



Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: [www.elsevier.com/locate/jco](http://www.elsevier.com/locate/jco)

# Approximation rates for the hierarchical tensor format in periodic Sobolev spaces

Reinhold Schneider<sup>a</sup>, André Uschmajew<sup>b,\*</sup><sup>a</sup> Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany<sup>b</sup> MATHICSE-ANCHP, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

## ARTICLE INFO

### Article history:

Available online 15 October 2013

Dedicated to J.F. Traub and G.W. Wasilkowski on the occasion of their 80th and 60th birthdays

### Keywords:

Approximation of multi-variate functions  
Hierarchical tensor format  
Hierarchical Tucker rank  
High-order SVD  
Bilinear approximation

## ABSTRACT

In this note we estimate the asymptotic rates for the  $L_2$ -error decay and the storage cost when approximating  $2\pi$ -periodic,  $d$ -variate functions from isotropic and mixed Sobolev classes by the recent hierarchical tensor format as introduced by Hackbusch and Kühn. To this end, we survey some results on bilinear approximation due to Temlyakov. The approach taken in this paper improves and generalizes recent results of Griebel and Harbrecht for the bi-variate case.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

In numerical mathematics, nonlinear tensor product approximation techniques in function spaces and their discretized analogs search for possibilities to escape the exponential scaling of classical numerical methods when applied to functions with a very large number of variables. The main idea is to represent such high-dimensional functions in certain *tensor formats*, which ultimately means to represent them as sums of separable functions (products of functions of few variables). The state of the art has been recently presented in a monograph by Hackbusch [20]. For a literature survey see [18].

Throughout the paper let  $d \geq 2$ . We denote by  $L_2(\pi_d)$  the Hilbert space of  $d$ -variate,  $2\pi$ -periodic  $L_2$ -functions  $f(x_1, x_2, \dots, x_d)$  on  $\mathbb{R}^d$  with norm

$$\|f\|_{L_2} = \left( \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} |f(\mathbf{x})|^2 \, d\mathbf{x} \right)^{1/2}.$$

\* Corresponding author.

E-mail addresses: [schneidr@math.tu-berlin.de](mailto:schneidr@math.tu-berlin.de) (R. Schneider), [andre.uschmajew@epfl.ch](mailto:andre.uschmajew@epfl.ch) (A. Uschmajew).

The most natural setting, on which we solely focus in this paper, is the approximation of  $f \in L_2(\pi_d)$  by linear combinations of products  $u^1(x_1)u^2(x_2) \cdots u^d(x_d)$  of univariate functions  $u_k^\mu \in L_2(\pi_1)$ . This problem fits into the abstract framework of low-rank approximation in tensor products of Hilbert spaces. The way how these linear combinations are built up is referred to as a *tensor format*. In this paper we focus on the *hierarchical Tucker format* as introduced by Hackbusch and Kühn [23]. The results include the extended version [35] of the *TT-format* of Oseledets and Tyrtshnikov [34,33], where bases are also installed in the leaves of the tree. We note that in quantum physics these tensor formats have been known earlier as *matrix product states* and *tensor networks*, see [42] for a recent survey including historical remarks.

The approximation by tensor products is *nonlinear* in the sense that only the tensor format structure of the approximant, but no particular approximation subspaces for the univariate functions, are prescribed. The approximation power of a tensor format and the number of terms needed for storing functions in this format is typically governed by a *rank* parameter (depending on the chosen format), and the order  $d$  of the tensor product (the number of variables to be separated). The approach has demonstrated its effectiveness in a number of applications. In particular, the hierarchical tensor format we will consider has been applied successfully to various high-dimensional problems. Molecular Schrödinger equation [24], parabolic partial differential equations and instationary Schrödinger equation in many space dimensions have been computed based on Dirac–Frenkel variational principles e.g. in [25], and for Fokker–Planck equation in [13]. The chemical master equation has been treated in [12]. Also, boundary value problems with uncertain coefficient boundary data lead to parametric partial differential equations in high dimensions. This problem has been attacked in [26,14]. In [15] the recent developments have been applied to numerical homogenization. Besides, within the DMRG (density matrix renormalization group) approach, see e.g. [31], tensor approximation has been applied to problems in quantum chemistry, e.g. in [5]. In [4] efficient tensor approximation is used for the fast evaluation of singular integrals appearing in the Galerkin discretization of boundary integral equations.

But despite some progress in special cases, the approach so far lacks rigorous answers to the following question: how do approximation classes of functions (tensors) look like with regard to the ranks of the considered format. Or, putting it the other way around, which properties must a function have for being well approximable with moderate ranks.

In the present note we at least investigate what one has to expect in classical approximation spaces. In particular, we determine the approximation rates in terms of the *hierarchical Tucker rank* for the Sobolev classes

$$B^s = \{f \in L_2(\pi_d) : \|f\|_s \leq 1\} \quad \text{and} \quad B^{s, \text{mix}} = \{f \in L_2(\pi_d) : \|f\|_{s, \text{mix}} \leq 1\},$$

which are balls with respect to the semi-norms

$$\|f\|_s^2 = \max_{\mu=1,2,\dots,d} \|f\|_{s,\mu}^2 \quad \text{and} \quad \|f\|_{s, \text{mix}}^2 = \sum_{\mathbf{k} \in \mathbb{Z}^d} \left( \prod_{\mu=1}^d \hat{k}_\mu \right)^{2s} |\hat{f}(\mathbf{k})|^2, \tag{1}$$

respectively, where we have set

$$\|f\|_{s,\mu}^2 = \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{k}_\mu^{2s} |\hat{f}(\mathbf{k})|^2, \quad \text{and} \quad \hat{k}_\mu = \begin{cases} |k_\mu|, & \text{for } k_\mu \neq 0, \\ 1 & \text{for } k_\mu = 0. \end{cases}$$

These definitions are analogous to those in Temlyakov’s monograph [44]. In case that  $s$  is an integer, the class  $B^s$  consists of functions which admits weak partial derivatives up to order  $s$ , while functions in  $B^{s, \text{mix}}$  can have mixed partial derivatives up to order  $ds$ , but only if every variable is differentiated to order  $s$  only. This class of *dominating mixed smoothness* has been first considered by Babenko [2].

In a way, the main content of the paper solely consists in a rather obvious combination of two known results: the quasi optimality of the high order SVD truncation to obtain hierarchical Tucker format approximations due to Grasedyck and others, and the known estimates for best bilinear approximation (SVD truncation) in  $L_2$  due to Temlyakov. However, we had the feeling it would be worth to piece these known results together in a rigorous way. From the result on the asymptotic

behavior with respect to the hierarchical rank we can deduce asymptotic estimates for the storage requirement (degrees of freedom) one needs to spend to achieve an approximation of accuracy  $\epsilon$  using a function which is encoded in the hierarchical format. We do not address the question how much work has to be spent to calculate such a function. The considerations follow in spirit those in the paper of Griebel and Harbrecht [19] where the case  $d = 2$  (when the hierarchical Tucker format is just the SVD) has been discussed, although we will rely on refined singular value estimates from the literature.

In particular, we are interested in a comparison with optimal linear approximation methods like sparse grids. To this end, let us briefly recall the background. We call an approximation method for a function  $f$  from some function space  $X$  *linear* if it consists in a series of best approximations from a fixed sequence of (typically nested) subspaces  $V_n$  of growing dimension  $n$ . The approximation rate in terms of the dimension  $n$  a linear method can at most achieve for a set of functions  $F \subseteq X$  is determined by the Kolmogorov width [27]

$$d_n(F, X) = \inf_{\substack{V_n \subseteq X \\ \dim V_n = n}} \sup_{f \in F} \inf_{g \in V_n} \|f - g\|_X.$$

Note that the dimension  $n$  is also the adequate parameter for the storage complexity of a linear method.

In the same paper [27], Kolmogorov proved the nowadays well-known result<sup>1</sup>

$$d_n(B^s, L_2) \sim n^{-s/d} \quad (n \rightarrow \infty)$$

for the case  $d = 1$ . A proof for  $d \geq 2$  can be found in [48] or [32]. The dimension  $n(f, \epsilon)$  needed to approximate an arbitrary function  $f \in B^s$  up to an error  $\epsilon > 0$  in the  $L_2$  norm hence behaves like

$$n(f, \epsilon) \sim \epsilon^{-d/s} \quad (\epsilon \rightarrow 0). \tag{2}$$

To keep this tolerable the regularity has to grow proportional to the space dimension,  $s \sim d$ , which is one version of the *curse of dimensionality*.

A partial way out is to let only mixed derivatives increase. Babenko [3,2] proved

$$d_n(B^{s, \text{mix}}, L_2) \sim n^{-s} (\log n)^{s(d-1)} \quad (n \rightarrow \infty),$$

see [44] for a full treatise. The dimension  $n(f, \epsilon)$  needed for  $f \in B^{s, \text{mix}}$  can be shown (cf. Lemma 1) to behave like

$$n(f, \epsilon) \lesssim s^{-(d-1)} \epsilon^{-1/s} |\log \epsilon|^{(d-1)} \quad (\epsilon \rightarrow \infty), \tag{3}$$

and this cannot really be improved without further knowledge of  $f$ . The occurrence of the  $(d - 1)$ -th power of the logarithm typically serves as explanation why the applicability of these methods is, as observed in practice, restricted to, say,  $d \leq 10$  variables, unless the regularity is not unrealistically high. Note however that there are also hidden constants scaling exponentially in  $d$ .

Interestingly, the known methods that realize the optimal rates (2) and (3), like best approximation by trigonometric polynomials of degree  $n$  for  $B^s$ , or the hyperbolic cross method, sparse grids, or wavelets for  $B^{s, \text{mix}}$ , can be seen as tensor product approximation methods, but with fixed basis sets. Namely, they construct approximations of the form

$$g(x_1, x_2, \dots, x_d) = \sum_{k=1}^r u_k^1(x_1) u_k^2(x_2) \cdots u_k^d(x_d) \tag{4}$$

(with  $r = n$ ), which is known as the *canonical tensor format* in multi-linear algebra. The idea of *nonlinear* tensor approximation is to not restrict the factors  $u_k^d$  in advance. This should be advantageous in general. However, the question is whether it makes a substantial difference in the considered Sobolev classes in the asymptotic regime. It is widely believed that these classical smoothness spaces do not

---

<sup>1</sup> In all places where notations  $\sim$  and  $\lesssim$  are used the hidden constants usually depend (in an unfavorable way) on  $d$ , and may depend on  $s$ .

form the right framework for nonlinear tensor approximation theory. In a certain way we will confirm this for the hierarchical tensor format later. That the general suspicion is not completely justified is shown by the following results.

Let  $\mathcal{C}_{\leq r}$  denote the set of all functions  $g$  representable in the form (4). One refers to  $r$  as the canonical rank. Approximation by functions from  $\mathcal{C}_{\leq r}$  can be regarded as best  $r$ -term approximation with the dictionary consisting of rank-one decomposable functions. For the case  $s = 1$  Hackbusch [22] showed how a sparse grid approximation with the complexity (3) can be compressed into the canonical format using rank

$$r \lesssim \epsilon^{-(d-1)/d} |\log \epsilon|^{d-3} \quad (\epsilon \rightarrow 0).$$

Much earlier, Temlyakov [45] proved<sup>2</sup> for the periodic case that

$$\sup_{f \in B^{s, \text{mix}}} \inf_{g \in \mathcal{C}_{\leq r}} \|f - g\|_{L_2} \lesssim r^{-sd/(d-1)} \tag{5}$$

which gives an asymptotic bound of  $\epsilon^{-(d-1)/(sd)}$  on the rank to achieve accuracy  $\epsilon$ . Measured in number of terms this is significantly better than (3), and, most importantly, does not scale unfavorably with  $d$ . Also, it is not claimed that this is not improvable – the exact rate is unknown. From the practical side, however, as correctly noted by Griebel and Harbrecht [19], one also has to determine and store the initially unknown functions  $u_k^\mu$  which is not the case with (3) where a basis has been fixed. In [49] it is shown that for best approximations  $g \in \mathcal{C}_{\leq r}$  of  $f \in B^s \supset B^{s, \text{mix}}$  the factors  $u_k^\mu$  also have Sobolev regularity  $s$ , and therefore might be nicely approximable. However, when  $d \geq 3$  no reasonable estimates for their Sobolev norms in terms of  $f$  could be given, which makes the argument not very convincing. The problem is related to the deeper fact that  $\mathcal{C}_{\leq r}$  is not closed for  $d \geq 3$  and  $r \geq 2$  [50], that is, the infimum in (5) is typically not attained. This phenomenon already occurs in finite dimension, see the article of De Silva and Lim [9], and is suspected of being responsible that no robust general method to calculate almost best approximations from  $\mathcal{C}_{\leq r}$  is known. This is one main motivation for considering alternative tensor formats.

In contrast, for  $d = 2$ , i.e., for the bilinear approximation problem

$$f(x, y) \approx \sum_{k=1}^r u_k(x)v_k(y),$$

the problem is completely solved by the singular value decomposition (also called Schmidt expansion) of the corresponding integral operator with kernel  $f$ , which can be implemented in a stable approximation procedure. The hierarchical tensor format is based on a hierarchy of bilinear decompositions which ultimately lead to a representation of a function as a sum of products, but in a recursive way. A stable, almost optimal technique, called the *high order SVD truncation (HOSVD)*, is known for obtaining approximations using successive truncation of singular value decompositions of integral operators

$$(Av)(\mathbf{x}) = \int f(\mathbf{x}, \mathbf{y})v(\mathbf{y}) \, d\mathbf{y} \tag{6}$$

corresponding to different “unfoldings” of  $f$  into an  $L_2$  function  $f(\mathbf{x}, \mathbf{y})$  of two block variables  $\mathbf{x} = (x_{\mu_1}, \dots, x_{\mu_n})$  and  $\mathbf{y} = (x_{\nu_1}, \dots, x_{\nu_m})$ ,  $n + m = d$ . The splittings of variables one has to consider are fixed by the choice of a *dimension tree* that defines the format as will be explained later. As it was first observed by De Lathauwer et al. [8], and later generalized by Grasedyck [17], the overall squared approximation error of such a successive singular value truncation procedure can be bounded by the sum (over the whole tree) of squares of deleted singular values of the integral operators associated

<sup>2</sup> In his papers, Temlyakov considered functions of bounded mixed derivatives having only non-vanishing Fourier coefficients when all  $k_\mu \neq 0$ . However, the same asymptotic can be obtained inductively for  $f \in B^{s, \text{mix}}$  by subtraction of the parts belonging to the coordinate hyperplanes in  $\mathbb{Z}^d$  and using the lower dimensional results. In the asymptotic regime ( $r \rightarrow \infty$ ) the involved constants change only mildly (by a polynomial factor), since one gains some small negative powers of  $r$  at lower dimensions.

to the unfoldings. Thus the approximation rate in terms of the hierarchical separation ranks (the ranks used for the SVD truncations) is immediately obtained when approximation rates for bilinear approximation are known. The latter have been determined by Temlyakov in [44,46,47].

In terms of storage requirement  $c(f, \epsilon)$  for achieving accuracy  $\epsilon$  with approximation in the hierarchical format we will see that a naive straightforward calculation for an arbitrary tree leads to the asymptotic bound

$$c(f, \epsilon) \lesssim \max(\epsilon^{-d/s}, \epsilon^{-(2+1/(2s))/s}) \quad (\epsilon \rightarrow 0) \quad (7)$$

for  $f \in B^s$ . For the case  $d = 2$  and  $s > 1/2$  Griebel and Harbrecht [19] obtained  $c(f, \epsilon) \lesssim \epsilon^{-2/(s-1/2)}$  which is much worse for  $s$  close to  $1/2$  than our bound  $\epsilon^{-(2+1/(2s))/s}$ . For large  $s$ , both are almost the same, although our bound is always smaller. The term  $\epsilon^{-d/s}$  in (7) is due to the storage requirement of the transfer tensors (explained later), while  $\epsilon^{-(2+1/(2s))/s}$  is the estimated number of degrees of freedom required to approximate the HOSVD bases in the leaves by a linear method. As observed by Griebel and Harbrecht, in the case  $d = 2$  the latter is more expensive (the transfer tensor is a diagonal matrix then). We see from (7) that when  $d \geq 3$ , it depends on the regularity  $s$  which part asymptotically deteriorates the bound for the overall cost.

For  $f \in B^{s, \text{mix}}$  we will obtain

$$c(f, \epsilon) \lesssim \epsilon^{-(3+1/(2s))/(2s)}$$

for  $d = 2$ , and

$$c(f, \epsilon) \lesssim \epsilon^{-\max(\deg(T), 3+1/(2s))/(2s)} \times \text{powers of } |\log \epsilon| \quad (8)$$

when  $d \geq 3$ , where  $\deg(T)$  is the maximum number of connections a node in the tree  $T$  can have, that is, the largest order of required transfer tensors. It is interesting to note that for a binary tree, as in the pure SVD case  $d = 2$ , the dominating part in the upper bound for a function  $f \in B^{s, \text{mix}}$  is always due to the approximation of the HOSVD bases in the leaves (the second term in the max in (8)), since  $\deg(T) = 3$  then. For other trees such as the Tucker tree the storage requirement of the transfer tensors (the first term in max) which are of size  $r^{\deg(T)}$  can take over (depending on the regularity).

The bounds (7) and (8) are our main result and stated in [Theorem 2](#). They are worse than (2) and (3), respectively (we have  $\deg(T) \geq 3$  when  $d \geq 3$ ). But we emphasize that they are only upper bounds and need to be interpreted sensibly. In particular, they are calculated using a *linear* approximation method for the HOSVD bases, and not looking for further structure in the transfer tensors. However, this “naive” way of counting the degrees of freedom is close to what one would do “in practice”. We do not expect any specific properties from the HOSVD bases except their regularity, so restricting to a fixed linear approximation method for them seems very reasonable. But it is not obviously excluded the possibility that the considered function classes always allow for hierarchical approximations with sparse or low-rank transfer tensors. This is for instance the case for the Tucker format (which is the opposite extreme to binary trees) where simply using hyperbolic cross approximation leads to better storage cost (cf. [Remark 1](#)).

Taking a more pessimistic (theoretical) viewpoint, the successive bilinear approximation might be regarded as an inappropriate approach in a space of dominating mixed smoothness which, by its definition, is tailored to hyperbolic cross approximation. On the other hand, one should emphasize the black box character of the hierarchical tensor approximation, and more specifically the HOSVD, in  $L_2$ . The class of functions providing good approximation properties is expected to be more general than the Sobolev classes, although a precise statement for this is unknown. In particular, it is not clear so far whether these classes are comparable at all.

In this sense, our restriction to Sobolev spaces is not intended to advocate for the sparse grids techniques (we are neutral in this question), but is due to the fact that this is still one of the few known frameworks in which one can say at least something fairly general without being too tautological. As we will also point out in [Section 2.1](#), the correct framework for treating the approximation property of the hierarchical tensor formats would be the theory of operator ideals, more precisely, the theory of singular numbers for various integral operators (6) with the same kernel, but different space splittings. Characterizations of classes of kernels that belong to several Schatten ideals simultaneously

(characterizations other than this one) would give rise to approximation results for the hierarchical format. In Section 2.1 we rely on some known results we have found in the literature and bring their implications to tensor product approximations to attention. With this aim in mind the paper is naturally more devoted to the tensor product community, although it is our hope that experts in the theory of operator ideals could be stimulated seeing new applications for their results.

Let us finally note that generalizations to approximations in other than the  $L_2$ -norm, even to  $L_p$ -norms, seems not to be straightforward. Although results on bilinear approximation rates are known, the quasi optimality of the high-order SVD truncation, which relies on the orthogonality of SVD projections, cannot be exploited. It is likely that periodic functions could be replaced by Sobolev functions on cubes. But since our results rely on approximation rates for bilinear approximation from the literature, and since for the mix norms we only found those of Temlyakov for the periodic case, we preferred staying on the safe side here. Considering anisotropic Sobolev spaces should be straightforward, but a little clumsy, therefore we leave it out here and only mention the required results on bilinear approximation (in the next section). In particular, it then would be of importance which variables are actually separated at each node of the dimension tree, and not just how many. Results for other function classes are obtained immediately as soon as results on bilinear approximation are available. A very recent result by Romanyuk and Romanyuk provides the same bilinear approximation rate as Eq. (10) in certain Nikol'skii–Besov classes, but for separations of the variables into two equally sized groups only [39].<sup>3</sup>

The rest of the paper is organized as follows. In Section 2 we recall the results on bilinear approximation we will use. We also show the link to the theory of operator and sequence ideals. In Section 3 we recall the high-order SVD truncation, the hierarchical tensor format, and the hierarchical tensor rank. Finally, in Section 4 the results on approximation rates are stated and proved.

## 2. Results on bilinear approximation

We keep the notation from the introduction. Let  $d \geq 2$ ,  $1 \leq a < d$ ,  $\mathbf{x} = (x_1, \dots, x_a)$ , and  $\mathbf{y} = (y_1, \dots, y_{d-a})$ . For a function  $f \in L_2(\pi_d)$  we define

$$\tau_r(f, a) = \inf \left\| f(\mathbf{x}, \mathbf{y}) - \sum_{k=1}^r u_k(\mathbf{x})v_k(\mathbf{y}) \right\|_{L_2} \tag{9}$$

where the infimum is taken over all sets of functions  $u_1, \dots, u_r \in L_2(\pi_a)$  and  $v_1, \dots, v_r \in L_2(\pi_{d-a})$ . Temlyakov [47] proved

$$\sup_{f \in B^s} \tau_r(f, a) \sim r^{-s \max(1/a, 1/(d-a))} \quad (r \rightarrow \infty). \tag{10}$$

In fact, in the cited paper refined results for different  $L_p$  norms and anisotropic regularity are given. In that case the order of decay is  $\max((1/s_1 + \dots + 1/s_a)^{-1}, (1/s_{a+1} + \dots + 1/s_d)^{-1})$ . For related results in the non-periodic case see [1].

For the spaces of dominating mixed smoothness, to the best of our knowledge, results are available for the case of even  $d$  and  $a = d/2$  only. In that case, Temlyakov [44,46] has shown<sup>4</sup>

$$\sup_{f \in B^{s, \text{mix}}} \tau_r(f, d/2) \sim r^{-2s} (\log r)^{2s(d/2-1)} \quad (r \rightarrow \infty). \tag{11}$$

<sup>3</sup> This reference was kindly pointed out by an anonymous referee after submission of the paper.

<sup>4</sup> In Theorem 4.3 of [44] the result is stated for  $s > 1/2$ . The result for  $s > 0$  is a remark to Theorem 2.1 in [46] which even deals with the case that the two blocks of variables possess different mixed regularity within themselves. Then  $s$  is to be replaced with  $(s_a + s_{d-a})/2$ . One should also note that Theorem 2.1 in [46] contains the typo  $2 < \mathbf{q}$  instead of  $2 \leq \mathbf{q}$ . Again, in these papers slightly different Sobolev classes are considered which however have the same asymptotic, cf. Footnote 2 (here one gains negative logarithmic powers in the lower dimensions).

However, we can apply this result to functions of  $2a$  and  $2(d - a)$  variables, which gives

$$r^{-2s}(\log r)^{2s(\min(a,d-a)-1)} \lesssim \sup_{f \in B^s, \text{mix}} \tau_r(f, a) \lesssim r^{-2s}(\log r)^{2s(\max(a,d-a)-1)} \quad (r \rightarrow \infty) \tag{12}$$

for arbitrary  $a$  and  $d$ . To see this, let  $1 \leq d_1 < d_2 \leq d_3$ , and let  $V$  denote the subspace of  $L_2(\pi_{d_3-d_1})$  containing functions which depend only on the first  $d_2 - d_1$  variables, that is, whose Fourier coefficients of index  $\mathbf{k} \in \mathbb{Z}^{d_3-d_1}$  are zero if some  $k_\nu \neq 0$  for  $\nu > d_3 - d_1$ . Then it is plain that  $L_2(\pi_{d_3}) = (L_2(\pi_{d_1}) \otimes V) \oplus (L_2(\pi_{d_1}) \otimes V^\perp)$  (in the sense of Hilbert spaces [20]) and that  $L_2(\pi_{d_2})$  is isometrically isomorphic to  $L_2(\pi_{d_1}) \otimes V$ . Moreover, the best bilinear approximation of a such embedded  $f \in L_2(\pi_{d_1}) \otimes V$  with respect to variable splitting after index  $d_1$  has to belong to  $L_2(\pi_{d_1}) \otimes V$  as well, which can be seen by applying the  $L_2(\pi_{d_3})$  orthogonal projection  $I \otimes P_V$  where  $P_V$  is the orthogonal projection on  $V$  in  $L_2(\pi_{d_3-d_1})$ . Also, by definition (1),  $f$  will have the same mix norm when regarded as a function of  $d_2$  or of  $d_3$  variables (since the Fourier basis of trigonometric polynomials is a tensor product basis). Summarizing, it holds

$$\sup_{f \in B^s, \text{mix}(\mathbb{R}^{d_2})} \tau_r(f, d_1) = \sup_{f \in (L_2(\pi_{d_1}) \otimes V) \cap B^s, \text{mix}(\mathbb{R}^{d_3})} \tau_r(f, d_1) \leq \sup_{f \in B^s, \text{mix}(\mathbb{R}^{d_3})} \tau_r(f, d_1).$$

Using (11) (and noting that  $\tau_r(f, a) = \tau_r(f, d - a)$ ), the lower bound in (12) is obtained by taking  $d_1 = \min(a, d - a)$ ,  $d_2 = 2d_1$ , and  $d_3 = d$ , while for the upper estimate we choose  $d_1 = \max(a, d - a)$ ,  $d_2 = d$ , and  $d_3 = 2d_1$ .

**Lemma 1.** For  $c, s, d > 0$ , let  $r(\epsilon)$  denote the smallest  $r$  such that  $cr^{-s}(\log r)^{sd} \leq \epsilon$ . Then it holds

$$r(\epsilon) \leq C(d, \epsilon)c^{1/s}s^{-d}\epsilon^{-1/s}|\log \epsilon|^d$$

with  $C(d, \epsilon) \searrow 1$  for  $\epsilon \rightarrow 0$ , but  $C(d, \epsilon) \rightarrow +\infty$  superexponentially for  $d \rightarrow \infty$ .

**Proof.** It suffices to prove this for  $c = s = 1$  and then invoking  $r((\epsilon/c)^{1/s})$  for the general case. For  $\rho = C\epsilon^{-1}|\log \epsilon|^d$  with  $C > 0$  it holds

$$\begin{aligned} \rho^{-1}(\log \rho)^d &= C^{-1}\epsilon|\log \epsilon|^{-d}(\log C + |\log \epsilon| + d \log |\log \epsilon|)^d \\ &= C^{-1} \left( 1 + \frac{\log C}{|\log \epsilon|} + \frac{d \log |\log \epsilon|}{|\log \epsilon|} \right)^d \epsilon. \end{aligned}$$

Choosing  $C = C(d, \epsilon)$  as the smallest constant such that the right hand side is less than or equal to  $\epsilon$  provides the required estimate, since  $r(\epsilon) = \lfloor \rho \rfloor$  then.  $\square$

Let  $c(d, s)$  be the constant for the upper bound of (12). Substituting  $s$  by  $2s$  and  $d$  by  $\max(a, d - a) - 1$  in Lemma 1, we obtain the following result.

**Lemma 2.** Let  $r(\epsilon)$  denote the smallest rank  $r$  such that  $\sup_{f \in B^s, \text{mix}} \tau_r(f, a) \leq \epsilon$ . Then it holds

$$r(\epsilon) \leq C(d, \epsilon)c(d, s)^{1/s}(2s)^{-(\max(a,d-a)-1)}\epsilon^{-1/(2s)}|\log \epsilon|^{\max(a,d-a)-1}$$

when  $\epsilon$  is small enough to satisfy the upper bound of (12) with constant  $c(d, s)$ . It holds  $C(d, \epsilon) \searrow 1$  for  $\epsilon \rightarrow 0$ .

### 2.1. The link to the theory of operator ideals

Given  $f$  and  $a$  as above, we denote by  $A_f$  the integral operator

$$(A_f w)(\mathbf{x}) = \int_{[-\pi, \pi]^{d-a}} f(\mathbf{x}, \mathbf{y}) \overline{w(\mathbf{y})} \, d\mathbf{y}. \tag{13}$$

This operator is compact, in fact Hilbert–Schmidt. The best bilinear approximation problem is equivalent to

$$\|A_f - A\|_{HS} = \min \quad \text{s.t. rank}(A) \leq r.$$

This problem can formally be solved by means of the singular value decomposition (Schmidt expansion [41])

$$A_f = \sum_{k=1}^{\infty} \sigma_k u_k \otimes v_k, \quad \sigma_1 \geq \sigma_2 \geq \dots \geq 0, \tag{14}$$

of  $A_f$ , where  $u_1, u_2, \dots$  and  $v_1, v_2, \dots$  are orthonormal systems in  $L_2(\pi_a)$  and  $L_2(\pi_{d-a})$ , respectively, and  $u_k \otimes v_k$  denotes rank-one operators with kernels  $u_k(\mathbf{x})v_k(\mathbf{y})$ . Namely, we have for  $A_f^r = \sum_{k=1}^r \sigma_k u_k \otimes v_k$  that

$$\tau_r(f, a) = \|A_f - A_f^r\|_{HS} = \left( \sum_{k=r+1}^{\infty} \sigma_k^2 \right)^{1/2}. \tag{15}$$

Thus singular value estimates lead to approximation rates.

For an  $\ell_2$ -sequence  $(\varsigma_k)$  we denote by  $(\varsigma_k^*)$  a rearrangement which is decreasing in modulus. The weak  $l_p$ -space  $\ell_p^w$  contains by definition all sequences  $(\varsigma_k)$  for which there exists a constant  $c$  such that  $\varsigma_k^* \leq ck^{-1/p}$ . As one can think about,  $\ell_p \subset \ell_p^w$ , and  $\ell_p^w$  is compactly embedded in  $\ell_q$  for  $q > p$ . A standard result in approximation theory is that for  $s > 0$  we have

$$\left( \sum_{k>r} (\varsigma_k^*)^2 \right)^{1/2} \lesssim r^{-s} \quad \text{if and only if} \quad (\varsigma_k) \in \ell_{p(s)}^w \quad \text{with} \quad p(s) = (s + 1/2)^{-1},$$

see, e.g., in DeVore [10] or Dahmen [7, Proposition 16] for a proof. In the special case that  $(\varsigma_k) \in \ell_{p(s)}$  we have the following estimate which is frequently called Stechkin’s lemma<sup>5</sup>:

$$\left( \sum_{k>r} (\varsigma_k^*)^2 \right)^{1/2} \leq r^{-s} \|(\varsigma_k)\|_{p(s)}.$$

We see from (15) that

$$\tau_r(f, a) \sim r^{-s} \quad \text{if and only if} \quad (\sigma_k) \in \ell_{p(s)}^w,$$

which is in line with singular value estimates also found by Temlyakov in the cited and other papers. Compact operators whose singular values lie in certain weak  $l_p$ -spaces form ideals. This follows from

$$\sigma_k(AB) \leq \sigma_k(A) \|B\|, \tag{16}$$

where  $A$  and  $B$  are bounded operators between Banach spaces, and  $A$  is compact. The definition of the singular numbers in this case is  $\sigma_k(A) = \inf \|A - A_k\|$ ,  $\text{rank} A_k < k$ , from which (16) easily follows. Among those ideals the Schatten classes [40] of compact operators whose singular values belong to strong  $\ell_p$  spaces are the most famous. Let us also refer to the classic books of Gohberg and Krein [16] and Pietsch [37].

As for singular value estimates of integral operators with kernels from Sobolev and Besov classes standard references seem to be the surveys of Birman and Solomjak [6] and Pietsch’s monograph [38], with the latter also containing an excellent historical survey. There is also a nice chapter in the book of König [28]. The typical approach in these references is to use the fact that, if a kernel belongs to, say, a Sobolev space with respect to variable  $\mathbf{x}$ , then the integral operator (13) can be regarded as a continuous mapping into a lower-variate Sobolev space. Using then (16) with  $A$  being a compact Sobolev embedding and  $B$  the integral operator, the problem reduces to estimation of singular values of Sobolev embeddings. New bounds have been obtained recently [30]. However, it seems to be difficult with this approach to take the regularity with respect to the variable  $\mathbf{y}$  into account.

<sup>5</sup> We were unable to find a reference where exactly this estimate has been stated by Stechkin. However, it is closely related to his pioneering results on  $N$ -term approximation in the case  $p(s) = 1$  [43], see also [11]. A direct elementary proof of the stated inequality can be found in [29].

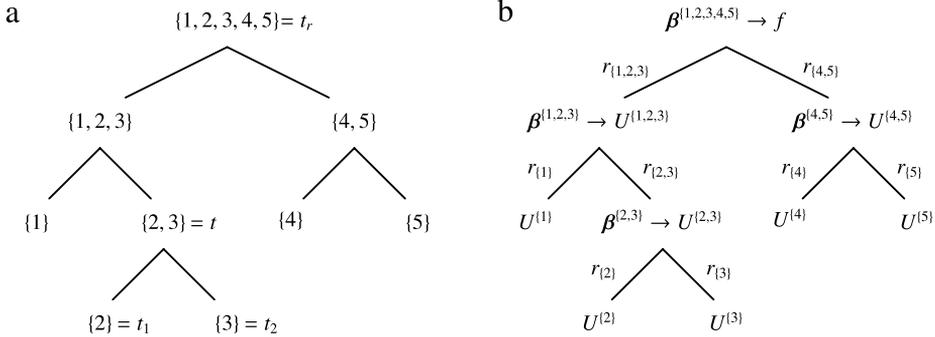


Fig. 1. (a) A binary dimension tree for {1, 2, 3, 4, 5}, (b) parameters of the (T, r)-decomposition.

We are only aware of one such result which gives an upper bound for “block mixed” spaces. Let  $H^{s_a}(\pi_a) \otimes H^{s_{d-a}}(\pi_{d-a})$  denote the space of  $2\pi$ -periodic functions  $f(\mathbf{x}, \mathbf{y})$  having weak derivatives up to order  $s_a, s_{d-a}$ , but only order  $s_a$  for the first block, and  $s_{d-a}$  for the second block are allowed. Then, if  $2s_a > a$  and  $2s_{d-a} > d - a$ , it holds [6, Theorems 2.5/9.4] that

$$\sigma_k(A_f) \leq Ck^{-1/p} \|f\|_{s_a, s_{d-a}, \text{mix}} \quad \text{with } 1/p = 1/2 + s_a/a + s_{d-a}/(d - a),$$

where  $\|f\|_{s_a, s_{d-a}, \text{mix}} = \|g\|_{s_{d-a}}, g(\mathbf{y}) = \|f(\cdot, \mathbf{y})\|_{s_a}$ , and  $C$  does not depend on  $f$ . According to the previous considerations this implies

$$\sup_{\|f\|_{s_a, s_{d-a}, \text{mix}} \leq 1} \tau_r(f, a) \lesssim r^{-s_a/a - s_{d-a}/(d-a)} \quad (r \rightarrow \infty).$$

In this statement we have an inbuilt, probably unnecessary growing smoothness requirement with the dimension which ensures compactness of certain Sobolev embeddings. For  $d = 2$  and  $s_a = s_{d-a} > 1/2$  one recovers the upper bound in (11).

### 3. The hierarchical tensor format, hierarchical rank, and HOSVD truncation

Here we recall the apparatus of the hierarchical tensor format, mainly following [51], but allowing for non-binary trees. More details can be found in Hackbusch’s monograph [20] and Grasedyck’s paper [17].

Let  $T$  denote a set of subsets of  $\{1, 2, \dots, d\}$  which can be arranged into a tree as follows: (i) the root is  $t_r = \{1, 2, \dots, d\} \in T$ , (ii) every node  $t \in T$  that is not a leaf has at least two nonempty sons  $t_1, t_2, \dots, t_{n_t} \in T$  such that  $t = t_1 \cup t_2 \cup \dots \cup t_{n_t}$  is a disjoint union, (iii) the leaves are  $\{\mu\}, \mu = 1, 2, \dots, d$ . Such a set  $T$  is called a *dimension tree*. Examples are shown in Figs. 1(a), and 2. It is convenient to also introduce  $t^c = \{1, 2, \dots, d\} \setminus t$ . Let us agree that in tuples  $\mathbf{x}^t = (x_{v_1}, x_{v_2}, \dots, x_{v_{|t|}})$ , where  $t = \{v_1, v_2, \dots, v_{|t|}\}$ , the variables are arranged by increasing index, i.e.,  $v_1 < v_2 < \dots < v_{|t|}$ . Likewise for  $\mathbf{x}^{t^c}$ , whereas in  $(\mathbf{x}^t, \mathbf{x}^{t^c})$  we assume all variables being merged into their initial ordering.

Let  $T$  be a dimension tree and  $\mathbf{r} = (r_t)_{t \in T \setminus \{t_r\}}$  an associated set of nonnegative values  $r_t \in \mathbb{N} \cup \{+\infty\}$ . For convenience let  $r_{t_r} = 1$  for the root. A function  $f \in L_2(\pi_d)$  is said to be  $(T, \mathbf{r})$ -decomposable if it can be decomposed in the following form.

- (i) To every node  $t \in T \setminus \{t_r\}$  an  $r_t$ -dimensional subspace  $U^t \subset L_2(\pi_{|t|})$  is associated in form of a basis  $u_1^t, u_2^t, \dots, u_{r_t}^t$ . For the root let  $u_1^{t_r} = f$ .
- (ii) For every node  $t \in T$  having sons  $t_1, t_2, \dots, t_{n_t}$  there exists a transfer tensor  $\beta^t \in \mathbb{R}^{r_t \times r_{t_1} \times r_{t_2} \times \dots \times r_{t_{n_t}}}$  such that it holds

$$u_k^t(\mathbf{x}^{t_1}, \mathbf{x}^{t_2}, \dots, \mathbf{x}^{t_{n_t}}) = \sum_{k_1=1}^{r_{t_1}} \sum_{k_2=1}^{r_{t_2}} \dots \sum_{k_{n_t}=1}^{r_{t_{n_t}}} \beta_{k, k_1, k_2, \dots, k_{n_t}}^t u_{k_1}^{t_1}(\mathbf{x}^{t_1}) u_{k_2}^{t_2}(\mathbf{x}^{t_2}) \dots u_{k_{n_t}}^{t_{n_t}}(\mathbf{x}^{t_{n_t}}). \quad (17)$$

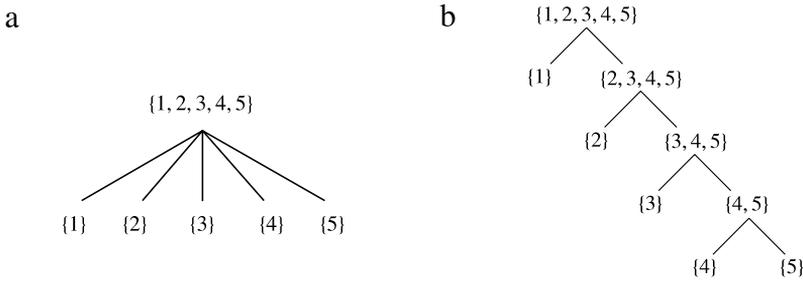


Fig. 2. (a) Tucker format, (b) (extended) tensor train format.

The second condition can be expressed as

$$U^t \subseteq U^{t_1} \otimes U^{t_2} \otimes \dots \otimes U^{t_{n_t}}$$

and is called *nestedness* [17].

The set of  $(T, \mathbf{r})$ -decomposable functions will be denoted by  $\mathcal{H}_{\leq \mathbf{r}, T}$ . In the cases of the dimension trees depicted in Fig. 2, the format is called *Tucker format* and *extended tensor train format* (which is essentially different from the ordinary TT format, see, e.g., [51]), respectively. In general, we speak about the *hierarchical Tucker format*.

The parameters that need to be stored for a function  $f \in \mathcal{H}_{\leq \mathbf{r}, T}$  are the basis functions  $u_{k_t}^t$  for the leaves and the transfer tensors  $\beta^t$  for all other nodes of the tree. We have illustrated this in Fig. 1(b) for a binary tree. Thus, as long as the transfer tensors do not possess a specific structure, the storage cost is

$$\sum_{t \in T \text{ not leaf}} r_t \prod_{i=1}^{n_t} r_{t_i} + \sum_{\mu=1}^d \text{cost to store approximations of } u_1^{[\mu]}, u_2^{[\mu]}, \dots, u_{r(\mu)}^{[\mu]}. \tag{18}$$

One should remark that typically, to ensure the computational stability of the recursion, the bases  $u_1^t, u_2^t, \dots, u_{r_t}^t$  at each node are chosen to be orthonormal. In this case, every “slice” of the transfer tensors  $\beta^t$  (fixed index  $k_t$ ) has Frobenius norm 1, except in the root  $t_r$ .

We now define the hierarchical  $T$ -rank of a function  $f \in L_2(\pi_d)$  associated to the tree  $T$ . For every  $t \in T \setminus \{t_r\}$  let  $A_f^t$  be the integral operator

$$(A_f^t v)(\mathbf{x}^t) = \int_{[-\pi, \pi]^{|t^c|}} f(\mathbf{x}^t, \mathbf{x}^{t^c}) v(\mathbf{x}^{t^c}) d\mathbf{x}^{t^c}$$

and

$$U_f^t = \mathcal{R}(A_f^t)$$

its range. The  $T$ -rank of  $f$  (notation  $\text{rank}_T(f)$ ) is the tuple  $\mathbf{r} = (r_t)_{t \in T \setminus \{t_r\}}$  with

$$r_t = \text{rank}(A_f^t) = \dim(U_f^t),$$

which can be countably infinite. Note that the two sons of the root of  $T$  have identical ranks.

When  $f$  is  $(T, \mathbf{r})$ -decomposable it is easy to see that  $U_f^t \subseteq U^t$  and hence  $\text{rank}_T(f) \leq \mathbf{r}$  (understood elementwise) is a necessary condition for  $f \in \mathcal{H}_{\leq \mathbf{r}, T}$ . It is also sufficient since we can choose  $U^t = U_f^t$  in every node which follows from the nestedness property

$$U_f^t \subseteq U_f^{t_1} \otimes U_f^{t_2} \otimes \dots \otimes U_f^{t_{n_t}}$$

whose easy proof we can omit. Consequently, we have

$$\mathcal{H}_{\leq \mathbf{r}, T} = \{f \in L_2(\pi_d) : \text{rank}_T(f) \leq \mathbf{r}\}.$$

Let us now come to the problem of approximation by  $(T, \mathbf{r})$ -decomposable functions. The main tool is the high order-singular value truncation. The following facts can be found in [17] or [20]. Let

$$A_f^t = \sum_{k=1}^{\infty} \sigma_k^t u_k^t \otimes v_k^{t^c}, \quad \sigma_1^t \geq \sigma_2^t \geq \dots \geq 0,$$

be the singular value decomposition of  $A_f^t$  as in (14). Given  $r_t$ , let  $U_f^{t,r_t} = \text{span}\{u_1^t, u_2^t, \dots, u_{r_t}^t\}$  and  $P_f^{t,r_t}$  be the orthogonal projection onto  $U_f^{t,r_t} \otimes L_2(\pi_{|t^c|})$ . Note that  $P_f^{t,r_t}(f)$  is the best bilinear rank- $r_t$  approximation of  $f$  with respect to the splitting  $(\mathbf{x}^t, \mathbf{x}^{t^c})$ .

Assume  $T$  has  $L$  levels (the root having level zero) and denote by  $\text{level}(t)$  the level of a node. Then the high-order SVD truncation/projection  $f^{\mathbf{r}}$  is defined as the projection

$$f^{\mathbf{r}} = P_f^{\mathbf{r}}(f), \quad \text{where } P_f^{\mathbf{r}} = P_{f,L}^{\mathbf{r}} P_{f,L-1}^{\mathbf{r}} \dots P_{f,1}^{\mathbf{r}}, \quad \text{and } P_{f,l}^{\mathbf{r}} = \prod_{\text{level}(t)=l} P_f^{t,r_t}.$$

The ordering within one level does not matter since for different values of  $t$  in the same level the projections  $P_f^{t,r_t}$  commute. Thus every  $P_{f,l}^{\mathbf{r}}$  itself is an orthogonal projection.

It is easily seen that  $f^{\mathbf{r}} \in \mathcal{H}_{\leq \mathbf{r}, T}$ . Note that from the second projection on, the performed bilinear approximations are not optimal anymore. This can be fixed by an on-line choice of the projection to use (calculated from the current approximation). We do not dwell on this because, astonishingly,  $f^{\mathbf{r}}$  itself is already a quasi optimal approximation from  $\mathcal{H}_{\leq \mathbf{r}, T}$ . The result is due to De Lathauwer et al. [8] for the Tucker format and Grasedyck [17] for the general case. (Oseledets and Tyrtshnikov [36] came to a similar conclusion for the TT format.)

**Lemma 3.** *Let*

$$\tau_{\mathbf{r}}(f, T) = \inf_{g \in \mathcal{H}_{\leq \mathbf{r}, T}} \|f - g\|_{L_2}, \tag{19}$$

then the following holds:

$$\tau_{\mathbf{r}}(f, T) \leq \|f - f^{\mathbf{r}}\| \leq \sqrt{\sum_{t \in T \setminus \{t_{\mathbf{r}}\}} \sum_{k_t \geq r_t + 1} (\sigma_{k_t}^t)^2} \leq \sqrt{|T| - 1} \tau_{\mathbf{r}}(f, T). \tag{20}$$

Note that for each  $t$  the sum  $\sum_{k_t \geq r_t + 1} (\sigma_{k_t}^t)^2$  is the exact error of the best bilinear rank- $r_t$  approximation of  $f(\mathbf{x}^t, \mathbf{x}^{t^c})$ . In the case that the root has two sons, the constant can be improved to  $\sqrt{|T| - 2}$ .

The HOSVD subspaces  $U_f^{t,r_t}$  are spanned by the first  $r_t$  left singular vectors  $u_k^t$  of the integral operator  $A_f^t$  which are eigenvectors of  $A_f^t (A_f^t)^*$ . Typically, at each node these singular vectors are taken as basis functions for  $U_f^{t,r_t}$  in a  $(T, \mathbf{r})$ -decomposition of  $f^{\mathbf{r}}$ . This is called the HOSVD basis. In the case that the root has two sons, the transfer tensor  $\beta^{\mathbf{r}}$  is a diagonal matrix.

In practice, we will have to approximate the HOSVD basis functions in the leaves based on their regularity. If the kernel  $f$  is in  $H^s(\pi_d)$ , then  $A_f^t$  will be a continuous map into  $H^s(\pi_{|t|})$ . Since the HOSVD basis functions are in the range of  $A_f^t$ , they belong to that space. In particular, it is straightforward to show that

$$\|u_k^{\{\mu\}}\|_s = \|A_f^t v_k^{\{\mu\}^c} / \sigma_k^{\{\mu\}}\|_s \leq \|f\|_{s, \mu} / \sigma_k^{\{\mu\}} \tag{21}$$

for the leaves, see, e.g., [49] for a proof.

Concerning the regularity of  $f^{\mathbf{r}}$ , Hackbusch [21, Corollary 3.2] proved by an elegant technique that

$$\|f^{\mathbf{r}}\|_s \leq \|f\|_s. \tag{22}$$

#### 4. Results on asymptotic complexity

We first state the asymptotic upper bound for  $\tau_{\mathbf{r}}(f, T)$  (defined in (19)) with respect to  $\mathbf{r}$ .

**Theorem 1** (Asymptotics With Respect to Rank). *For a fixed dimension tree  $T$ , we have the following asymptotics:*

$$\sup_{f \in B^s} \tau_{\mathbf{r}}(f, T) \sim \max_{t \in T \setminus \{t_r\}} r_t^{-s \max(1/|t|, 1/(d-|t|))} \quad (\mathbf{r} \rightarrow \infty),$$

and

$$\begin{aligned} \max_{t \in T \setminus \{t_r\}} r_t^{-2s} (\log r_t)^{2s(\min(|t|, d-|t|)-1)} &\lesssim \sup_{f \in B^s, \text{mix}} \tau_{\mathbf{r}}(f, T) \\ &\lesssim \max_{t \in T \setminus \{t_r\}} r_t^{-2s} (\log r_t)^{2s(\max(|t|, d-|t|)-1)} \quad (\mathbf{r} \rightarrow \infty). \end{aligned}$$

The hidden constants depend on  $d, s$ , and  $T$ .

**Proof.** By (20), it holds

$$C_1 \max_{t \in T \setminus \{t_r\}} \sum_{k \geq r_t+1} (\sigma_k^t)^2 \leq \tau_{\mathbf{r}}(f, T)^2 \leq C_2 \max_{t \in T \setminus \{t_r\}} \sum_{k \geq r_t+1} (\sigma_k^t)^2$$

for constants depending only on  $T$ . Since  $\sum_{k \geq r_t+1} (\sigma_k^t)^2 = \tau_{r_t}(f, |t|)$  in (9) for a suitable rearrangement of the variables, and since the maximum is taken over a finite set, we immediately obtain

$$C_1 \max_{t \in T \setminus \{t_r\}} \sup_{f \in B} \tau_{r_t}(f, |t|) \leq \sup_{f \in B} \tau_{\mathbf{r}}(f, T) \leq C_2 \max_{t \in T \setminus \{t_r\}} \sup_{f \in B} \tau_{r_t}(f, |t|) \quad (23)$$

for every  $\mathbf{r}$  for which  $\mathcal{H}_{\leq \mathbf{r}, T}$  is not empty. Here  $B$  stands either for  $B^s$  or for  $B^s, \text{mix}$ . The claims follow from (10) and (12).  $\square$

Now let  $c(f, \epsilon)$  denote the minimum storage cost (required degrees of freedom) for approximating a function  $f$  to an  $L_2$ -accuracy  $\epsilon$  by  $(T, \mathbf{r})$ -decomposable functions. In the estimates below the storage cost for the transfer tensors  $\beta^t$  is just estimated by their size. We leave it open whether there is some further structure to be exploited. For the leaves we have to determine the dimension required to approximate the exact HOSVD bases by a fixed linear method, e.g., approximation by trigonometric polynomials.

**Lemma 4** (Approximation of Bases in Leaves). *Let  $f \in B^s \cap \mathcal{H}_{\leq \mathbf{r}, T}$  and  $\epsilon > 0$ . Then there exists  $\tilde{f} \in \mathcal{H}_{\leq \mathbf{r}, T}$  with  $\|f - \tilde{f}\|_{L_2} \leq \epsilon$  such that  $\tilde{f}$  can be encoded with the basis functions in leaf  $\{\mu\}$  being trigonometric polynomials of degree  $\lesssim d^{1/s} r_{\{\mu\}}^{1/(2s)} \epsilon^{-1/s}$ . Thus it follows from (18) that*

$$c(f, \epsilon) \lesssim \sum_{t \in T \text{ not leaf}} r_t \prod_{i=1}^{n_t} r_{t_i} + d^{1/s} \epsilon^{-1/s} \sum_{\mu=1}^d r_{\{\mu\}}^{1+1/(2s)} \quad (\epsilon \rightarrow 0).$$

**Proof.** Assume  $f$  is encoded using HOSVD bases  $u_k^t$  in all nodes. By (2) and (21), we can approximate the basis functions  $u_k^{\{1\}}$  by their best approximations by trigonometric polynomials, denoted by  $\tilde{u}_k^{\{1\}}$ , to an accuracy  $\epsilon r_{\{1\}}^{-1/2} / (d \sigma_{k(1)}^{\{1\}})$  using polynomial degree

$$n \lesssim (\epsilon/d)^{-1/s} r_{\{1\}}^{1/(2s)}.$$

Keeping the bases in the other leaves, and using the same transfer tensors, we obtain a reconstruction  $f^{(1)} \in \mathcal{H}_{\leq \mathbf{r}, T}$ . Denoting by  $\sum_{k=1}^{r_1} \sigma_k^{\{1\}, \{1\}} u_k^{\{1\}} \otimes v_k$  an SVD of  $A_f^{\{1\}}$ , it follows from the hierarchical decomposition (17) that the  $v_k$  are multilinear combinations of the basis functions in leaves  $\{2\}$  to

$\{d\}$  only, involving the transfer tensors as coefficients. Replacing the  $u_k^{(1)}$  in the first leaf by  $\tilde{u}_k^{(1)}$  we find that it has to hold that

$$A_{f^{(1)}}^{(1)} = \sum_{k=1}^{r_1} \sigma_k^{(1)} \tilde{u}_k^{(1)} \otimes v_k. \tag{24}$$

Hence, since the  $v_k$  are orthonormal,

$$\|A_f^{(1)} - A_{f^{(1)}}^{(1)}\|_{HS}^2 = \left\| \sum_{k=1}^{r_1} \sigma_k^{(1)} (u_k^{(1)} - \tilde{u}_k^{(1)}) \otimes v_k \right\|_{HS}^2 = \sum_{k=1}^{r_1} \sigma_k^2 \|u_k^{(1)} - \tilde{u}_k^{(1)}\|_{L_2}^2 \leq (\epsilon/d)^2.$$

By isometry,  $\|f - f^{(1)}\|_{L_2} \leq \epsilon/d$ .<sup>6</sup> Since the  $\tilde{u}_k^{(1)}$  are obtained from  $u_k^{(1)}$  by Fourier truncation, (24) also shows that  $f^{(1)}$  is a Fourier truncation of  $f$ . In particular, it holds  $f^{(1)} \in B^s$ , so that we can continue with the further leaves in the following way: before step  $\mu + 1$  we change the representation of  $f^{(\mu)}$  into one using HOSVD bases of  $f^{(\mu)}$  in every node, and then manipulate the basis at node  $\{\mu + 1\}$  as in the first step such that  $\|f^{(\mu)} - f^{(\mu+1)}\|_{L_2} \leq \epsilon/d$ . Note that establishing HOSVD bases in all nodes does neither increase the  $T$ -rank, nor does it change the optimal subspaces in the leaves (simply because it does not change the represented function). Therefore the HOSVD bases for the leaves already treated remain trigonometrical polynomials of at most the same degree. The final result of this process is  $\tilde{f} = f^{(d)} \in \mathcal{H}_{\leq r, T}$  satisfying

$$\|f - \tilde{f}\|_{L_2} \leq \|f - f^{(1)}\|_{L_2} + \|f^{(1)} - f^{(2)}\|_{L_2} + \dots + \|f^{(d-1)} - \tilde{f}\|_{L_2} \leq \epsilon$$

as required.  $\square$

We now state the main result. For every node  $t \in T$  let  $\text{deg}(t)$  be the number of connected nodes (number of sons plus number of parents (one or zero)). We set

$$\text{deg}(T) = \max_{t \in T} \text{deg}(t).$$

In the estimates below, the first term in the max function is a bound for the transfer tensors, while the second is the approximation of the HOSVD bases.

**Theorem 2** (Asymptotic Rate of Approximation Cost).

(a) It holds

$$\sup_{f \in B^s} c(f, \epsilon) \lesssim \max(\epsilon^{-d/s}, \epsilon^{-(2+1/(2s))/s}) \quad (\epsilon \rightarrow 0)$$

with a constant depending on  $d, s$ , and  $T$ .

(b) When  $d = 2$ ,

$$\sup_{f \in B^s, \text{mix}} c(f, \epsilon) \lesssim \epsilon^{-(3+1/(2s))/(2s)} \quad (\epsilon \rightarrow 0),$$

whereas for  $d \geq 3$  at most

$$\sup_{f \in B^s, \text{mix}} c(f, \epsilon) \lesssim \max(\epsilon^{-\text{deg}(T)/(2s)} |\log \epsilon|^{N(T)}, \epsilon^{-(3+1/(2s))/(2s)} |\log \epsilon|^{(1+1/(2s))(d-2)}) \quad (\epsilon \rightarrow 0)$$

with

$$N(T) = \max_{t \in T \text{ not leaf}} \left( \max(|t|, d - |t|) + \sum_{i=1}^{n_t} \max(|t_i|, d - |t_i|) \right) - \text{deg}(T),$$

with constants depending on  $d, s$ , and  $T$ . These constants improve for growing  $s$  and fixed  $d$  and  $T$ . In both cases, the degrees of freedom needed for approximating the HOSVD bases in the leaves is asymptotically bounded by  $\epsilon^{-(2+1/(2s))/s} |\log \epsilon|^{(1+1/(2s))(d-2)}$ .

<sup>6</sup> A rigorous and detailed proof of precisely this perturbation bound is given in [20, Theorem 11.53].

**Remark 1.** For a binary dimension tree (Fig. 1) and  $d \geq 3$  we have  $\text{deg}(T) = 3$  and thus for  $f \in B^{s, \text{mix}}$  always obtain the second bound

$$\sup_{f \in B^{s, \text{mix}}} c(f, \epsilon) \lesssim \epsilon^{-(3+1/(2s))/(2s)} |\log \epsilon|^{(1+1/(2s))(d-2)} \quad (\epsilon \rightarrow 0),$$

which is the bound for approximating the basis functions in the leaves. This is in line with the findings of Griebel and Harbrecht [19] for two variables (a simple binary tree of depth one).

**Remark 2.** It appears plausible, yet difficult to prove, that the cost for approximating the HOSVD bases by a linear method as done in the proof of Lemma 4 cannot be reduced in general. But even if this is true, it would not necessarily imply that the found rates are optimal for approximation by the hierarchical tensor format (they might be optimal when restricting to the HOSVD algorithm). In fact, this seems unlikely as can be seen from the bi-variate case, where the optimal linear approximation by trigonometric polynomials actually is a bilinear approximation whose rank might not be optimal, but which has an asymptotically lower cost  $\epsilon^{-1/s} |\log \epsilon|$ . How to transfer this counter-example to higher dimension is however not clear to us.

**Proof of Theorem 2.** We again make use of the fact that  $\tau_r(f, T) \sim \|f - f^r\|_{L_2}$  for  $r \rightarrow \infty$  (Lemma 3). Let  $r_t(\epsilon/2)$  be the smallest ranks required to achieve  $\|f - f^r\|_{L_2} \leq \epsilon/2$ . From (23) and (10) we see that we have

$$r_t(\epsilon/2) \lesssim \epsilon^{-\min(|t|, d-|t|)/s} \quad (\epsilon \rightarrow 0)$$

for  $f \in B^s$ . For  $f \in B^{s, \text{mix}}$  we apply Lemma 2 to get from (23) that

$$r_t(\epsilon/2) \lesssim \epsilon^{-1/(2s)} |\log \epsilon|^{\max(|t|, d-|t|)-1}$$

with a constant depending on  $d, s$ , and  $T$ . When  $d = 2$  we simply have  $r(\epsilon/2) \lesssim \epsilon^{-1/(2s)}$ . The behavior of this constant with regard to  $s$  has been stated in Lemma 2. In both cases it holds that  $\|f^r\|_s \leq \|f\|_s \leq 1$  according to (22). Therefore we can apply Lemma 4 with  $\epsilon/2$  for further approximating the basis functions in the leaves. For  $f \in B^s$  this provides an overall estimate of

$$c(f, \epsilon) \lesssim \sum_{t \in T \text{ not leaf}} \epsilon^{-M(t)/s} + (2d)^{1/s} \epsilon^{-1/s} \sum_{\mu=1}^d \epsilon^{-(1+1/(2s))/s} \quad (\epsilon \rightarrow 0)$$

with  $M(t) = \min(|t|, d - |t|) + \sum_{i=1}^{n_t} \min(|t_i|, d - |t_i|)$ , while for  $f \in B^{s, \text{mix}}$  we have

$$c(f, \epsilon) \lesssim \sum_{t \in T \text{ not leaf}} \epsilon^{-\text{deg}(t)/(2s)} |\log \epsilon|^{N(t)} + (2d)^{1/s} \epsilon^{-1/s} \sum_{\mu=1}^d \epsilon^{-(1+1/(2s))/(2s)} |\log \epsilon|^{(1+1/(2s))(d-2)} \quad (\epsilon \rightarrow 0)$$

with  $N(t) = \max(|t|, d - |t|) - 1 + \sum_{i=1}^{n_t} (\max(|t_i|, d - |t_i|) - 1)$ . This estimate for  $c(f, \epsilon)$  is also correct for  $d = 2$  although the first term is not sharp then (it can be replaced by  $\epsilon^{-1/(2s)}$  due to the diagonal SVD core matrix). The trivial bound  $M(t) \leq d$  completes the proof.  $\square$

**Acknowledgment**

The authors were supported by the DFG-Priority Program 1324.

**References**

[1] M.B.A. Babaev, Approximation of the Sobolev classes  $W_q^r$  of functions of several variables by bilinear forms in  $L_p$  for  $2 \leq q \leq p \leq \infty$ , Mat. Zametki 62 (1997) 18–34. <http://dx.doi.org/10.1007/BF02356060> (in Russian); Math. Notes 62 (1992) 15–29. English translation.

- [2] K.I. Babenko, Approximation by trigonometric polynomials in a certain class of periodic functions of several variables, Dokl. Akad. Nauk SSSR 132 (1960) 982–985 (in Russian); Sov. Math. Dokl. 1 (1960) 672–675. English translation.
- [3] K.I. Babenko, Approximation of periodic functions of many variables by trigonometric polynomials, Dokl. Akad. Nauk SSSR 132 (1960) 247–250 (in Russian); Sov. Math Dokl. 1 (1960) 513–516. English translation.
- [4] J. Ballani, Fast evaluation of singular BEM integrals based on tensor approximations, Numer. Math. 121 (2012) 433–460. <http://dx.doi.org/10.1007/s00211-011-0436-6>.
- [5] U. Benedikt, A.A. Auer, M. Espig, W. Hackbusch, Tensor decomposition in post-hartree-fock methods, I. two-electron integrals and mp2, J. Chem. Phys. 134 (2011) 054118. <http://dx.doi.org/10.1063/1.3514201>.
- [6] M.S. Birman, M.Z. Solomjak, Estimates for the singular numbers of integral operators, Uspekhi Mat. Nauk 32 (1977) 17–84. 271.
- [7] W. Dahmen, Multiscale and wavelet methods for operator equations, in: Multiscale Problems and Methods in Numerical Simulations, Springer, Berlin, 2003, pp. 31–96. [http://dx.doi.org/10.1007/978-3-540-39810-3\\_2](http://dx.doi.org/10.1007/978-3-540-39810-3_2).
- [8] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, SIAM J. Matrix Anal. Appl. 21 (2000) 1253–1278. <http://dx.doi.org/10.1137/S0895479896305696>.
- [9] V. De Silva, L.H. Lim, Tensor rank and the ill-posedness of the best low-rank approximation problem, SIAM J. Matrix Anal. Appl. 30 (2008) 1084–1127. <http://dx.doi.org/10.1137/06066518X>.
- [10] R.A. DeVore, Nonlinear approximation, Acta Numer. 7 (1998) 51–150. <http://dx.doi.org/10.1017/S0962492900002816>.
- [11] R.A. DeVore, V.N. Temlyakov, Some remarks on greedy algorithms, Adv. Comput. Math. 5 (1996) 173–187. <http://dx.doi.org/10.1007/BF02124742>.
- [12] S.V. Dolgov, B.N. Khoromskij, Tensor-product approach to global time-space-parametric discretization of chemical master equation, Preprint 68/2012, 2012.
- [13] S.V. Dolgov, B.N. Khoromskij, I.V. Oseledets, Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker–Planck equation, SIAM J. Sci. Comput. 34 (2012) A3016–A3038. <http://dx.doi.org/10.1137/120864210>.
- [14] M. Espig, W. Hackbusch, A. Litvinenko, H.G. Matthies, P. Wähnert, Efficient low-rank approximation of the stochastic galerkin matrix in tensor formats, Comput. Math. Appl. (2012) <http://dx.doi.org/10.1016/j.camwa.2012.10.008> in press.
- [15] L. Giraldi, A. Nouy, G. Legrain, P. Cartraud, Tensor-based methods for numerical homogenization from high-resolution images, Comput. Methods Appl. Mech. Engrg. 254 (2013) 154–169. <http://dx.doi.org/10.1016/j.cma.2012.10.012>.
- [16] I.C. Gohberg, M.G. Krein, Introduction to the Theory of Linear Nonselfadjoint Operators, American Mathematical Society, Providence, RI, 1969.
- [17] L. Grasedyck, Hierarchical singular value decomposition of tensors, SIAM J. Matrix Anal. Appl. 31 (2009/10) 2029–2054. <http://dx.doi.org/10.1137/090764189>.
- [18] L. Grasedyck, D. Kressner, C. Tobler, A literature survey of low-rank tensor approximation techniques, GAMM-Mitt. 36 (1) (2013) 53–78. <http://dx.doi.org/10.1002/gamm.201310004>.
- [19] M. Griebel, H. Harbrecht, Approximation of bi-variate functions: singular value decomposition versus sparse grids, IMA J. Numer. Anal. (2013) <http://dx.doi.org/10.1093/imanum/drs047>, in press.
- [20] W. Hackbusch, Tensor Spaces and Numerical Tensor Calculus, Springer-Verlag, Heidelberg, 2012.
- [21] W. Hackbusch,  $L^\infty$  estimation of tensor truncations, Numer. Math. (2013) <http://dx.doi.org/10.1007/s00211-013-0544-6> in press.
- [22] W. Hackbusch, The use of sparse grid approximation for the  $r$ -term tensor representation, in: J. Garcke, M. Griebel (Eds.), Sparse Grids and Applications, Springer, Berlin, 2013, pp. 151–159. [http://dx.doi.org/10.1007/978-3-642-31703-3\\_7](http://dx.doi.org/10.1007/978-3-642-31703-3_7).
- [23] W. Hackbusch, S. Kühn, A new scheme for the tensor representation, J. Fourier Anal. Appl. 15 (2009) 706–722. <http://dx.doi.org/10.1007/s00041-009-9094-9>.
- [24] B.N. Khoromskij, I.V. Oseledets, DMORG+QTT approach to computation of the ground state for the molecular Schrödinger operator, Preprint 69/2010, MPI MIS, 2010. URL: [http://www.mis.mpg.de/preprints/2010/preprint2010\\_69.pdf](http://www.mis.mpg.de/preprints/2010/preprint2010_69.pdf).
- [25] B.N. Khoromskij, I.V. Oseledets, R. Schneider, Efficient time-stepping scheme for dynamics on TT-manifolds, Preprint 24/2012, MPI MIS, 2012. URL: [http://www.mis.mpg.de/preprints/2012/preprint2012\\_24.pdf](http://www.mis.mpg.de/preprints/2012/preprint2012_24.pdf).
- [26] B.N. Khoromskij, C. Schwab, Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs, SIAM J. Sci. Comput. 33 (2011) 364–385. <http://dx.doi.org/10.1137/100785715>.
- [27] A. Kolmogoroff, Über die beste annäherung von funktionen einer gegebenen funktionenklasse, Ann. of Math. (2) 37 (1936) 107–110. <http://dx.doi.org/10.2307/1968691> (in German).
- [28] H. König, Eigenvalue Distribution of Compact Operators, Birkhäuser-Verlag, Basel, 1986.
- [29] D. Kressner, C. Tobler, Low-rank tensor Krylov subspace methods for parametrized linear systems, SIAM J. Matrix Anal. Appl. 32 (2011) 1288–1316. <http://dx.doi.org/10.1137/100799010>.
- [30] T. Kühn, W. SICKEL, T. Ullrich, Approximation numbers of Sobolev embeddings – sharp constants and tractability, J. Complexity 30 (2) (2014) 95–116.
- [31] O. Legeza, T. Rohwedder, R. Schneider, Tensor methods in quantum chemistry, in: Encyclopedia of Applied and Computational Mathematics, 2011, in preparation.
- [32] Y. Liu, G. Xu, Widths and average widths of Sobolev classes, Acta Math. Sci. Ser. B Engl. Ed. 23 (2003) 178–184.
- [33] I.V. Oseledets, Tensor-train decomposition, SIAM J. Sci. Comput. 33 (2011) 2295–2317. <http://dx.doi.org/10.1137/090752286>.
- [34] I.V. Oseledets, E.E. Tyrtyshnikov, Recursive decomposition of multidimensional tensors, Dokl. Akad. Nauk 427 (2009) 14–16. <http://dx.doi.org/10.1134/S1064562409040036> (in Russian); Dokl. Math. 80 (2009) 460–462. English translation.
- [35] I.V. Oseledets, E.E. Tyrtyshnikov, Tensor tree decomposition does not need a tree, Preprint 2009-04, INM RAS, 2009. URL: <http://spring.inm.ras.ru/osel/>.
- [36] I.V. Oseledets, E.E. Tyrtyshnikov, TT-cross approximation for multidimensional arrays, Linear Algebra Appl. 432 (2010) 70–88. <http://dx.doi.org/10.1016/j.laa.2009.07.024>.
- [37] A. Pietsch, Operator Ideals, VEB Deutscher Verlag der Wissenschaften, Berlin, 1978.
- [38] A. Pietsch, Eigenvalues and  $s$ -Numbers, Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1987.

- [39] A.S. Romanyuk, V.S. Romanyuk, Best bilinear approximations of functions from Nikol'skii–Besov classes, *Ukrain. Mat. Zh.* 64 (2012) 685–697. <http://dx.doi.org/10.1007/s11253-012-0678-1> (in Ukrainian); *Ukrainian Math. J.* 64 (2012) 781–796. English translation.
- [40] R. Schatten, *Norm Ideals of Completely Continuous Operators*, Springer, Berlin, 1970.
- [41] E. Schmidt, Zur theorie der linearen und nichtlinearen integralgleichungen, *Math. Ann.* 63 (1907) 433–476. <http://dx.doi.org/10.1007/BF01449770>.
- [42] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Physics* 326 (2011) 96–192. <http://dx.doi.org/10.1016/j.aop.2010.09.012>.
- [43] S.B. Stečkin, On absolute convergence of orthogonal series, *Dokl. Akad. Nauk SSSR* 102 (1955) 37–40 (in Russian).
- [44] V.N. Temlyakov, Approximation of functions with bounded mixed derivative, *Tr. Mat. Inst. Steklova* 178 (1986) (in Russian); *Proc. Steklov Inst. Math.* 1 (178) (1989) English translation.
- [45] V.N. Temlyakov, Estimates of best bilinear approximations of periodic functions, *Tr. Mat. Inst. Steklova* 181 (1988) 250–267, 272 (in Russian); *Proc. Steklov Inst. Math.* (4) (1989) English translation.
- [46] V.N. Temlyakov, Bilinear approximation and related questions, *Tr. Mat. Inst. Steklova* 194 (1992) 229–248 (in Russian); *Proc. Steklov Inst. Math.* 4 (194) (1993) 245–265. English translation.
- [47] V.N. Temlyakov, Estimates for the best bilinear approximations of functions and approximation numbers of integral operators, *Mat. Zametki* 51 (1992) 125–134. <http://dx.doi.org/10.1007/BF01262187>. 159 (in Russian); *Math. Notes* 51 (1992) 510–517. English translation.
- [48] V.N. Temlyakov, *Approximation of Periodic Functions*, Nova Science Publishers Inc., Commack, NY, 1993.
- [49] A. Uschmajew, Regularity of tensor product approximations to square integrable functions, *Constr. Approx.* 34 (2011) 371–391. <http://dx.doi.org/10.1007/s00365-010-9125-4>.
- [50] A. Uschmajew, *Zur Theorie der Niedrigrangapproximation in Tensorprodukten von Hilberträumen*, Ph.D. Thesis, Technische Universität Berlin, 2013 (in German).
- [51] A. Uschmajew, B. Vandereycken, The geometry of algorithms using hierarchical tensors, *Linear Algebra Appl.* 439 (2013) 133–166. <http://dx.doi.org/10.1016/j.laa.2013.03.016>.