_r \otimes I)f\|_{L^2(\Omega_1 \times \Omega_2)}^2 = \operatorname{trace} \mathcal{K}_f - \operatorname{trace} \mathcal{K}_f^r = \sum_{\alpha=1}^r \big(\lambda(\alpha) - \lambda_r(\alpha)\big) + \sum_{\alpha=r+1}^\infty \lambda(\alpha).$$

By combining this lemma with the approximation estimate (2.6) and in view of $\lambda(\alpha) - \lambda_r(\alpha) \geq 0$ for all $\alpha \in \{1, 2, \ldots, r\}$ according to the min-max theorem of Courant-Fischer, see [1] for example, we conclude that the truncation error of the singular value decomposition can be bounded by

$$\left\| f - \sum_{\alpha=1}^r \sqrt{\lambda(\alpha)}\big(\varphi(\alpha) \otimes \psi(\alpha)\big) \right\|_{L^2(\Omega_1 \times \Omega_2)} = \sqrt{\sum_{\alpha=r+1}^\infty \lambda(\alpha)} \leq c_k r^{-\frac{k}{n_1}} |f|_{H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)}.$$

Since the eigenvalues of the integral operator $\mathcal{K}_f$ and its adjoint $\widetilde{\mathcal{K}}_f$ are the same, we can also exploit the smoothness of f in the second coordinate by interchanging the roles of Ω_1 and Ω_2 in the above considerations. We thus obtain the following theorem:

Theorem 1 *Let $f \in H^k(\Omega_1 \times \Omega_2)$ for some fixed $k \geq 0$ and let*

$$f_r^{\mathrm{SVD}} = \sum_{\alpha=1}^r \sqrt{\lambda(\alpha)}\big(\varphi(\alpha) \otimes \psi(\alpha)\big).$$

Then, it holds

$$\|f - f_r^{\mathrm{SVD}}\|_{L^2(\Omega_1 \times \Omega_2)} = \sqrt{\sum_{\alpha=r+1}^\infty \lambda(\alpha)} \leq c_k r^{-\frac{k}{\min\{n_1, n_2\}}} |f|_{H^k(\Omega_1 \times \Omega_2)}. \tag{2.7}$$

Remark 1 Theorem 1 implies that the eigenvalues $\{\lambda(\alpha)\}_{\alpha \in \mathbb{N}}$ in case of a function $f \in H^k(\Omega_1 \times \Omega_2)$ decay like

$$\lambda(\alpha) \lesssim \alpha^{-\frac{2k}{\min\{n_1, n_2\}} - 1} \quad \text{as } \alpha \to \infty. \tag{2.8}$$

Having the decay rate of the eigenvalues at hand, we are able to improve the result of Lemma 2 by sacrificing some regularity. Note that the proof of this result is based upon an argument from [43].

Lemma 4 *Assume that $f \in H^{k+\min\{n_1,n_2\}}(\Omega_1 \times \Omega_2)$ for some fixed $k \geq 0$. Then, it holds*

$$\sum_{\alpha=1}^{\infty} \lambda(\alpha) \|\varphi(\alpha)\|_{H^k(\Omega_1)}^2 = \|f\|_{H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)}^2$$

and

$$\sum_{\alpha=1}^{\infty} \lambda(\alpha) \|\psi(\alpha)\|_{H^k(\Omega_2)}^2 = \|f\|_{H_{\mathrm{mix}}^{0,k}(\Omega_1 \times \Omega_2)}^2.$$

Proof Without loss of generality, we assume $n_1 \leq n_2$. Then, since $f \in H^{k+n_1}(\Omega_1 \times \Omega_2)$, we conclude from (2.8) that

$$\lambda(\alpha) \lesssim \alpha^{-\frac{2(k+n_1)}{n_1}-1} \quad \text{as } \alpha \to \infty,$$

where we used that $n_1 = \min\{n_1, n_2\}$. Moreover, by interpolating between $L^2(\Omega_1)$ and $H^{k+n_1}(\Omega_1)$, compare [33,45] for example, we find

$$\|\varphi(\alpha)\|_{H^k(\Omega_1)}^2 \lesssim \lambda(\alpha)^{-\frac{k}{k+n_1}},$$

that is

$$\lambda(\alpha) \|\varphi(\alpha)\|_{H^k(\Omega_1)}^2 \lesssim \lambda(\alpha)^{\frac{n_1}{k+n_1}}.$$

As a consequence, we infer that

$$\lambda(\alpha) \|\varphi(\alpha)\|_{H^k(\Omega_1)}^2 \lesssim \alpha^{-\left(\frac{2(k+n_1)}{n_1}+1\right) \cdot \frac{n_1}{k+n_1}} = \alpha^{-(2+\delta)}$$

with $\delta = \frac{n_1}{k+n_1} > 0$. Therefore, it holds

$$\sum_{\alpha=1}^{\infty} \alpha^{(1+\delta')} \lambda(\alpha) \|\varphi(\alpha)\|_{H^k(\Omega_1)}^2 < \infty$$

for any $\delta' \in (0, \delta)$. Hence, the series

$$A(\boldsymbol{x}) := \sum_{\alpha=1}^{\infty} \alpha^{(1+\delta')} \lambda(\alpha) |\partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} \varphi(\alpha, \boldsymbol{x})|^2$$

converges for almost all $\boldsymbol{x} \in \Omega_1$, provided that $|\boldsymbol{\beta}| \leq k$. Likewise, the series

$$B(\boldsymbol{y}) := \sum_{\alpha=1}^{\infty} \alpha^{-(1+\delta')} |\psi(\alpha, \boldsymbol{y})|^2$$

converges for almost all $\boldsymbol{y} \in \Omega_2$. Thus, the series

$$\sum_{\alpha=1}^{\infty} \sqrt{\lambda(\alpha)} |\partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} \varphi(\alpha, \boldsymbol{x})| |\psi(\alpha, \boldsymbol{y})| \leq \sqrt{A(\boldsymbol{x})} \sqrt{B(\boldsymbol{y})}$$

converges for almost all $\boldsymbol{x} \in \Omega_1$ and $\boldsymbol{y} \in \Omega_2$, provided that $|\boldsymbol{\beta}| \leq k$. Because of Egorov's theorem, the pointwise absolute convergence almost everywhere implies uniform convergence. Hence, we can switch differentiation and summation to get

$$
\begin{aligned}
\|f\|_{H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)}^2 &= \sum_{|\boldsymbol{\beta}| \leq k} \left\| \partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} \sum_{\alpha=1}^{\infty} \sqrt{\lambda(\alpha)} \big(\varphi(\alpha) \otimes \psi(\alpha) \big) \right\|_{L^2(\Omega_1 \times \Omega_2)}^2 \\
&= \sum_{|\boldsymbol{\beta}| \leq k} \left\| \sum_{\alpha=1}^{\infty} \sqrt{\lambda(\alpha)} \big(\partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} \varphi(\alpha) \otimes \psi(\alpha) \big) \right\|_{L^2(\Omega_1 \times \Omega_2)}^2 .
\end{aligned}
$$

Finally, we exploit the product structure of $L^2(\Omega_1 \times \Omega_2)$ and the orthonormality of $\{\psi(\alpha)\}_{\alpha \in \mathbb{N}}$ to derive the first assertion, i.e.,

$$
\begin{aligned}
\|f\|_{H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)}^2 &= \sum_{\alpha=1}^{\infty} \lambda(\alpha) \sum_{|\boldsymbol{\beta}| \leq k} \big\| \partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} \varphi(\alpha) \big\|_{L^2(\Omega_1)}^2 \|\psi(\alpha)\|_{L^2(\Omega_2)}^2 \\
&= \sum_{\alpha=1}^{\infty} \lambda(\alpha) \|\varphi(\alpha)\|_{H^k(\Omega_1)}^2 .
\end{aligned}
$$

The second assertion follows in complete analogy.

2.4 Vector-valued functions

In addition to the aforementioned results, we will also need the following result which concerns the approximation of vector-valued functions. Here and in the sequel, the vector-valued function $\boldsymbol{w} = [w(\alpha)]_{\alpha=1}^m$ is an element of $[L^2(\Omega)]^m$ and $[H^k(\Omega)]^m$ for some domain $\Omega \subset \mathbb{R}^n$, respectively, if the norms

$$
\|\boldsymbol{w}\|_{[L^2(\Omega)]^m} = \sqrt{\sum_{\alpha=1}^m \|w(\alpha)\|_{L^2(\Omega)}^2}, \quad \|\boldsymbol{w}\|_{[H^k(\Omega)]^m} = \sqrt{\sum_{\alpha=1}^m \|w(\alpha)\|_{H^k(\Omega)}^2}
$$

are finite. Likewise, the seminorm is defined in $[H^k(\Omega)]^m$.

Consider now a vector-valued function $\boldsymbol{w} \in [H^k(\Omega_1)]^m$ of dimension m. Then, instead of (2.6), we find

$$
\|(I - P_r)\boldsymbol{w}\|_{[L^2(\Omega_1)]^m} \leq c_k \left(\frac{r}{m} \right)^{-k/n_1} |\boldsymbol{w}|_{[H^k(\Omega_1)]^m},
$$

since $\boldsymbol{w}$ consists of m components and we thus need m-times as many ansatz functions for our approximation argument. Hence, in case of a vector-valued function $\boldsymbol{f} \in [H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)]^m \simeq [H^k(\Omega_1)]^m \otimes L^2(\Omega_2)$, we conclude by exploiting the smoothness in the first variable[3] that the truncation error of the singular value decomposition can be estimated by

$$
\left\| \boldsymbol{f} - \sum_{\alpha=1}^r \sqrt{\lambda(\alpha)} \big(\boldsymbol{\varphi}(\alpha) \otimes \psi(\alpha) \big) \right\|_{[L^2(\Omega_1 \times \Omega_2)]^m} \leq c_k \left(\frac{r}{m} \right)^{-k/n_1} |\boldsymbol{f}|_{[H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)]^m} . \quad (2.9)
$$

---

[3] Note that the kernel function of $\mathcal{S}_f^\star \mathcal{S}_f$ is matrix-valued while the kernel function of $\mathcal{S}_f \mathcal{S}_f^\star$ is scalar-valued.

Hence, the decay rate of the singular values is considerably reduced. Finally, we like to remark that Lemma 4 holds also in the vector case, i.e.,

$$\sum_{\alpha=1}^{\infty} \lambda(\alpha) \|\boldsymbol{\varphi}(\alpha)\|_{[H^k(\Omega_1)]^m}^2 = \|\boldsymbol{f}\|_{[H_{\mathrm{mix}}^{k,0}(\Omega_1 \times \Omega_2)]^m}^2$$

and

$$\sum_{\alpha=1}^{\infty} \lambda(\alpha) \|\psi(\alpha)\|_{H^k(\Omega_2)}^2 = \|\boldsymbol{f}\|_{[H_{\mathrm{mix}}^{0,k}(\Omega_1 \times \Omega_2)]^m}^2, \tag{2.10}$$

provided that $\boldsymbol{f}$ has extra regularity in terms of $\boldsymbol{f} \in [H^{k+n_1}(\Omega_1 \times \Omega_2)]^m$. Here, analogously to above, $[H_{\mathrm{mix}}^{0,k}(\Omega_1 \times \Omega_2)]^m \simeq [L^2(\Omega_1)]^m \otimes H^k(\Omega_2)$.

After these preparations, we now introduce and analyze two types of continuous analogues of tensor formats, namely of the Tucker format [26, 49] and of the tensor train format [36, 34], and discuss their approximation properties for functions $f \in H^k(\Omega_1 \times \cdots \times \Omega_m)$.

3 Tucker tensor format

3.1 Tucker decompostion

We shall consider from now on a product domain which consists of m different domains $\Omega_j \subset \mathbb{R}^{n_j}$, $j = 1, \ldots, m$. For given $f \in L^2(\Omega_1 \times \cdots \times \Omega_m)$ and $j \in \{1, 2, \ldots, m\}$, we apply the singular value decomposition to separate the variables $\boldsymbol{x}_j \in \Omega_j$ and $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_m) \in \Omega_1 \times \cdots \times \Omega_{j-1} \times \Omega_{j+1} \times \cdots \times \Omega_m$. We hence get

$$\begin{aligned}
&f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_j, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_m) \\
&= \sum_{\alpha_j=1}^{\infty} \sqrt{\lambda_j(\alpha_j)} \varphi_j(\boldsymbol{x}_j, \alpha_j) \psi_j(\alpha_j, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_m),
\end{aligned} \tag{3.11}$$

where the left eigenfunctions $\{\varphi_j(\alpha_j)\}_{\alpha_j \in \mathbb{N}}$ form an orthonormal basis in $L^2(\Omega_j)$. Consequently, if we iterate over all $j \in \{1, 2, \ldots, m\}$, this yields an orthonormal basis $\{\varphi_1(\alpha_1) \otimes \cdots \otimes \varphi_m(\alpha_m)\}_{\boldsymbol{\alpha} \in \mathbb{N}^m}$ of $L^2(\Omega_1 \times \cdots \times \Omega_m)$, and we arrive at the representation

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \sum_{|\boldsymbol{\alpha}|=1}^{\infty} \omega(\boldsymbol{\alpha}) \varphi_1(\alpha_1, \boldsymbol{x}_1) \cdots \varphi_m(\alpha_m, \boldsymbol{x}_m). \tag{3.12}$$

Herein, the tensor $\left[\omega(\boldsymbol{\alpha})\right]_{\boldsymbol{\alpha} \in \mathbb{N}^m}$ is the *core tensor*, where a single coefficient is given by

$$\omega(\alpha_1, \ldots, \alpha_m) = \int_{\Omega_1 \times \cdots \times \Omega_m} f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \varphi_1(\alpha_1, \boldsymbol{x}_1) \cdots \varphi_m(\alpha_m, \boldsymbol{x}_m) \, \mathrm{d}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m).$$

3.2 Truncation error

If we intend to truncate the singular value decomposition (3.11) after r_j terms such that the truncation error is bounded by ε, we have to choose

$$\sqrt{\sum_{\alpha_j=r_j+1}^{\infty} \lambda_j(\alpha_j)} \lesssim r_j^{-k/n_j} |f|_{H^k(\Omega_1 \times \cdots \times \Omega_m)} \overset{!}{\lesssim} \varepsilon \qquad \Longrightarrow \qquad r_j = \varepsilon^{-n_j/k} \quad (3.13)$$

according to Theorem 1. Doing so for all $j \in \{1, 2, \ldots, m\}$, we obtain the approximation

$$f_{r_1,\ldots,r_m}^{\mathrm{TF}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m) = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_m=1}^{r_m} \omega(\alpha_1,\ldots,\alpha_m)\varphi_1(\alpha_1,\boldsymbol{x}_1)\cdots\varphi_m(\alpha_m,\boldsymbol{x}_m).$$

We have the following result on the Tucker decomposition:

Theorem 2 *Let $f \in H^k(\Omega_1 \times \cdots \times \Omega_m)$ for some fixed $k > 0$ and $0 < \varepsilon < 1$. If the ranks are chosen according to $r_j = \varepsilon^{-n_j/k}$ for all $j = 1,\ldots,m$. Then, the truncation error of the truncated Tucker decomposition is*

$$\left\| f - f_{r_1,\ldots,r_m}^{\mathrm{TF}} \right\|_{L^2(\Omega_1 \times \cdots \times \Omega_m)} \leq \sqrt{\sum_{j=1}^{m} \sum_{\alpha_j=r_j+1}^{\infty} \lambda_j(\alpha_j)} \lesssim \sqrt{m}\varepsilon,$$

while the storage cost for the core tensor of $f_{r_1,\ldots,r_m}^{\mathrm{TF}}$ are $\prod_{j=1}^{m} r_j = \varepsilon^{-(n_1+\cdots+n_m)/k}$.

Proof For the approximation of the core tensor, the sets of the univariate eigenfunctions $\{\varphi_j(\alpha_j)\}_{\alpha_j=1}^{r_j}$ are used for all $j = 1,\ldots,m$, cf. (3.12). Due to orthonormality, we find

$$\left\| f - f_{r_1,\ldots,r_m}^{\mathrm{TF}} \right\|_{L^2(\Omega_1 \times \cdots \times \Omega_m)}^2$$
$$= \sum_{j=1}^{m} \left\| f_{r_1,\ldots,r_{j-1},\infty,\ldots,\infty}^{\mathrm{TF}} - f_{r_1,\ldots,r_j,\infty,\ldots,\infty}^{\mathrm{TF}} \right\|_{L^2(\Omega_1 \times \cdots \times \Omega_m)}^2,$$

where we obtain $f_{\infty,\ldots,\infty}^{\mathrm{TF}} = f$ in case of $j = 1$. Since

$$\left\| f_{r_1,\ldots,r_{j-1},\infty,\ldots,\infty}^{\mathrm{TF}} - f_{r_1,\ldots,r_j,\infty,\ldots,\infty}^{\mathrm{TF}} \right\|_{L^2(\Omega_1 \times \cdots \times \Omega_m)}^2$$
$$\leq \left\| f_{\infty,\ldots,\infty}^{\mathrm{TF}} - f_{\infty,\ldots,\infty,r_j,\infty,\ldots,\infty}^{\mathrm{TF}} \right\|_{L^2(\Omega_1 \times \cdots \times \Omega_m)}^2 = \sum_{\alpha_j=r_j+1}^{\infty} \lambda_j(\alpha_j)$$

for all $j \in \{1, 2, \ldots, m\}$, we arrive with (3.13) and the summation over $j = 1,\ldots,m$ at the desired error estimate. This completes the proof, since the estimate on the number of coefficients in the core tensor is obvious.

3.3 Sobolev spaces with dimension weights

The cost of the core tensor of the Tucker decomposition exhibit the curse of dimension as the number m of subdomains increases. This can be seen most simply for the example $n_j = n$. Then, the cost are $\varepsilon^{-nm/k}$, which expresses the curse of dimension as long as k is not proportional to m. Nonetheless, in case of Sobolev spaces with dimension weights, the curse of dimension can be beaten.

For $f \in H^{k+n}(\Omega^m)$, we shall discuss the situation $m \to \infty$ in more detail. To this end, we assume that all subdomains are identical to a single domain $\Omega \subset \mathbb{R}^n$ of dimension n and note that the limit $m \to \infty$ only makes sense when weights are included in the underlying Sobolev spaces which ensure that higher dimensions become less important. For our proofs we choose as usual m arbitrary but fixed and show the existence of m-independent constants in the convergence and cost estimates.

The Sobolev spaces $H_{\boldsymbol{\gamma}}^k(\Omega^m)$ with dimension weights $\boldsymbol{\gamma} \in \mathbb{R}^m$ which we consider are given by all functions $f \in H^k(\Omega^m)$ such that

$$\left\| \frac{\partial^k f}{\partial \boldsymbol{x}_j^{\boldsymbol{\beta}}} \right\|_{L^2(\Omega^m)} \lesssim \gamma_j^k \|f\|_{H^k(\Omega^m)} \text{ for all } |\boldsymbol{\beta}| = k \text{ and } j = 1, 2, \ldots, m. \tag{3.14}$$

The definition in (3.14) means that, given a function f with norm $\|f\|_{H^k(\Omega^m)} < \infty$, the partial derivatives with respect to $\boldsymbol{x}_j$ become less important as the dimension j increases. Such functions appear for example in uncertainty quantification. Let be given a Karhunen-Loève expansion

$$u(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^m \sigma_j \varphi_j(\mathbf{x}) y_j, \quad y_j \in [-1/2, 1/2],$$

and insert it into a function $b : \mathbb{R} \to \mathbb{R}$ of *finite* smoothness $W^{k,\infty}(\mathbb{R})$. Then, the function $b\big(u(\mathbf{x}, \mathbf{y})\big)$ satisfies (3.14) with respect to the $\mathbf{y}$-variable, where $\gamma_j = \sigma_j$. Hence, the solution of a given partial differential equation would satisfy a decay estimate similar to (3.14) whenever the stochastic field enters the partial differential equation through a non-smooth coefficient function b, compare [14, 23, 24] for example.

It turns out that algebraically decaying weights (3.15) are sufficient to beat the curse of dimension in case of the Tucker tensor decomposition.[4]

Theorem 3 *Given $\delta > 0$, let $f \in H_{\boldsymbol{\gamma}}^k(\Omega^m)$ for some fixed $k > 0$ with weights* (3.14) *that decay like*

$$\gamma_j \lesssim j^{-(1+\delta')/k} \text{ for some } \delta' > \delta + \frac{k}{n}. \tag{3.15}$$

Then, for all $0 < \varepsilon < 1$, the error of the continuous Tucker decomposition with ranks

$$r_j = \lceil \gamma_j^n j^{(1+\delta)n/k} \varepsilon^{-n/k} \rceil \tag{3.16}$$

is of order ε while the storage cost for the core tensor of $f_{r_1, \ldots, r_m}^{\mathrm{TF}}$ are bounded by $\varepsilon^{-n/k}$ independent of the dimension m.

---

[4] In Theorem 3, no truncation of the dimension is applied, as it would be required in practice if the number m of domains tends to infinity. Note that the dimension truncation is indeed here the same as for the tensor train decomposition later on, see also Theorem 5.

Proof In view of Theorem 1 and (3.14), we deduce by choosing the ranks as in (3.13) that

$$\sqrt{\sum_{\alpha_j = r_j+1}^{\infty} \lambda_j(\alpha_j)} \lesssim r_j^{-k/n} \gamma_j^k \|f\|_{H^k(\Omega^m)} \lesssim \frac{\varepsilon}{j^{1+\delta}}.$$

Therefore, we reach the desired over-all truncation error

$$\sum_{j=1}^{m} \frac{\varepsilon}{j^{1+\delta}} \lesssim \varepsilon \text{ as } m \to \infty. \tag{3.17}$$

When the weights γ_j decay as in (3.16), then the cost of the core tensor are

$$C := \prod_{j=1}^{m} r_j \leq \prod_{j=1}^{m} \left(1 + \gamma_j^n j^{(1+\delta)n/k} \varepsilon^{-n/k}\right) \lesssim \prod_{j=1}^{m} \left(1 + j^{-\theta} \varepsilon^{-n/k}\right)$$

with $\theta = (\delta' - \delta)n/k > 1$. Hence, the cost of the core tensor stay independently of m since

$$\log C \lesssim \sum_{j=1}^{m} \log \left(1 + j^{-\theta} \varepsilon^{-n/k}\right) \leq \varepsilon^{-n/k} \sum_{j=1}^{m} j^{-\theta} \lesssim \varepsilon^{-n/k} \text{ as } m \to \infty.$$

4 Tensor train format

4.1 Tensor train decomposition

For the discussion of the continuous tensor train decomposition, we should assume that the domains $\Omega_j \subset \mathbb{R}^{n_j}$, $j = 1, \ldots, m$, are arranged in such a way that it holds $n_1 \leq \cdots \leq n_m$.[5]

Now, consider $f \in H^k(\Omega_1 \times \cdots \times \Omega_m)$ and separate 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

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

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 $(\alpha_1, \boldsymbol{x}_2) \in \mathbb{N} \times \Omega_2$ from $(\boldsymbol{x}_3, \ldots, \boldsymbol{x}_m) \in \Omega_3 \times \cdots \times \Omega_m$ by means of a second singular value decomposition and arrive at

$$\left[\sqrt{\lambda_1(\alpha_1)} \psi_1(\alpha_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)\right]_{\alpha_1 = 1}^{\infty}$$
$$= \sum_{\alpha_2 = 1}^{\infty} \sqrt{\lambda_2(\alpha_2)} \left[\varphi_2(\alpha_1, \boldsymbol{x}_2, \alpha_2)\right]_{\alpha_1 = 1}^{\infty} \psi_2(\alpha_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_m). \tag{4.18}$$

---

[5] The considerations in this section are based upon [5]. Nonetheless, the results derived there are not correct. The authors did not consider the impact of the vector-valued singular value decomposition in a proper way, which indeed does result in the curse of dimension.

By repeating the last 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

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

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 contrast to the Tucker format, we do not obtain a huge core tensor since each of the $m-1$ singular value decompositions of the tensor train decomposition removes the actual first spatial domain from the approximant. We just obtain a product of *matrix*-valued functions (except for the first and last factor which are vector-valued functions), each of which is related with a specific domain Ω_j. This especially results in only $m-1$ sums in contrast to the m sums for the Tucker format.

4.2 Truncation error

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

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

One readily infers by using Pythogoras' theorem that the truncation error is bounded by

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

see also [35]. 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 original continuous tensor train decomposition due to the truncation.

   We next shall give bounds on the truncation error. In the j-th step of the algorithm, $j = 2, 3, \ldots, m-1$, one needs to approximate the vector-valued function

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

by a vector-valued singular value decomposition. This means that we consider the singular value decomposition (2.9) for a vector-valued function in case of the domains Ω_j and $\Omega_{j+1} \times \cdots \times \Omega_m$.

For $f \in H^{k+n_{m-1}}(\Omega_1 \times \cdots \times \Omega_m)$, it holds $\boldsymbol{g}_2 \in [H^{k+n_{m-1}}(\Omega_2 \times \cdots \times \Omega_m)]^{r_1}$ and

$$|\boldsymbol{g}_2|_{[H^k(\Omega_2 \times \cdots \times \Omega_m)]^{r_1}} \leq \sqrt{\sum_{\alpha_1=1}^{\infty} \lambda_1(\alpha_1)|\psi_1(\alpha_1)|^2_{H^k(\Omega_2 \times \cdots \times \Omega_m)}}$$

$$\leq |f|_{H^k(\Omega_1 \times \cdots \times \Omega_m)}$$

according to Lemma 4, precisely in its vectorized version (2.10). It follows $\boldsymbol{g}_3 \in [H^{k+n_{m-1}}(\Omega_3 \times \cdots \times \Omega_m)]^{r_2}$ and, again by (2.10),

$$|\boldsymbol{g}_3|_{[H^k(\Omega_3 \times \cdots \times \Omega_m)]^{r_2}} \leq \sqrt{\sum_{\alpha_2=1}^{\infty} \lambda_2(\alpha_2)|\psi_2(\alpha_2)|^2_{H^k(\Omega_3 \times \cdots \times \Omega_m)}}$$

$$\leq |\boldsymbol{g}_2|_{[H^k(\Omega_2 \times \cdots \times \Omega_m)]^{r_1}}.$$

We hence conclude recursively $\boldsymbol{g}_j \in [H^{k+n_{m-1}}(\Omega_j \times \cdots \times \Omega_m)]^{r_{j-1}}$ and

$$|\boldsymbol{g}_j|_{[H^k(\Omega_j \times \cdots \times \Omega_m)]^{r_{j-1}}} \leq |f|_{H^k(\Omega_1 \times \cdots \times \Omega_m)} \text{ for all } j = 2, 3, \ldots, m-1. \quad (4.19)$$

Estimate (4.19) shows that the H^k-seminorm of the vector-valued functions $\boldsymbol{g}_j$ stays bounded by $|f|_{H^k(\Omega_1 \times \cdots \times \Omega_m)}$. But according to (2.9), we have in the j-th step only the truncation error estimate

$$\left\| \boldsymbol{g}_j - \sum_{\alpha_j=1}^{r_j} \sqrt{\lambda_j(\alpha_j)} \big(\boldsymbol{\varphi}_j(\alpha_j) \otimes \psi_j(\alpha_j) \big) \right\|_{[L^2(\Omega_j \times \cdots \times \Omega_m)]^{r_{j-1}}}$$

$$\lesssim \left(\frac{r_j}{r_{j-1}} \right)^{-k/n_j} |\boldsymbol{g}_j|_{[H^k(\Omega_j \times \cdots \times \Omega_m)]^{r_{j-1}}}.$$

Hence, in view of (4.19), to achieve the target accuracy ε per truncation, the truncation ranks need to be increased in accordance with

$$r_1 = \varepsilon^{-n_1/k}, \quad r_2 = \varepsilon^{-(n_1+n_2)/k}, \quad \ldots, \quad r_{m-1} = \varepsilon^{-(n_1+\cdots+n_{m-1})/k}. \quad (4.20)$$

We summarize our findings in the following theorem, which holds in this form also if the subdomains $\Omega_j \subset \mathbb{R}_{n_j}$ are not ordered in such a way that $n_1 \leq \cdots \leq n_m$.

Theorem 4 *Let $f \in H^{k+\max\{n_1,\ldots,n_m\}}(\Omega_1 \times \cdots \times \Omega_m)$ for some fixed $k > 0$ and $0 < \varepsilon < 1$. Then, the over-all truncation error of the tensor train decomposition with truncation ranks (4.20) is*

$$\|f - f^{\mathrm{TT}}_{r_1,\ldots,r_{m-1}}\|_{L^2(\Omega_1 \times \cdots \times \Omega_m)} \lesssim \sqrt{m}\varepsilon.$$

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

$$r_1 + \sum_{j=2}^{m-1} r_{j-1}r_j = \varepsilon^{-n_1/k} + \varepsilon^{-(2n_1+n_2)/k} + \cdots + \varepsilon^{-(2n_1+\cdots+2n_{m-2}+n_{m-1})/k} \quad (4.21)$$

and hence are bounded by $\mathcal{O}(\varepsilon^{-(2m-1)\max\{n_1,\ldots,n_{m-1}\}/k})$.

Remark 2 If $n := n_1 = \cdots = n_m$, then the cost of the tensor train decomposition are $\mathcal{O}(\varepsilon^{-(2m-1)n/k})$. Thus, the cost are quadratic compared to the cost of the Tucker decomposition. However, in practice, one performs $m/2$ forward steps and $m/2$ backward steps. This means one computes $m/2$ steps as described above to successively separate $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{m/2}$ from the other variables. Then, one performs the algorithm in the opposite direction, i.e., one successively separates $\boldsymbol{x}_m, \boldsymbol{x}_{m-1}, \ldots, \boldsymbol{x}_{m/2+1}$ from the other variables. This way, the over-all cost are reduced to the order $\mathcal{O}(\varepsilon^{-mn/k}).$[6]

4.3 Sobolev spaces with dimension weights

Like for the Tucker decomposition, the cost of the tensor train decomposition suffer from the curse of dimension as the number m of subdomains increases. We therefore discuss again appropriately Sobelev spaces with dimension weights, where we assume for reasons of simplicity that all subdomains are identical to a single domain $\Omega \subset \mathbb{R}^n$ of dimension n.

Theorem 5 *Given $\delta > 0$, let $f \in H_{\boldsymbol{\gamma}}^{k+n}(\Omega^m)$ for some fixed $k > 0$ with weights (3.14) that decay like (3.15). For $0 < \varepsilon < 1$, choose the ranks successively in accordance with*

$$r_j = \lceil r_{j-1} \gamma_j^n j^{(1+\delta)n/k} \varepsilon^{-n/k} \rceil \qquad (4.22)$$

if $j \leq M$ and $r_j = 0$ if $j > M$. Here, M is given by

$$M = \varepsilon^{-1/(1+\delta')}. \qquad (4.23)$$

Then, the error of the continuous tensor train decomposition is of order ε while the storage cost of $f_{r_1,\ldots,r_m}^{\mathrm{TT}}$ stay bounded by $M \exp(\varepsilon^{-n/k})^2$ independent of the dimension m.

Proof The combination of Theorem 1, (3.14) and (4.22) implies

$$\sqrt{\sum_{\alpha_j=r_j+1}^{\infty} \lambda_j(\alpha_j)} \lesssim \left(\frac{r_j}{r_{j-1}}\right)^{-k/n} \gamma_j^k \|f\|_{H^k(\Omega^m)} \lesssim \frac{\varepsilon}{j^{1+\delta}}, \quad j = 1, 2, \ldots, M,$$

and

$$\sqrt{\sum_{\alpha_{M+1}=1}^{\infty} \lambda_{M+1}(\alpha_{M+1})} \lesssim \gamma_{M+1}^k \|f\|_{H^k(\Omega^m)} \lesssim \varepsilon \|f\|_{H^k(\Omega^m)}.$$

Hence, as in the proof of Theorem 3, the approximation error of the continuous tensor train decomposition is bounded by a multiple of ε independent of m.

Next, we observe for all $j \leq M$ that

$$r_j \lesssim \lceil r_{j-1} \gamma_j^n j^{(1+\delta)n/k} \varepsilon^{-n/k} \rceil \lesssim r_{j-1} \gamma_j^n j^{(1+\delta)n/k} \varepsilon^{-n/k} + 1.$$

---

[6] If the spatial dimensions n_j, $j = 1, \ldots, m$, of the subdomains are different, one can balance the number of forward and backward steps in a better way to reduce the cost further.

This recursively yields

$$r_j - 1 \lesssim \sum_{p=1}^{j} \prod_{q=p}^{j} \gamma_q^n q^{(1+\delta)n/k} \varepsilon^{-n/k}$$

$$\lesssim \sum_{p=1}^{j} \varepsilon^{(p-j-1)n/k} \prod_{q=p}^{j} q^{-\theta}$$

$$= \sum_{p=1}^{j} \varepsilon^{(p-j-1)n/k} \left(\frac{(p-1)!}{j!} \right)^{\theta}$$

$$= \sum_{p=1}^{j} \varepsilon^{-pn/k} \left(\frac{(j-p)!}{j!} \right)^{\theta}.$$

Hence, by using that $\theta = (\delta' - \delta)n/k > 1$, we obtain

$$r_j \lesssim \sum_{p=0}^{j} \varepsilon^{-pn/k} \frac{(j-p)!}{j!} \leq \sum_{p=0}^{j} \frac{\varepsilon^{-pn/k}}{p!} \leq \exp(\varepsilon^{-n/k}).$$

Therefore, the cost (4.21) are

$$r_1 + \sum_{j=2}^{M} r_{j-1} r_j \leq \sum_{j=1}^{M} r_j^2 \lesssim M \exp(\varepsilon^{-n/k})^2$$

and are hence bounded independently of m in view of (4.23).

5 Discussion and conclusion

In the present article, we considered the continuous versions of the Tucker tensor format and of the tensor train format for the approximation of functions which live on an m-fold product of arbitrary subdomains. By considering (isotropic) Sobolev smoothness, we derived estimates on the ranks to be chosen in order to realize a prescribed target accuracy. These estimates exhibit the curse of dimension.

Both tensor formats have in common that always only the variable with respect to a single domain is separated from the other variables by means of the singular value decomposition. This enables cheaper storage schemes, while the influence of the over-all dimension of the product domain is reduced to a minimum.

We also examined the situation of Sobolev spaces with dimension weights. Having sufficiently fast decaying weights helps to beat the curse of dimension as the number of subdomains tends to infinity. It turned out that algebraically decaying weights are appropriate for both, the Tucker tensor format and the tensor train format.

We finally remark that we considered here only the ranks of the tensor decomposition in the *continuous* case, i.e., for functions and not for tensors of discrete data. Of course, an additional projection step onto suitable finite dimensional trial spaces on the individual domains would be necessary to arrive at a fully discrete approximation scheme that can really be used in computer simulations. This would impose a further error of discretization type which needs to be balanced with the truncation error of the particular continuous tensor format.

References

1. I. Babuška and J. Osborn. Eigenvalue problems. In *Handbook of Numerical Analysis*, vol. II, North-Holland, Amsterdam, 1991, pp. 641–784.
2. M. Bachmayr and W. Dahmen. Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. *Found. Comput. Math.* **15**(4) (2015), 839–898.
3. M. Bachmayr and W. Dahmen. Adaptive low-rank methods for problems on Sobolev spaces with error control in L_2. *ESAIM Math. Model. Numer. Anal.* **50**(4) (2016), 1107–1136.
4. 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.
5. D. Bigoni, A. Engsig-Karup, and Y. Marzouk. Spectral tensor-train decomposition. *SIAM J. Sci. Comput.* **48**(4) (2016), A2405–A2439.
6. D. Braess. *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge, NY, 2001.
7. S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, Berlin, 2008.
8. H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica* **13** (2004), 147–269.
9. A. Chernov and C. Schwab. First order k-th moment Finite Element analysis of nonlinear operator equations with stochastic data. *Math. Comput.* **82**(284) (2013), 1859–1888.
10. D. Cioranescu, A. Damlamian, and G. Griso. The periodic unfolding method in homogenization. *SIAM J. Appl. Math.* **40** (2008), 1585–1620.
11. N. Cohen, O. Sharir, and A. Shashua. On the expressive power of deep learning: A tensor analysis. 29th Annual Conference on Learning Theory (COLT), *PMLR* **49** (2016), 698–728.
12. A. Falcó and W. Hackbusch. On minimal subspaces in tensor representations. *Found. Comput. Math.* **12** (2012), 765–803.
13. A. Falcó, W. Hackbusch, and A. Nouy. Tree-based tensor formats. *SeMA Journal* **78**(2) (2021), 159–173.
14. R. Gantner and M. Peters. Higher order quasi-Monte Carlo for Bayesian shape inversion. *SIAM/ASA J. Uncertain. Quantif.* **6** (2018), 707–736.
15. L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.* **31** (2010), 2029–2054.
16. L. Grasedyck, D. Kressner, and C. Tobler. A literature survey of low-rank tensor approximation techniques, *GAMM-Mitteilungen* **36**(1) (2013), 53–78.
17. M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Math. Comput.* **82**(282) (2013), 975–994.
18. M. Griebel and H. Harbrecht. A note on the construction of L-fold sparse tensor product spaces. *Constr. Approx.* **38**(2) (2013), 235–251.
19. M. Griebel and H. Harbrecht. Approximation of bi-variate functions. Singular value decomposition versus sparse grids. *IMA J. Numer. Anal.* **34** (2014), 28–54.
20. M. Griebel and H. Harbrecht. Singular value decomposition versus sparse grids. Refined complexity estimates. *IMA J. Numer. Anal.* **39** (2019), 1652–1671.
21. W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer, Berlin-Heidelberg, 2012.
22. W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *J. Fourier Anal. Appl.* **15**(5) (2009), 706–722.
23. H. Harbrecht, M. Peters, and M. Siebenmorgen. Analysis of the domain mapping method for elliptic diffusion problems on random domains. *Numer. Math.* **134** (2016), 823–856.
24. H. Harbrecht and M. Schmidlin. Multilevel methods for uncertainty quantification of elliptic PDEs with random anisotropic diffusion. *Stoch. Partial Differ. Equ. Anal. Comput.* **8** (2020), 54–81.

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

26. F. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.* **6** (1927), 164–189.

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

28. V. Khrulkov, A. Novikov, and I. Oseledets. Expressive power of recurrent neural networks. In *6th International Conference on Learning Representations, ICLR 2018 – Conference Track Proceedings*, 6th International Conference on Learning Representations, ICLR 2018, Vancouver, 30 April 2018–3 May 2018, 149806.

29. 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.

30. M. Lòeve. *Probability Theory*, Vol. I+II, Springer, New York, 1978.

31. 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.

32. A. Matache. Sparse two-scale FEM for homogenization problems. *SIAM J. Sci. Comput.* **17** (2002), 709–720.

33. W. McLean. *Strongly elliptic systems and boundary integral equations.* Cambridge University Press, Cambridge, NY, 2000.

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

35. I. Oseledets and E. Tyrtyshnikov. Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.* **31**(5) (2009), 3744–3759.

36. D. Pérez-García, F. Verstraete, M. Wolf, and I. Cirac. Matrix product state representations. *Quantum Inf. Comput.* **7** (2007), 401–430.

37. A. Rüttgers and M. Griebel. Multiscale simulation of polymeric fluids using the sparse grid combination technique. *Appl. Math. Comput.* **319** (2018), 425–443.

38. E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. *Math. Ann.* **63** (1907), 433–476.

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

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

41. C. Schwab and R.-A. Todor. Sparse finite elements for stochastic elliptic problems. Higher order moments. *Computing* **71** (2003), 43–63.

42. J. Šimša. The best L^2-approximation by finite sums of functions with separate variables. *Aequationes Math.* **43** (1992), 248–263.

43. F. Smithies. The eigen-values and singular values of integral functions. *Proc. London Math. Soc. (2)* **43**(1) (1937), 255–279.

44. G. Stewart. On the early history of the singular value decomposition. *SIAM Rev.* **35**(4) (1993), 551–566.

45. O. Steinbach. *Numerical approximation methods for elliptic boundary value problems. Finite and boundary elements.* Springer, New York, 2008.

46. V. Temlyakov. Approximation of functions with bounded mixed derivative. *Tr. Mat. Inst. Steklov.*, **178** (1986), 3–113 (in Russian); *Proc. Steklov Inst. Math.*, **1** (1989), 1–121 (English translation).

47. V. Temlyakov. Estimates for the best bilinear approximations of periodic functions. *Tr. Mat. Inst. Steklov.* **181** (1988), 250–267 (in Russian); *Proc. Steklov Inst. Math.*, **4** (1989), 275–293 (English translation).

48. V. Temlyakov. Bilinear approximation and related questions. *Tr. Mat. Inst. Steklov.* **194** (1992), 229–248 (1992)(in Russian); *Proc. Steklov Inst. Math.* **4**(194) (1993), 245–265 (English translation).

49. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**(3) (1966), 279–311.

50. C. Zenger. Sparse grids. In *Parallel algorithms for partial differential equations*, Proceedings of the 6th GAMM-Seminar, Kiel/Germany, 1990, Notes Numer. Fluid Mech. 31, Vieweg, Braunschweig, 1991, pp. 241–251.