APPROXIMATION THEORY OF TREE TENSOR NETWORKS: TENSORIZED UNIVARIATE FUNCTIONS

MAZEN ALI AND ANTHONY NOUY

Fraunhofer ITWM, 67663 Kaiserslautern, Germany

Centrale Nantes, Nantes Université, LMJL UMR CNRS 6629, France

Abstract. We study the approximation of univariate functions by combining tensorization of functions with tensor trains (TTs) – a commonly used type of tensor networks (TNs). Lebesgue L^p-spaces in one dimension can be identified with tensor product spaces of arbitrary order through tensorization. We use this tensor product structure to define different approximation tools and corresponding approximation spaces of TTs, associated with different measures of complexity. The approximation tools are shown to have (near to) optimal approximation rates for functions with classical Besov smoothness. We then use classical interpolation theory to show that a scale of interpolated smoothness spaces is continuously embedded into the scale of TT approximation spaces and, vice versa, we show that the TT approximation spaces are, in a sense, much larger than smoothness spaces when the depth of the tensor network is not restricted but are embedded into a scale of interpolated smoothness spaces if one restricts the depth.

The results of this work can be seen as both an analysis of the approximation spaces of a type of TNs and a study of the expressivity of a particular type of neural networks (NNs) – namely feed-forward sum-product networks with sparse architecture. We point out interesting parallels to recent results on the expressivity of rectifier networks. □

1. Introduction

Approximation of functions is an integral part of mathematics with many important applications in various other fields of science, engineering and economics. Many classical approximation methods – such as approximation with polynomials, splines, wavelets, rational functions, etc. – are by now thoroughly understood. In recent decades new families of methods have gained increased popularity due to their success in various applications – tensor and neural networks (TNs and NNs). See, e.g., [8, 46, 14, 15, 51, 37, 56, 27] and references therein for an overview.

In this work, we will define approximation classes of TTs – a commonly used type of TNs – and study their properties. Our goal is to investigate the expressivity of tensor networks in the framework of classical approximation theory, as was done in [32, 2] for deep rectifier NNs.

E-mail address: mazen.ali@itwm.fraunhofer.de, anthony.nouy@ec-nantes.fr.

Date: February 21, 2023.

2010 *Mathematics Subject Classification.* 41A65, 41A15, 41A10 (primary); 68T05, 42C40, 65D99 (secondary).

Key words and phrases. Tensor Networks, Tensor Trains, Matrix Product States, Neural Networks, Approximation Spaces, Besov Spaces, direct (Jackson) and inverse (Bernstein) inequalities.

1.1. **Approximation of Functions.** In this work, we focus on approximating one-dimensional real-valued functions $f : \Omega \to \mathbb{R}$ on bounded intervals $\Omega \subset \mathbb{R}$. We address the multi-dimensional setting separately in [3].

The approximation of general functions by simpler "building blocks" has been a central topic in mathematics for centuries with many arising methods: algebraic polynomials, trigonometric polynomials, splines, wavelets or rational functions are among some of the by now established tools. Recently, more sophisticated tools such as TNs or NNs have proven to be powerful techniques.

In the 20th century, a fully fledged mathematical theory of approximation has been established. It is by now well understood that approximability properties of a function by more standard tools – such as polynomials or splines – are closely related to its smoothness. Moreover, functions that can be approximated with a certain rate can be grouped to form quasi-Banach spaces. Varying the approximation rate then generates an entire scale of spaces that turn out to be so-called *interpolation spaces*. See [19, 17] for more details.

In this work, we address the classical question of function approximation but with a new set of tools relying on tensorization of functions and the use of rank-structured tensor formats (or TNs). We analyze the resulting approximation classes: we will show that many known classical spaces of smoothness are embedded in these newly defined approximation classes. On the other hand, we will also show that these classes are, in a sense, much larger than classical smoothness spaces.

1.2. **Tensor Networks.** TNs have been studied in parallel in different fields, sometimes under different names: e.g., hierarchical tensor formats in numerical analysis or sum-product networks in machine learning. TNs are commonly applied and studied in condensed matter physics, where understanding phenomena in quantum many-body systems has proven to be a challenging problem, to a large extent due to the sheer amount of dependencies that cannot be simulated even on the most powerful computers (see [50] for a non-technical introduction). In all these fields, a common challenge is the approximation of functions of a very large number of variables. This led to the development of tools tailored to so-called *high-dimensional* problems.

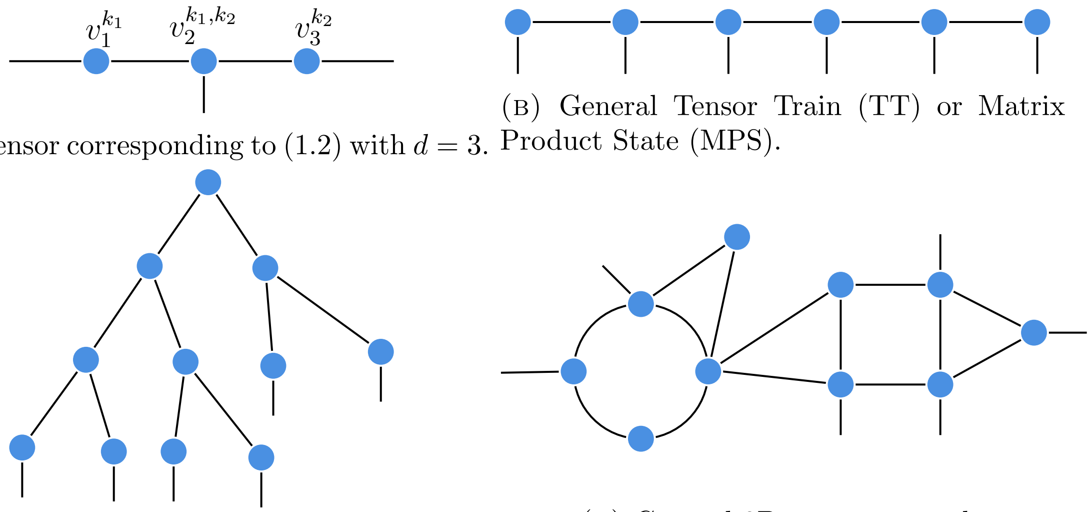
For approximating a d-variate function f, there are several types of tensor formats. The simplest is the so-called *r-term* or *CP* format, where f is approximated as

$$f(x_1, \dots, x_d) \approx \sum_{k=1}^{r} v_1^k(x_1) \cdots v_d^k(x_d). \tag{1.1}$$

If each factor v_ν^k is encoded with N parameters, the total number of parameters is thus dNr, which is linear in the number of variables. The approximation format (1.1) is successful in many applications (chemometrics, inverse problems in signal processing...) but due to a few unfavorable properties (see [34, Chapter 9]), different types of tensor formats are frequently used in numerical approximation. In particular, with the so-called *tensor train* (TT) format or *matrix product state* (MPS), the function f is approximated as

$$f(x_1, \dots, , x_d) \approx \sum_{k_1=1}^{r_1} \dots \sum_{k_{d-1}=1}^{r_{d-1}} v_1^{k_1}(x_1) v_2^{k_1,k_2}(x_2) \dots v_{d-1}^{k_{d-2},k_{d-1}}(x_{d-1}) v_d^{k_{d-1}}(x_d). \tag{1.2}$$

The numbers r_ν are referred to as *TT-ranks*, *multi-linear ranks* or *hierarchical ranks*. The rank r_ν is related to the classical notion of rank for bi-variate functions, by identifying a d-variate function as a function of two complementary groups of variables $(x_1, \dots, x_\nu)$ and $(x_{\nu+1}, \dots, x_d)$. It corresponds to the so-called β-rank r_β, with $\beta = \{1, \dots, \nu\}$. The format in (1.2) is a particular case of *tree-based tensor formats*, or *tree tensor networks (TNs)* [34, 24], the TT format being associated with a linear dimension partition tree. The TT format is the main focus of this work, although many results extend to more general TNs. Numerically, such formats have favorable stability properties and robust algorithms (see [28, 54, 47, 30]). Moreover, the corresponding TNs and decompositions have a physical interpretation in the context of entangled many-body systems, see [50, 51]. For more general TNs, we refer to Figure 1 for graphical representations. The specific choice of a TN is sometimes suggested by the problem at hand: e.g., in quantum physics by the entanglement/interaction structure of the quantum system that f is to model – if f is a vector – see, e.g., [36, 1, 58, 4].



(A) Tensor corresponding to (1.2) with $d = 3$.

(B) General Tensor Train (TT) or Matrix Product State (MPS).

(C) Hierarchical Tucker (HT) or a tree-based format.

(D) General 2D tensor network.

FIGURE 1. Examples of tensor networks. The vertices in Figure 1 represent the low-dimensional functions (tensors) in the decomposition, such as $v^1, \ldots, v^d$ in (1.2). The edges between the vertices represent summation over an index (*contraction*) between two functions (tensors), such as summation over k_ν in (1.2). The free edges represent input variables $x_1, \ldots, x_d$ in (1.2).

At first glance, it seems that TNs are a tool suited only for approximating high-dimensional functions. However, such formats can be applied in any multi-variate setting and this multi-variate setting can be applied even if $d = 1$ by identifying a one-dimensional function with a multi-variate function (or tensor). This identification is the *tensorization* of functions which is at the core of the approximation tools considered in this work. It was originally applied for matrices in [52] and later coined as *quantized tensor format* when tensorization is combined with the use of a tensor format.

In high-dimensional approximation, keeping the ranks r_β small relies on the correct choice of the tensor network that "fits" the interaction structure as hinted above. For the approximation of tensorized functions a different type of structure is required. In [28], it was shown that, if f is a vector of evaluations of a polynomial on a grid, then choosing the TT format yields hierarchical ranks that are bounded by the degree of the polynomial plus one. Similar statements were shown for trigonometric polynomials and exponential functions. In [42], tensorized TT formats were applied in numerical modeling and, in particular, it was shown that discrete solutions to high-dimensional elliptic and parabolic problems can be approximated with a complexity that is logarithmic in the number of grid points. In [41], it was shown that a finite element approximation of two-dimensional functions with singularities, where the coefficient vector was stored in a tensorized TT format, automatically recovers an exponential rate of convergence, analogous to that of an hp-approximation.

The pioneering observations of [28, 52] as well as the subsequent developments in [53, 41, 42] provide the crucial tools for our work. We utilize these techniques to propose an approximation theory – in the spirit of classical approximation theory (see [19, 17]) – for tensorized functions, treating the univariate case in depth here and addressing the multivariate case in [3]. We first show that Lebesgue spaces of p-integrable functions are isometric to tensor product spaces of any order and analyze some basic properties of this identification. We then define and analyze the approximation classes of L^p functions that can be approximated by rank-structured functions in the TT format with a certain rate, showing that these classes are quasi-Banach spaces under appropriate conditions. Finally, we show that classical smoothness spaces are continuously embedded in TT approximation classes, while vice versa TT approximation classes are embedded in smoothness spaces only under additional assumptions about the TT length (or depth of the corresponding linear tree).

1.3. **Tensor vs. Neural Networks.** Recently multiple connections between TTs and NNs have been discovered, see, e.g., [10, 55, 44, 16, 13, 43]. Tree tensor networks can be seen as convolutional feedforward neural networks with nonlinear feature maps, product pooling, a number of layers equal to the depth of the dimension partition tree and a number of neurons equal to the sum of tree tensor ranks (see [16]). Our work was partly motivated by current developments in the field of deep learning, particularly [32], where the authors analyzed the approximation spaces of deep rectifier networks. In this spirit, our work can be seen as a result on the approximation power of a particular type of NNs, where the TT format is a feed-forward sum-product NN with a recurrent neural network architecture. When compared to the results of [32, 2] on approximation classes of ReLU networks, we observe that both TNs and NNs are able to recover optimal or close to optimal approximation rates for functions with any order of Sobolev or Besov smoothness. This is to be contrasted with more standard approximation tools, such as splines or wavelets, where the approximation class (and thus the approximation method) has to be adapted to the smoothness notion in question, i.e., for a given a spline or wavelet order, the approximation tool can optimally approximate only functions with the same smoothness order. Moreover, both tools will frequently perform better than predicted on specific instances of functions that possess structural features that are not captured by classical smoothness theory[1].

Of course, this is simply to say that both tools do a good job when it comes to classical notions of smoothness. We still expect that NNs approximation classes are very different than those of TNs, in an appropriate sense. We also show that TT approximation classes are not embedded in any Besov space, as was shown in [32] for RePU networks, unless the depth of tensor networks is restricted.

1.4. **Main Results.** First, we show that any L^p-function f defined on the interval $[0, 1)$ can be identified with a tensor. For a given $b \in \mathbb{N}$ (the base) and $d \in \mathbb{N}$ (the level or resolution), we first note that any $x \in [0, 1)$ can be uniquely decomposed as

$$x = \sum_{k=1}^{d} i_k b^{-k} + b^{-d} y := t_{b,d}(i_1, \ldots, i_d, y),$$

[1]A fundamental theory of these structures remains an open question for both tools.

where $(i_1, ..., i_d)$ is the representation of $\lfloor b^d x \rfloor$ in base b and $y = b^d x - \lfloor b^d x \rfloor$. This allows to identify a function with a tensor (or multivariate function)

$$\boldsymbol{f}(i_1, \dots, i_d, y) = f(t_{b,d}(i_1, \dots, i_d, y)) := T_{b,d} f(i_1, \dots, i_d, y),$$

and to define different notions of ranks for a univariate function. A function f can be tensorized at different levels $d \in \mathbb{N}$. We analyze the relation between tensorization maps at different levels, and the relation between the ranks of the corresponding tensors of different orders. When looking at $T_{b,d}$ as a map on $L^p([0,1))$, an important observation is given by Theorem 2.15 and Lemma B.1.

Result 1.1. *For any $0 < p \leq \infty$, $b \in \mathbb{N}$ ($b \geq 2$) and $d \in \mathbb{N}$, the map $T_{b,d}$ is a linear isometry from $L^p([0,1))$ to the algebraic tensor space $\mathbf{V}_{b,d,L^p} := (\mathbb{R}^b)^{\otimes d} \otimes L^p([0,1))$, where $\mathbf{V}_{b,d,L^p}$ is equipped with a crossnorm, which is a reasonable crossnorm for $p \geq 1$.*

1.4.1. *Approximation Tools.* For later use in approximation, we introduce the tensor subspace

$$\mathbf{V}_{b,d,S} := (\mathbb{R}^b)^{\otimes d} \otimes S,$$

where $S \subset L^p([0,1))$ is some finite-dimensional subspace. Then, we can identify $\mathbf{V}_{b,d,S}$ with a finite-dimensional subspace of L^p as

$$V_{b,d,S} := T_{b,d}^{-1}(\mathbf{V}_{b,d,S}) \subset L^p([0,1)).$$

We introduce the crucial assumption that S is closed under b-adic dilation, i.e., for any $f \in S$ and any $k \in \{0, \dots, b-1\}$, $f(b^{-1}(\cdot + k)) \in S$. Under this assumption, which is reminiscent of multi-resolution analysis (MRA), we obtain bounds for TT-ranks that are related to the dimension of S. Also, under this assumption on S, we obtain the main results given by Propositions 2.19 and 2.20 and Theorem 2.21.

Result 1.2. *The spaces $V_{b,d,S}$ form a hierarchy of L^p-subspaces, i.e.*

$$S := V_{b,0,S} \subset V_{b,1,S} \subset V_{b,2,S} \subset \dots,$$

and $V_{b,S} := \bigcup_{d \in \mathbb{N}} V_{b,d,S}$ is a linear space. If we further assume that S contains the constant function one, $V_{b,S}$ is dense in L^p for $0 < p < \infty$.

For the approximation of multivariate functions (or tensors), we use the set $\Phi_{b,d,S,\boldsymbol{r}}$ of functions φ whose tensorization $\boldsymbol{\varphi} = T_{b,d}(\varphi)$ admits a representation in TT format with TT-ranks $\boldsymbol{r} = (r_\nu)_{\nu=1}^d$, i.e. given a basis $\{\varphi_k\}_{k=1}^{\dim S}$ of S,

$$\boldsymbol{\varphi}(i_1, \dots, i_d, y) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} \sum_{k_{d+1}=1}^{\dim S} v_1^{k_1}(i_1) v_2^{k_1,k_2}(i_2) \cdots v_d^{k_{d-1},k_d}(i_d) v_{d+1}^{k_d,k_{d+1}} \varphi_{k_{d+1}}(y), \tag{1.3}$$

where the parameters $\mathbf{v} := (v_1, \dots, v_{d+1})$ form a tensor network (a collection of low-order tensors). Then an approximation tool for univariate functions is defined as

$$\Phi := (\Phi_n)_{n \in \mathbb{N}}, \quad \Phi_n = \{\varphi \in \Phi_{b,d,S,\boldsymbol{r}} : d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d, \mathrm{compl}(\varphi) \leq n\},$$

where $\mathrm{compl}(\varphi)$ is some measure of complexity of a function φ. We introduce three different measures of complexity, $\mathrm{compl}_{\mathcal{N}}$, $\mathrm{compl}_{\mathcal{C}}$ and $\mathrm{compl}_{\mathcal{S}}$, that are respectively the sum of ranks, the number of parameters (or number of entries of the tensors from the tensor network) and the number of non-zero parameters (or number of non-zero entries of the tensors from the tensor network). Consequently, this defines three types of approximation tools $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$. When interpreting a tensor network $\mathbf{v}$ as a sum-product neural network, $\mathrm{compl}_{\mathcal{N}}$ corresponds to the number of neurons, $\mathrm{compl}_{\mathcal{C}}$ to the number of weights of a fully connected network, and $\mathrm{compl}_{\mathcal{S}}$ the number of non-zero weights (or number of connections of a sparsely connected network).

1.4.2. *Approximation Rates.* For each approximation tool $\Phi_n \in \left\{\Phi_n^{\mathcal{N}}, \Phi_n^{\mathcal{C}}, \Phi_n^{\mathcal{S}}\right\}$ we consider the corresponding best approximation error

$$E_n(f)_p := \inf_{\varphi\in\Phi_n} \|f-\varphi\|_p$$

for functions f in $L^p([0,1))$. The main results for approximation rates are Theorems 5.2, 5.8 and 5.11 and can be summarized as follows.

Result 1.3. *Let $1 \le p \le \infty$ and $k \in \mathbb{N}$. For any $f \in W^{k,p}$ we have*

$$E_n^{\mathcal{N}}(f)_p \le Cn^{-2k}\|f\|_{W^{k,p}},$$
$$E_n^{\mathcal{S}}(f)_p \le E_n^{\mathcal{C}}(f)_p \le Cn^{-k}\|f\|_{W^{k,p}},$$

with constants C depending on k, m, b.

Result 1.4. *Let $1 \le p < \infty$, $0 < \tau < p$, $\alpha > 1/\tau - 1/p$, and assume $f \in B^\alpha_{\tau,\tau}$. Then, for any $\sigma > 0$,*

$$E_n^{\mathcal{N}}(f)_p \le C\,|f|_{B^\alpha_{\tau,\tau}}\, n^{-\frac{\alpha}{1+\sigma}},$$
$$E_n^{\mathcal{C}}(f)_p \le C\,|f|_{B^\alpha_{\tau,\tau}}\, n^{-\frac{\alpha}{2+\sigma}},$$
$$E_n^{\mathcal{S}}(f)_p \le C\,|f|_{B^\alpha_{\tau,\tau}}\, n^{-\frac{\alpha}{1+\sigma}},$$

where the constants C depend on $\alpha > 0$, $\sigma > 0$, b and m. In particular, they diverge to infinity as $\sigma \to 0$ or $\alpha \to 1/\tau - 1/p$.

Result 1.5 (Spectral Approximation)**.** *For $S = \mathbb{P}_m$ with a fixed $m \in \mathbb{N}_0$, we show that if f is analytic on an open domain containing* $[0,1]$,

$$E_n^{\mathcal{N}}(f)_\infty \le C\rho^{-n^{1/2}},$$
$$E_n^{\mathcal{S}}(f)_\infty \le E_n^{\mathcal{C}}(f)_\infty \le C\rho^{-n^{1/3}},$$

for constants $C, \rho > 1$. This can be extended to analytic functions with singularities using ideas from [41].

1.4.3. *Approximation spaces.* An approximation tool (Φ_n) is associated with an approximation class $A^\alpha_q(L^p, (\Phi_n))$ of functions f such that the best approximation error $E_n(f)_p$ decays with an algebraic rate α for $q = \infty$ or slightly faster for $q < \infty$ (the sequence $(n^{\alpha-1/q}E_n(f)_p)_{n\ge1}$ is in ℓ_q), see Section 3.4 for a precise definition. The approximation tools $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$ are associated with approximation classes $N^\alpha_q(L^p)$, $C^\alpha_q(L^p)$ and $S^\alpha_q(L^p)$, respectively. We first obtain the following properties and relationship given by Theorems 3.17 and 3.19.

Result 1.6. *For any $\alpha > 0$, $0 < p \le \infty$ and $0 < q \le \infty$, the classes $N^\alpha_q(L^p)$, $C^\alpha_q(L^p)$ and $S^\alpha_q(L^p)$ are quasi-normed vector spaces and satisfy the continuous embeddings*

$$C^\alpha_q(L^p) \hookrightarrow S^\alpha_q(L^p) \hookrightarrow N^\alpha_q(L^p) \hookrightarrow C^{\alpha/2}_q(L^p).$$

Direct Embeddings of Smoothness Spaces. The main results on embeddings of smoothness spaces are Theorems 6.4 and 6.6 and can be summarized as follows.

Result 1.7 (Direct Embedding for Sobolev Spaces $W^{k,p}$ and Besov spaces $B^\alpha_{p,q}$)**.** *Let $W^{k,p}$ denote the Sobolev space of $k \in \mathbb{N}$ times weakly differentiable, p-integrable functions and $B^\alpha_{p,q}$ the Besov space of smoothness $\alpha > 0$, with primary parameter p and secondary parameter q. Then, for $S = \mathbb{P}_m$ (space of polynomials of degree m) with a fixed $m \in \mathbb{N}_0$, we show that for $1 \le p \le \infty$ and any $k \in \mathbb{N}$*

$$W^{k,p} \hookrightarrow N^{2k}_\infty(L^p), \quad W^{k,p} \hookrightarrow C^k_\infty(L^p) \hookrightarrow S^k_\infty(L^p),$$

and for $0 < q \leq \infty$ *and any* $\alpha > 0$

$$B^{\alpha}_{p,q} \hookrightarrow N^{2\alpha}_{q}(L^p), \quad B^{\alpha}_{p,q} \hookrightarrow C^{\alpha}_{q}(L^p) \hookrightarrow S^{\alpha}_{q}(L^p).$$

Remark 1.8. *Note that for* $p = q$ *and non-integer* $\alpha > 0$, $B^{\alpha}_{p,p} = W^{\alpha,p}$ *is the fractional Sobolev space. Moreover, these results can also be extended to the range* $0 < p < 1$, *see Remark 5.3.*

Result 1.9 (Direct Embedding for Besov Spaces $B^{\alpha}_{\tau,q}$)**.** *Let* $B^{\alpha}_{p,q}$ *denote the Besov space of smoothness* $\alpha > 0$, *with primary parameter* p *and secondary parameter* q. *Then, for* $S = \mathbb{P}_m$ *with a fixed* $m \in \mathbb{N}_0$, *we show that for any* $1 \leq p < \infty$, *any* $0 < \tau < p$, *any* $\gamma > 1/\tau - 1/p$ *and any* $0 < \bar{\gamma} < \gamma$,

$$B^{\gamma}_{\tau,\tau} \hookrightarrow N^{\bar{\gamma}}_{\infty}(L^p) \hookrightarrow C^{\bar{\gamma}/2}_{\infty}(L^p), \quad B^{\gamma}_{\tau,\tau} \hookrightarrow S^{\bar{\gamma}}_{\infty}(L^p),$$

and for any $0 < q \leq \infty$, *any* $0 < \alpha < \bar{\gamma}$

$$(L^p, B^{\gamma}_{\tau,\tau})_{\alpha/\bar{\gamma},q} \hookrightarrow N^{\alpha}_{q}(L^p) \hookrightarrow C^{\alpha/2}_{q}(L^p), \quad (L^p, B^{\gamma}_{\tau,\tau})_{\alpha/\bar{\gamma},q} \hookrightarrow S^{\alpha}_{q}(L^p),$$

where $(X, Y)_{\theta,q}$, $0 < \theta < 1$, *is the real* K*-interpolation space between* X *and* $Y \hookrightarrow X$.

Remark 1.10. *Note that both Result 1.7 and Result 1.9 apply to Besov spaces. The Besov spaces in Result 1.7 are of the type* $B^{\alpha}_{p,q}$, *where* p *is the same for the error measure. Such Besov spaces are captured by linear approximation and for* $p \geq 1$ *these are equal to or are very close to Sobolev spaces.*

On the other hand, the Besov spaces $B^{\alpha}_{\tau,\tau}$ *for* $1/\tau = \alpha + 1/p$ *are much larger and these correspond to the critical embedding line. These Besov spaces can only be captured by nonlinear approximation. Our results require* $\alpha > 1/\tau - 1/p$, *i.e., Besov spaces that are strictly above the critical line.*

Inverse Embeddings. The main results on inverse embeddings are Theorems 7.1 and 7.2 and can be summarized as follows.

Result 1.11 (No Inverse Embedding)**.** *Let* $B^{\alpha}_{p,q}$ *denote the Besov space of smoothness* α, *with primary parameter* p *and secondary parameter* q. *We show that for any* $\alpha > 0$, $0 < p, q \leq \infty$, *and any* $\tilde{\alpha} > 0$,

$$C^{\alpha}_{q}(L^p) \not\hookrightarrow B^{\tilde{\alpha}}_{p,q}.$$

Result 1.12 (Inverse Embedding For Restricted Depth)**.** *Define for* $n \in \mathbb{N}$, $k_{\mathrm{B}} \geq 1$ *and* $c_{\mathrm{B}} > 0$ *the restricted sets*

$$\Phi^{\mathrm{B}}_{n} := \{\varphi \in V_{b,m} : \mathrm{compl}_{\mathcal{N}}(\varphi) \leq n \quad \text{and} \quad d(\varphi) \leq k_{\mathrm{B}} \log_b(n) + c_{\mathrm{B}}\},$$

where $d(\varphi)$ *is the minimal possible level (depth) for a tensorized representation of* φ. *For* $1 \leq p < \infty$, *the depth restricted approximation classes* $A^{\alpha}_{q}(L^p, (\Phi^{\mathrm{B}}_{n}))$ *satisfy the continuous embeddings*

$$A^{\alpha}_{q}(L^p, (\Phi^{\mathrm{B}}_{n})) \hookrightarrow (L^p, B^{m+1}_{\tau,\tau})_{\frac{\alpha}{k_{\mathrm{B}}(m+1)},q},$$
$$A^{k_{\mathrm{B}}(m+1)}_{\infty}(L^p, (\Phi^{\mathrm{B}}_{n})) \hookrightarrow B^{m+1}_{\tau,\tau}.$$

In words:

- For the approximation tools $\Phi^{\mathcal{C}}_{n}$ and $\Phi^{\mathcal{S}}_{n}$ (of fixed polynomial degree $m \in \mathbb{N}_0$), we obtain optimal approximation rates for Sobolev spaces $W^{\alpha,p}$ of *any* order $\alpha > 0$.
- For the approximation tool $\Phi^{\mathcal{N}}_{n}$, we obtain twice the optimal rate. Note, however, that the corresponding complexity measure only reflects the number of neurons in a corresponding neural network. It does not reflect the representation or computational complexity. Moreover, from [18] we know that an optimal approximation

tool with *continuous* parametrization for the Sobolev space $W^{\alpha,p}$ cannot exceed the rate α, see also [61].

- For the approximation tools $\Phi_n^{\mathcal{N}}$ and $\Phi_n^{\mathcal{S}}$, we obtain *near to* optimal rates[2] for the Besov space $B_{\tau,\tau}^{\alpha}$, for any order $\alpha > 0$. For $\Phi_n^{\mathcal{C}}$, the approximation rate is near to half the optimal rate.
- More explicitly, for a given approximation accuracy $\varepsilon > 0$ in the L^p-norm, a target function f in the Sobolev space $W^{k,p}$ can be approximated with a tensorized function φ with TT-ranks of the order $\varepsilon^{-1/(2k)}$ and overall number of TT-parameters of the order $\varepsilon^{-1/k}$ – hence, optimal in the sense of nonlinear widths.

  If instead $f \in B_{\tau,\tau}^{\alpha}$, the TT-ranks of φ are of the order $\varepsilon^{-1/\alpha}$ and the overall number of TT-parameters is of the order $|\log(\varepsilon)|\varepsilon^{-2/\alpha}$. However, the number of *nonzero* TT-parameters of φ is $|\log(\varepsilon)|\varepsilon^{-1/\alpha}$.
- Particularly the tool $\Phi_n^{\mathcal{S}}$ is interesting, as it corresponds to deep, sparsely connected networks. The above results imply that deep, sparsely connected tensor networks can optimally replicate both h-uniform and h-adaptive approximation of any order.
- All approximation tools achieve exponential approximation rates for analytic target functions. Together with the previous result, this implies that deep, sparsely connected tensor networks can optimally replicate hp-adaptive approximation, while the underlying polynomial degree of the tensor network remains fixed.
- Finally, an arbitrary function from any of the three approximation classes possesses no Besov smoothness. This can be mainly attributed to the depth of the tensor network and smoothness can be recovered if we restrict the latter to grow not too fast with the complexity.

We restrict ourselves in this work to approximation of functions on intervals in one dimension to focus on the presentation of the basic concepts and postpone the multi-dimensional case to [3].

We base our approximation tool on the TT format. Although some of our results would remain unchanged for other tree-based tensor formats, ranks are generally affected by the choice of the format. This is known for multi-dimensional non-tensorized approximation with tensor formats, see, e.g., [12, 11]. In the non-tensorized case, ranks remain low if the format “fits” the problem at hand, e.g., if the format mimics the interaction structure dictated by the differential operator, see [1]. In the context of tensorized 1D approximation, the tensor format would have to fit the self-similarity, periodicity or other algebraic features of the target function, see, e.g., [29, 5].

We thus stress the following point concerning the approximation power of tree-based tensor networks: on one hand, when comparing approximation classes of different tensor networks to spaces of classical smoothness – the distinction between different tree-based formats seems insignificant. On the other hand, when comparing approximation classes of different tensor networks *to each other* – we expect these to be substantially different.

1.5. **Outline.** In Section 2, we discuss how one-dimensional functions can be identified with tensors and analyze some basic properties of this identification. In Section 3, we introduce the approximation tool, briefly review general results from approximation theory, and analyze several approximation classes of rank-structured functions. In particular, we show that these classes are quasi-Banach spaces. In Section 4, we discuss how classical approximation tools can be encoded with a TT format and estimate the resulting complexity. Among the classical tools considered are fixed knot splines, free knot splines,

[2] The approximation rates are arbitrary close to optimal.

polynomials (of higher order)[3]. In Section 5, we show approximation rates for our approximation tool that lead to direct embeddings in Section 6. We show in Section 7 that inverse embeddings can only hold if we restrict the depth (or resolution) of the TT format. In Sections 8 and 9, we discuss the role of tensorization, seen as a particular featuring step, and the role of depth and sparsity.

2. Tensorization of Measurable Functions

We begin by introducing how one-dimensional measurable functions can be identified with tensors of arbitrary dimension. We then introduce finite-dimensional subspaces of tensorized functions and show that these form a hierarchy of subspaces that are dense in L^p. This will be the basis for our approximation tool in Section 3.

2.1. The Tensorization Map.

Consider one-dimensional functions on the unit interval, $f : [0,1) \to \mathbb{R}$, that we tensorize as in Section 1.4. We deduce the following property.

Lemma 2.1. *The conversion map $t_{b,d}$ defines a linear bijection from the set $I_b^d \times [0,1)$ to the interval $[0,1)$, with inverse defined for $x \in [0,1)$ by*

$$t_{b,d}^{-1}(x) = (\lfloor bx \rfloor, \lfloor b^2 x \rfloor \text{ mod } b, \ldots, \lfloor b^d x \rfloor \text{ mod } b, b^d x - \lfloor b^d x \rfloor).$$

Definition 2.2 (Tensorization Map)**.** *We define the* tensorization map

$$T_{b,d} : \mathbb{R}^{[0,1)} \to \mathbb{R}^{I_b^d \times [0,1)}, \quad f \mapsto f \circ t_{b,d} := \boldsymbol{f}$$

which associates to a function $f \in \mathbb{R}^{[0,1)}$ the multivariate function $\boldsymbol{f} \in \mathbb{R}^{I_b^d \times [0,1)}$ such that

$$\boldsymbol{f}(i_1, \ldots, i_d, y) := f(t_{b,d}(i_1, \ldots, i_d, y)).$$

From Lemma 2.1, we directly deduce the following property of $T_{b,d}$, which allows to identify the spaces $\mathbb{R}^{[0,1)}$ and $\mathbb{R}^{I_b^d \times [0,1)}$.

Proposition 2.3. *The tensorization map $T_{b,d}$ is a linear bijection from $\mathbb{R}^{[0,1)}$ to $\mathbb{R}^{I_b^d \times [0,1)}$, with inverse given for $\boldsymbol{f} \in \mathbb{R}^{I_b^d \times [0,1)}$ by $T_{b,d}^{-1} \boldsymbol{f} = \boldsymbol{f} \circ t_{b,d}^{-1}$.*

The space $\mathbb{R}^{I_b^d \times [0,1)}$ can be identified with the algebraic tensor space

$$\mathbf{V}_{b,d} := \mathbb{R}^{I_b^d} \otimes \mathbb{R}^{[0,1)} = \underbrace{\mathbb{R}^{I_b} \otimes \ldots \otimes \mathbb{R}^{I_b}}_{d \text{ times}} \otimes \mathbb{R}^{[0,1)} =: (\mathbb{R}^{I_b})^{\otimes d} \otimes \mathbb{R}^{[0,1)},$$

which is the set of functions $\boldsymbol{f}$ defined on $I_b^d \times [0,1)$ that admit a representation

$$\boldsymbol{f}(i_1, \ldots, i_d, y) = \sum_{k=1}^{r} v_1^k(i_1) \ldots v_d^k(i_d) g^k(y) := \sum_{k=1}^{r} (v_1^k \otimes \ldots \otimes v_d^k \otimes g^k)(i_1, \ldots, i_d, y) \tag{2.1}$$

for some $r \in \mathbb{N}$ and for some functions $v_\nu^k \in \mathbb{R}^{I_b}$ and $g^k \in \mathbb{R}^{[0,1)}$, $1 \le k \le r$, $1 \le \nu \le d$. Letting $\{\delta_{j_\nu} : j_\nu \in I_b\}$ be the canonical basis of $\mathbb{R}^{I_b}$, defined by $\delta_{j_\nu}(i_\nu) = \delta_{i_\nu, j_\nu}$, a function $\boldsymbol{f} \in \mathbb{R}^{I_b^d \times [0,1)}$ admits the particular representation

$$\boldsymbol{f} = \sum_{j_1 \in I_b} \ldots \sum_{j_d \in I_b} \delta_{j_1} \otimes \ldots \otimes \delta_{j_d} \otimes \boldsymbol{f}(j_1, \ldots, j_d, \cdot). \tag{2.2}$$

The following result provides an interpretation of the above representation.

[3] See [3] for the encoding of wavelets.

Lemma 2.4. *Let $f \in \mathbb{R}^{[0,1)}$ and $\boldsymbol{f} = T_{b,d}f \in \mathbf{V}_{b,d}$. For $(j_1,\dots,j_d) \in I_b^d$ and $j = \sum_{k=1}^d j_k b^{d-k}$, we have*

$$T_{b,d}(f\mathbb{1}_{[b^{-d}j,b^{-d}(j+1))}) = \delta_{j_1} \otimes \ldots \otimes \delta_{j_d} \otimes \boldsymbol{f}(j_1,\dots,j_d,\cdot), \tag{2.3}$$

and

$$\boldsymbol{f}(j_1,\dots,j_d,\cdot) = f(b^{-d}(j+\cdot)), \tag{2.4}$$

where $f(b^{-d}(j+\cdot))$ is the restriction of f to the interval $[b^{-d}j, b^{-d}(j+1))$ rescaled to $[0,1)$.

Proof. For $x = t_{b,d}(i_1,\dots,i_d,y)$,

$$\begin{aligned} f(x)\mathbb{1}_{[b^{-d}j,b^{-d}(j+1))}(x) &= \delta_{j_1}(i_1)\ldots\delta_{j_d}(i_d)f(t_{b,d}(i_1,\dots,i_d,y)) \\ &= \delta_{j_1}(i_1)\ldots\delta_{j_d}(i_d)f(t_{b,d}(j_1,\dots,j_d,y)) \\ &= \delta_{j_1}(i_1)\ldots\delta_{j_d}(i_d)\boldsymbol{f}(j_1,\dots,j_d,y). \end{aligned}$$

The property (2.4) simply results from the definition of $\boldsymbol{f}$. □

From Lemma 2.4, we deduce that the representation (2.2) corresponds to the decomposition of $f = T_{b,d}^{-1}(\boldsymbol{f})$ as a superposition of functions with disjoint supports,

$$f(x) = \sum_{j=0}^{b^d-1} f_j(x), \quad f_j(x) = \mathbb{1}_{[b^{-d}j,b^{-d}(j+1))}(x)f(x), \tag{2.5}$$

where f_j is the function supported on the interval $[b^{-d}j, b^{-d}(j+1))$ and equal to f on this interval. Also, Lemma 2.4 yields the following result.

Corollary 2.5. *A function $f \in \mathbb{R}^{[0,1)}$ defined by*

$$f(x) = \begin{cases} g(b^d x - j) & \text{for} \quad x \in [b^{-d}j, b^{-d}(j+1)) \\ 0 & \text{elsewhere} \end{cases}$$

with $g \in \mathbb{R}^{[0,1)}$ and $0 \le j < b^d$ admits a tensorization $T_{b,d}f = \delta_{j_1} \otimes \ldots \otimes \delta_{j_d} \otimes g$, which is an elementary tensor.

We now provide a useful result on compositions of tensorization maps for changing the representation level of a function.

Lemma 2.6. *Let $\bar{d}, d \in \mathbb{N}$ such that $\bar{d} > d$. For any $(i_1,\dots,i_{\bar{d}},y) \in I_b^{\bar{d}} \times [0,1)$, we have*

$$t_{b,\bar{d}}(i_1,\dots,i_{\bar{d}},y) = t_{b,d}(i_1,\dots,i_d,t_{b,\bar{d}-d}(i_{d+1},\dots,i_{\bar{d}},y)),$$

and the operator $T_{b,\bar{d}} \circ T_{b,d}^{-1}$ from $\mathbf{V}_{b,d}$ to $\mathbf{V}_{b,\bar{d}}$ is such that

$$T_{b,\bar{d}} \circ T_{b,d}^{-1} = id_{\{1,\dots,d\}} \otimes T_{b,\bar{d}-d}$$

where $id_{\{1,\dots,d\}}$ is the identity operator on $\mathbb{R}^{I_b^d}$ and $T_{b,\bar{d}-d}$ is the tensorization map from $\mathbb{R}^{[0,1)}$ to $\mathbf{V}_{b,\bar{d}-d}$. Also, the operator $T_{b,d} \circ T_{b,\bar{d}}^{-1}$ from $\mathbf{V}_{b,\bar{d}}$ to $\mathbf{V}_{b,d}$ is such that

$$T_{b,d} \circ T_{b,\bar{d}}^{-1} = id_{\{1,\dots,d\}} \otimes T_{b,\bar{d}-d}^{-1}.$$

Proof. See Appendix B. □

For $d = 0$, we adopt the conventions that $t_{b,0}$ is the identity on $[0,1)$, $T_{b,0}$ is the identity operator on $\mathbb{R}^{[0,1)}$, and $\mathbf{V}_{b,0} = \mathbb{R}^{[0,1)}$.

2.2. **Measurable functions.** We now look at $T_{b,d}$ as a linear map between spaces of measurable functions, by equipping the interval $[0,1)$ with the Lebesgue measure.

Proposition 2.7. *The Lebesgue measure λ on $[0,1)$ is the push-forward measure through the map $t_{b,d}$ of the product measure $\mu_{b,d} := \mu_b^{\otimes d} \otimes \lambda$, where μ_b is the uniform probability measure on I_b. Then the tensorization map $T_{b,d}$ defines a linear isomorphism from the space $\mathcal{M}([0,1))$ of measurable functions on $[0,1)$ to the space $\mathcal{M}(I_b^d \times [0,1))$ of measurable functions on $I_b^d \times [0,1)$, where $[0,1)$ is equipped with the Lebesgue measure and $I_b^d \times [0,1)$ is equipped with the product measure $\mu_{b,d}$.*
Furthermore, the algebra $\mathcal{M}(I_b^d \times [0,1))$ is identified with the tensor product of algebras $\mathcal{M}(I_b)^{\otimes d} \otimes \mathcal{M}([0,1))$, which is itself identified with $(\mathbb{R}^{I_b})^{\otimes d} \otimes \mathcal{M}([0,1))$.

Proof. See Appendix B. □

In the sequel, when considering measurable functions, the notation $\mathbf{V}_{b,d}$ will stand for the tensor space $(\mathbb{R}^{I_b})^{\otimes d} \otimes \mathcal{M}([0,1))$, with the convention $\mathbf{V}_{b,0} = \mathcal{M}([0,1))$.

2.3. **Ranks and Minimal Subspaces.** The minimal integer r such that $\boldsymbol{f} \in \mathbf{V}_{b,d}$ admits a representation of the form (2.1) is the *canonical tensor rank* of $\boldsymbol{f}$ denoted $r(\boldsymbol{f})$. We deduce from the representation (2.2) that

$$r(\boldsymbol{f}) \leq b^d.$$

Other notions of ranks can be defined from the classical notion of rank by identifying a tensor with a tensor of order two (through unfolding). Letting $V_\nu := \mathbb{R}^{I_b}$ for $1 \leq \nu \leq d$, and $V_{d+1} := \mathbb{R}^{[0,1)}$ (or $\mathcal{M}([0,1))$ in the case of measurable functions), we have

$$\mathbf{V}_{b,d} = \bigotimes_{\nu=1}^{d+1} V_\nu.$$

Then for any $\beta \subset \{1,\ldots,d+1\}$ and its complementary set $\beta^c = \{1,\ldots,d+1\}\setminus\beta$, a tensor $\boldsymbol{f} \in \mathbf{V}_{b,d}$ can be identified with an order-two tensor in $\mathbf{V}_\beta \otimes \mathbf{V}_{\beta^c}$, where $\mathbf{V}_\gamma = \bigotimes_{\nu\in\gamma} V_\nu$, called the β-unfolding of $\boldsymbol{f}$. This allows us to define the notion of β-rank.

Definition 2.8 (β-rank)**.** *For $\beta \subset \{1,\ldots,d+1\}$, the β-rank of $\boldsymbol{f} \in \mathbf{V}_{b,d}$, denoted $r_\beta(\boldsymbol{f})$, is the minimal integer such that $\boldsymbol{f}$ admits a representation of the form*

$$\boldsymbol{f} = \sum_{k=1}^{r_\beta(\boldsymbol{f})} \boldsymbol{v}_\beta^k \otimes \boldsymbol{v}_{\beta^c}^k, \tag{2.6}$$

where $\boldsymbol{v}_\beta^k \in \mathbf{V}_\beta$ and $\boldsymbol{v}_{\beta^c}^k \in \mathbf{V}_{\beta^c}$.

Since $\mathbf{V}_{b,d}$ is an algebraic tensor space, the β-rank is finite and we have $r_\beta(\boldsymbol{f}) \leq r(\boldsymbol{f})$ (though the β-rank can be much smaller). Moreover, we have the following straightforward property

$$r_\beta(\boldsymbol{f}) = r_{\beta^c}(\boldsymbol{f}),$$

and the bound

$$r_\beta(\boldsymbol{f}) \leq \min\left\{\prod_{\nu\in\beta} \dim V_\nu, \prod_{\nu\in\beta^c} \dim V_\nu\right\}, \tag{2.7}$$

which can be useful for small b and either very small or very large $\#\beta$.

Representation (2.6) is not unique but the space spanned by the $\boldsymbol{v}_k^\beta$ is unique and corresponds to the β-*minimal subspace* of $\boldsymbol{f}$.

Definition 2.9 (β-minimal subspace)**.** *For $\beta \subset \{1,\dots,d+1\}$, the β-minimal subspace of $\boldsymbol{f}$, denoted $U_\beta^{\min}(\boldsymbol{f})$, is the smallest subspace $\mathbf{U}_\beta \subset \mathbf{V}_\beta$ such that $\boldsymbol{f} \in \mathbf{U}_\beta \otimes \mathbf{V}_{\beta^c}$, and its dimension is*

$$\dim(U_\beta^{\min}(\boldsymbol{f})) = r_\beta(\boldsymbol{f}).$$

We have the following useful characterization of minimal subspaces from partial evaluations of a tensor.

Lemma 2.10. *For $\beta \subset \{1,\dots,d\}$ and $\boldsymbol{f} \in \mathbf{V}_{b,d}$,*

$$U_{\beta^c}^{\min}(\boldsymbol{f}) = \operatorname{span}\{\boldsymbol{f}(j_\beta,\cdot) : j_\beta \in I_b^{\#\beta}\} \subset \mathbf{V}_{b,d-\#\beta},$$

where $\boldsymbol{f}(j_\beta,\cdot) \in \mathbf{V}_{\beta^c} = \mathbf{V}_{b,d-\#\beta}$ is a partial evaluation of $\boldsymbol{f}$ along dimensions $\nu \in \beta$.

Proof. See Appendix B. □

Next we define a notion of (β,d)-rank for univariate functions.

Definition 2.11 ((β,d)-rank)**.** *For a (measurable) function $f : [0,1) \to \mathbb{R}$, $d \in \mathbb{N}$ and $\beta \subset \{1,\dots,d+1\}$, we define the (β,d)-rank of f, denoted $r_{\beta,d}(f)$, as the β-rank of its tensorization in $\mathbf{V}_{b,d}$,*

$$r_{\beta,d}(f) = r_\beta(T_{b,d}f).$$

In the rest of this work, we will essentially consider subsets β of the form $\{1,\dots,\nu\}$ or $\{\nu+1,\dots,d+1\}$ for some $\nu \in \{1,\dots,d\}$. For the corresponding β-ranks, we will use the shorthand notations

$$r_\nu(\boldsymbol{f}) := r_{\{1,\dots,\nu\}}(\boldsymbol{f}), \quad r_{\nu,d}(f) = r_{\{1,\dots,\nu\},d}(f).$$

Note that $r_\nu(\boldsymbol{f})$ should not be confused with $r_{\{\nu\}}(\boldsymbol{f})$. The ranks $(r_\nu(\boldsymbol{f}))_{1\le\nu\le d}$ of a tensor $\boldsymbol{f} \in \mathbf{V}_{b,d}$ have to satisfy some relations, as seen in the next lemma.

Lemma 2.12 (Ranks Admissibility Conditions)**.** *Let $\boldsymbol{f} \in \mathbf{V}_{b,d}$. For any set $\beta \subset \{1,\dots,d+1\}$ and any partition $\beta = \gamma \cup \alpha$, we have*

$$r_\beta(\boldsymbol{f}) \le r_\gamma(\boldsymbol{f}) r_\alpha(\boldsymbol{f})$$

and in particular

$$r_{\nu+1}(\boldsymbol{f}) \le b r_\nu(\boldsymbol{f}) \quad \text{and} \quad r_\nu(\boldsymbol{f}) \le b r_{\nu+1}(\boldsymbol{f}), \quad 1 \le \nu \le d-1, \tag{2.8}$$

Proof. See Appendix B. □

A function f admits infinitely many tensorizations of different levels. The following result provides a relation between minimal subspaces.

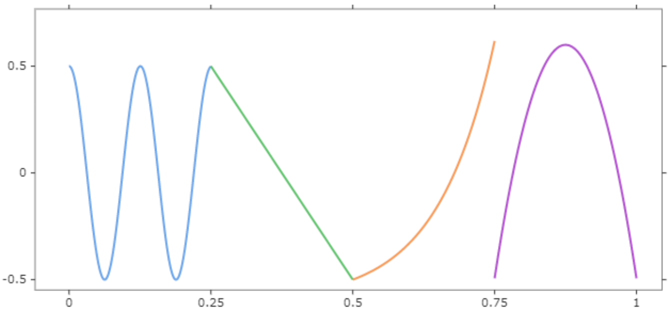
Lemma 2.13. *Consider a (measurable) function $f : [0,1) \to \mathbb{R}$ and its tensorization $\boldsymbol{f}^d = T_{b,d}f$ at level d. For any $1 \le \nu \le d$,*

$$T_{b,d-\nu}^{-1}(U_{\{\nu+1,\dots,d+1\}}^{\min}(\boldsymbol{f}^d)) = \operatorname{span}\{\boldsymbol{f}^\nu(j_1,\dots,j_\nu,\cdot) : \ (j_1,\dots,j_\nu) \in I_b^\nu\} = U_{\{\nu+1\}}^{\min}(\boldsymbol{f}^\nu),$$

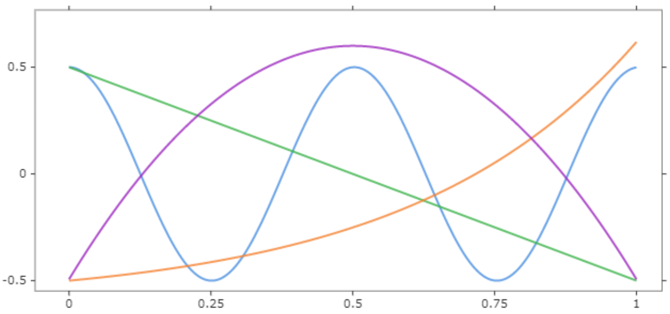
where $\boldsymbol{f}^\nu = T_{b,\nu}f$ is the tensorization of f at level ν.

Proof. See Appendix B. □

For $j = \sum_{k=1}^\nu j_k b^{\nu-k}$, since $\boldsymbol{f}^\nu(j_1,\dots,j_\nu,\cdot) = f(b^{-\nu}(j+\cdot))$ is the restriction of f to the interval $[b^{-\nu}j, b^{-\nu}(j+1))$ rescaled to $[0,1)$, Lemma 2.13 provides a simple interpretation of minimal subspace $U_{\{\nu+1\}}^{\min}(\boldsymbol{f}^\nu)$ as the linear span of contiguous pieces of f rescaled to $[0,1)$, see the illustration in Figure 2.



(A) Function $f : [0, 1) \to \mathbb{R}$



(B) Partial evaluations $\boldsymbol{f}^{\nu}(j_1, j_2, \cdot)$ for $(j_1, j_2) \in \{0, 1\}^2$.

FIGURE 2. A function $f : [0, 1) \to \mathbb{R}$ and partial evaluations of $\boldsymbol{f}^{\nu} \in \mathbf{V}_{b,d}$ for $b = d = 2$.

Corollary 2.14. *Consider a (measurable) function $f : [0, 1) \to \mathbb{R}$ and $d \in \mathbb{N}$. For any $1 \le \nu \le d$,*

$$r_{\nu,d}(f) = r_{\nu,\nu}(f)$$

and

$$r_{\nu,\nu}(f) = \dim \operatorname{span}\{f(b^{-\nu}(j + \cdot)) : 0 \le j \le b^{\nu} - 1\}.$$

Proof. We have

$$r_{\nu,d}(f) = r_{\{1,\ldots,\nu\}}(\boldsymbol{f}^d) = r_{\{\nu+1,\ldots,d+1\}}(\boldsymbol{f}^d) = \dim U^{\min}_{\{\nu+1,\ldots,d+1\}}(\boldsymbol{f}^d)$$

and

$$r_{\nu,\nu}(f) = r_{\{1,\ldots,\nu\}}(\boldsymbol{f}^{\nu}) = r_{\nu+1}(\boldsymbol{f}^{\nu}) = \dim U^{\min}_{\{\nu+1\}}(\boldsymbol{f}^{\nu}).$$

Lemma 2.13 then implies that $r_{\nu,d}(f) = r_{\nu,\nu}(f)$ and provides the characterization from the linear span of $\boldsymbol{f}^{\nu}(j_1, \ldots, j_{\nu}, \cdot) = f(b^{-\nu}(j + \cdot))$, with $j = \sum_{k=1}^{\nu} j_k b^{\nu-k}$, which is linearly identified with the restriction $f_{|[b^{-\nu}j, b^{-\nu}(j+1))}$ shifted and rescaled to $[0, 1)$. □

2.4. **Lebesgue Spaces.** For $0 < p \le \infty$, we consider the Lebesgue space $L^p([0, 1))$ of functions defined on $[0, 1)$ equipped with its standard (quasi-)norm. Then we consider the algebraic tensor space

$$\mathbf{V}_{b,d,L^p} := \mathbb{R}^{I_b}{}^{\otimes d} \otimes L^p([0, 1)) \subset \mathbf{V}_{b,d},$$

which is the space of multivariate functions $\boldsymbol{f}$ on $I_b^d \times [0, 1)$ with partial evaluations $\boldsymbol{f}(j_1, \ldots, j_d, \cdot) \in L^p([0, 1))$. From hereon we frequently abbreviate $L^p := L^p([0, 1))$.

Theorem 2.15 (Tensorization is an L^p-Isometry)**.** *For any $0 < p \le \infty$, $T_{b,d}$ is a linear isometry from $L^p([0, 1))$ to $\mathbf{V}_{b,d,L^p}$ equipped with the (quasi-)norm $\|\cdot\|_p$ defined by*

$$\|\boldsymbol{f}\|_p^p = \sum_{j_1 \in I_b} \cdots \sum_{j_d \in I_b} b^{-d} \|\boldsymbol{f}(j_1, \ldots, j_d, \cdot)\|_p^p$$

for $p < \infty$*, or*

$$\|\boldsymbol{f}\|_\infty = \max_{j_1 \in I_b} \dots \max_{j_d \in I_b} \|\boldsymbol{f}(j_1, \dots, j_d, \cdot)\|_\infty.$$

Proof. The result follows from Proposition 2.7 and by noting that for $\boldsymbol{f} = T_{b,d} f = f \circ t_{b,d}$,

$$\|f\|_p^p = \int_{[0,1)} |f(x)|^p d\lambda(x) = \int_{I_b^d \times [0,1)} |\boldsymbol{f}(j_1, \dots, j_d, y)|^p d\mu_{b,d}(j_1, \dots, j_d, y) = \|\boldsymbol{f}\|_p^p$$

for $p < \infty$, and $\|f\|_\infty = \operatorname{ess\,sup}_x |f(x)| = \operatorname{ess\,sup}_{(j_1,\dots,j_d,y)} |\boldsymbol{f}(j_1, \dots, j_d, y)| = \|\boldsymbol{f}\|_\infty$. $\square$

We denote by $\ell^p(I_b)$ the space $\mathbb{R}^{I_b}$ equipped with the (quasi-)norm $\|\cdot\|_{\ell^p}$ defined for $v = (v_k)_{k \in I_b}$ by

$$\|v\|_{\ell^p}^p := b^{-1} \sum_{k=0}^{b-1} |v_k|^p \quad (p < \infty), \quad \|v\|_{\ell^\infty} := \max_{0 \le k \le b-1} |v_k|.$$

The space $\mathbf{V}_{b,d,L^p}$ can then be identified with the algebraic tensor space

$$(\ell^p(I_b))^{\otimes d} \otimes L^p([0,1)).$$

and $\|\cdot\|_p$ is a crossnorm, i.e., satisfying for an elementary tensor $v^1 \otimes \dots \otimes v^{d+1} \in \mathbf{V}_{b,d,L^p}$,

$$\|v^1 \otimes \dots \otimes v^{d+1}\|_p = \|v^1\|_{\ell^p} \dots \|v^d\|_{\ell^p} \|v^{d+1}\|_p.$$

and even a reasonable crossnorm for $1 \le p \le \infty$ (see Lemma B.1 in the appendix). We let $\{e_k^p\}_{k \in I_b}$ denote the normalized canonical basis of $\ell^p(I_b)$, defined by

$$e_k^p = b^{1/p} \delta_k \text{ for } 0 < p < \infty, \quad \text{and} \quad e_k^\infty = \delta_k \text{ for } p = \infty. \tag{2.9}$$

The tensorization $\boldsymbol{f} = T_{b,d} f$ of a function $f \in L^p([0,1))$ admits a representation

$$\boldsymbol{f} = \sum_{j_1 \in I_b} \dots \sum_{j_d \in I_b} e_{j_1}^p \otimes \dots \otimes e_{j_d}^p \otimes f_{j_1,\dots,j_d}^p, \tag{2.10}$$

with $f_{j_1,\dots,j_d}^p = b^{-d/p} \boldsymbol{f}(j_1, \dots, j_d, \cdot)$ for $p < \infty$ and $f_{j_1,\dots,j_d}^\infty = \boldsymbol{f}(j_1, \dots, j_d, \cdot)$. The crossnorm property implies that

$$\|e_{j_1}^p \otimes \dots \otimes e_{j_d}^p \otimes f_{j_1,\dots,j_d}^p\|_p = \|f_{j_1,\dots,j_d}^p\|_p,$$

so that Theorem 2.15 implies

$$\|f\|_p = \Big(\sum_{(j_1,\dots,j_d) \in I_b^d} \|f_{j_1,\dots,j_d}^p\|_p^p \Big)^{1/p} \quad (p < \infty), \quad \|f\|_\infty = \max_{(j_1,\dots,j_d) \in I_b^d} \|f_{j_1,\dots,j_d}^\infty\|_\infty.$$

2.5. **Tensor Subspaces and Corresponding Function Spaces.** For a linear space of functions $S \subset \mathbb{R}^{[0,1)}$, we define the tensor subspace

$$\mathbf{V}_{b,d,S} := (\mathbb{R}^{I_b})^{\otimes d} \otimes S \subset \mathbf{V}_{b,d},$$

and the corresponding linear subspace of functions in $\mathbb{R}^{[0,1)}$,

$$V_{b,d,S} = T_{b,d}^{-1}(\mathbf{V}_{b,d,S}).$$

In the majority of this work we will be using finite-dimensional subspaces S for approximation. In particular, we will frequently use $S = \mathbb{P}_m$ where $\mathbb{P}_m$ is the space of polynomials of degree up to $m \in \mathbb{N}_0$. In this case we use the shorthand notation

$$V_{b,d,m} := V_{b,d,\mathbb{P}_m}.$$

The tensorization $\boldsymbol{f} = T_{b,d}(f)$ of a function $f \in V_{b,d,S}$ admits a representation (2.2) with functions $\boldsymbol{f}(j_1, \dots, j_d, \cdot) := f_{j_1,\dots,j_d}$ in S. For $x = t_{b,d}(j_1, \dots, j_d, y)$ in the interval $[x_j, x_{j+1})$, with $j = \sum_{k=1}^b j_k b^{d-k}$, we have $f(x) = f_{j_1,\dots,j_d}(y) = f_{j_1,\dots,j_d}(b^d x - j)$. Therefore, the functions $f \in V_{b,d,S}$ have restrictions on intervals $[x_j, x_{j+1})$ that are obtained by shifting and scaling functions in S. In particular, the space $V_{b,d,m}$ corresponds to the space of piecewise polynomials of degree m over the uniform partition of $[0,1)$ with b^d intervals.

For considering functions with variable levels $d \in \mathbb{N}$, we introduce the set

$$V_{b,S} := \bigcup_{d\in\mathbb{N}} V_{b,d,S}.$$

It is straight-forward to see that, in general,

$$V_{b,d,S} \not\subset V_{b,\bar{d},S}$$

for $\bar{d} < d$. E.g., take $S = \mathbb{P}_m$ and let $f \in V_{b,d,m}$ be a piece-wise polynomial but discontinuous function. Then, clearly f does not have to be a polynomial over the intervals

$$[kb^{-\bar{d}}, (k+1)b^{-\bar{d}}), \quad 0 \le k \le b^{\bar{d}} - 1,$$

for $\bar{d} < d$. The same holds for the other inclusion, as the following example demonstrates.

Example 2.16. *Consider the one-dimensional subspace* $S := \operatorname{span}\{\cos(2\pi\cdot)\}$. *A function* $0 \neq f \in V_{b,d,S}$ *is thus a piece-wise cosine. Take for simplicity* $b = 2$, $d = 0$ *(i.e.,* $V_{2,0,S} = S$*) and* $\bar{d} = 1$. *Then,* $f \notin V_{2,1,S}$ *due to* $\operatorname{span}\{\cos(2\pi\cdot)\} \not\supset \operatorname{span}\{\cos(\pi\cdot), \cos(\pi + \pi\cdot)\}$, *since cosines of different frequencies are linearly independent. The same reasoning can be applied to any* $b \ge 2$ *and* $d, \bar{d} \in \mathbb{N}$ *with* $d < \bar{d}$.

This motivates the following definition that is reminiscent of MRAs.

Definition 2.17 (Closed under b-adic dilation)**.** *We say that a linear space S is closed under b-adic dilation if for any $f \in S$ and any $k \in \{0, \ldots, b-1\}$,*

$$f(b^{-1}(\cdot + k)) \in S.$$

Lemma 2.18. *If S is closed under b-adic dilation, then for all $f \in S$,*

$$f(b^{-d}(\cdot + k)) \in S$$

for all $d \in \mathbb{N}$ and $k \in \{0, \ldots, b^d - 1\}$.

Proof. See Appendix B. □

Important examples of spaces S that satisfy the above property include spaces of polynomials and MRAs. The closedness of S under b-adic dilation implies a hierarchy between spaces $V_{b,d,S}$ with different levels, and provides $V_{b,S}$ with a linear space structure.

Proposition 2.19. *If S is closed under b-adic dilation, then*

$$S := V_{b,0,S} \subset V_{b,1,S} \subset \ldots \subset V_{b,d,S} \subset \ldots.$$

Proof. See Appendix B. □

Proposition 2.20 ($V_{b,S}$ is a linear space)**.** *If S is closed under b-adic dilation, then $V_{b,S}$ is a linear space.*

Proof. See Appendix B. □

If $S \subset L^p([0,1))$, then $V_{b,S}$ is clearly a subspace of $L^p([0,1))$. However, it is not difficult to see that, in general, $V_{b,S}$ is not a closed subspace of $L^p([0,1))$. On the other hand, we have the following density result.

Theorem 2.21 ($V_{b,S}$ dense in L^p)**.** *Let $0 < p < \infty$. If $S \subset L^p([0,1))$ and S contains the constant function one, then $V_{b,S} = \bigcup_{d\in\mathbb{N}} V_{b,d,S}$ is dense in $L^p([0,1))$.*

Proof. See Appendix B. □

Now we provide bounds for ranks of functions in $V_{b,S}$, directly deduced from (2.7).

Lemma 2.22. *For $f \in V_{b,d,S}$ and any $\beta \subset \{1,\dots,d\}$,*

$$r_{\beta,d}(f) \le \min\{b^{\#\beta}, b^{d-\#\beta}\dim S\}.$$

In particular, for all $1 \le \nu \le d$,

$$r_{\nu,d}(f) \le \min\{b^{\nu}, b^{d-\nu}\dim S\}.$$

In the case where S is closed under b-adic dilation, we can obtain sharper bounds for ranks.

Lemma 2.23. *Let S be closed under b-adic dilation.*

(i) If $f \in S$, then for any $d \in \mathbb{N}$, $f \in V_{b,d,S}$ and we have

$$r_{\nu,d}(f) \le \min\{b^{\nu}, \dim S\}, \quad 1 \le \nu \le d.$$

(ii) If $f \in V_{b,d,S}$, then for any $\bar{d} \ge d$, $f \in V_{b,\bar{d},S}$ and we have

$$r_{\nu,\bar{d}}(f) = r_{\nu,d}(f) \le \min\left\{b^{\nu}, b^{d-\nu}\dim S\right\}, \quad 1 \le \nu \le d,$$

$$r_{\nu,\bar{d}}(f) \le \min\left\{b^{\nu}, \dim S\right\}, \quad d < \nu \le \bar{d}. \tag{2.11}$$

Proof. See Appendix B. □

Remark 2.24. *Lemma 2.23(ii) shows that the ranks are independent of the representation level of a function φ, so that we will frequently suppress this dependence and simply note $r_{\nu,d}(\varphi) = r_\nu(\varphi)$ for any d such that $\varphi \in V_{b,d,S}$.*

We end this section by introducing projection operators based on local projection. Let $\mathcal{I}_S$ be a linear projection operator from $L^p([0,1))$ to a finite-dimensional space S. Then, let $\mathcal{I}_{b,d,S}$ be a linear operator defined for $f \in L^p([0,1))$ by

$$(\mathcal{I}_{b,d,S}f)(b^{-d}(j+\cdot)) = \mathcal{I}_S(f(b^{-d}(j+\cdot))), \quad 0 \le j < b^d. \tag{2.12}$$

Lemma 2.25 (Local projection)**.** *The operator $\mathcal{I}_{b,d,S}$ is a linear operator from $L^p([0,1))$ to $V_{b,d,S}$ and satisfies*

$$T_{b,d} \circ \mathcal{I}_{b,d,S} \circ T_{b,d}^{-1} = id_{\{1,\dots,d\}} \otimes \mathcal{I}_S. \tag{2.13}$$

Proof. See Appendix B. □

We now provide a result on the ranks of projections.

Lemma 2.26 (Local projection ranks)**.** *For any $f \in L^p$, $\mathcal{I}_{b,d,S}f \in V_{b,d,S}$ satisfies*

$$r_{\nu,d}(\mathcal{I}_{b,d,S}f) \le r_{\nu,d}(f), \quad 1 \le \nu \le d.$$

Proof. Lemma 2.25 implies that $T_{b,d} \circ \mathcal{I}_{b,d,S} \circ T_{b,d}^{-1}$ is a rank one operator. Since a rank-one operator can not increase β-ranks, we have for all $1 \le \nu \le d$

$$r_{\nu,d}(\mathcal{I}_{b,d,S}(f)) = r_\nu(T_{b,d} \circ \mathcal{I}_{b,d,S} \circ T_{b,d}^{-1}\boldsymbol{f}) = r_\nu((id_{\{1,\dots,d\}} \otimes \mathcal{I}_S)\boldsymbol{f}) \le r_\nu(\boldsymbol{f}) = r_{\nu,d}(f).$$

□

3. Tree Tensor Networks and Their Approximation spaces

In this section, we begin by describing particular tensor formats, namely tree tensor networks that will constitute our approximation tool. We then briefly review classical approximation spaces (see [19]). We conclude by introducing different measures of complexity of tree tensor networks and analyze the resulting approximation classes.

3.1. **Tree Tensor Networks and The Tensor Train Format.** Let S be a finite-dimensional subspace of $\mathbb{R}^{[0,1)}$ and A a dimension partition tree (or a subtree of such a tree). A hierarchical or tree-based tensor format [35, 24] in the tensor space $\mathbf{V}_{b,d,S} = (\mathbb{R}^{I_b})^{\otimes d} \otimes S$ is defined as a set of tensors with β-ranks bounded by some integers r_β, for a certain collection A of subsets $\beta \subset \{1, \ldots, d+1\}$,

$$\mathcal{T}^A_{\boldsymbol{r}}(\mathbf{V}_{b,d,S}) = \{\boldsymbol{f} \in \mathbf{V}_{b,d,S} : r_\beta(\boldsymbol{f}) \leq r_\beta, \beta \in A\}.$$

A tensor $\boldsymbol{f} \in \mathcal{T}^A_{\boldsymbol{r}}(\mathbf{V}_{b,d,S})$ in a tree-based tensor format admits a parametrization in terms of a collection of low-order tensors v_β, $\beta \in A$. Hence, the interpretation as a tree tensor network (see [47, Section 4]).

For the most part we will work with the tensor train format with the exception of a few remarks. This format considers the collection of subsets $A = \{\{1\}, \{1,2\}, \ldots, \{1, \ldots, d\}\}$.

Definition 3.1 (Tensor Train Format). *The set[4] of tensors in $\mathbf{V}_{b,d}$ in tensor train (TT) format with ranks at most $\boldsymbol{r} := (r_\nu)_{\nu=1}^d$ is defined as*

$$\mathcal{TT}_{\boldsymbol{r}}(\mathbf{V}_{b,d,S}) := \{\boldsymbol{f} \in \mathbf{V}_{b,d,S} : \ r_\nu(\boldsymbol{f}) \leq r_\nu, \ 1 \leq \nu \leq d\},$$

where we have used the shorthand notation $r_\nu(\boldsymbol{f}) := r_{\{1,\ldots,\nu\}}(\boldsymbol{f})$. This defines a set of univariate functions

$$\Phi_{b,d,S,\boldsymbol{r}} = T^{-1}_{b,d}(\mathcal{TT}_{\boldsymbol{r}}(\mathbf{V}_{b,d,S})) = \{f \in V_{b,d,S} : r_\nu(f) \leq r_\nu, 1 \leq \nu \leq d\},$$

where $r_\nu(f) := r_{\nu,d}(f)$, that we further call the tensor train format for univariate functions.

Letting $\{\varphi_k\}_{k=1}^{\dim S}$ be a basis of S, a tensor $\boldsymbol{f} \in \mathcal{TT}_{\boldsymbol{r}}(\mathbf{V}_{b,d,S})$ admits a representation

$$\boldsymbol{f}(i_1, \ldots, i_d, y) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} \sum_{k=1}^{\dim S} v_1^{k_1}(i_1) v_2^{k_1,k_2}(i_2) \cdots v_d^{k_{d-1},k_d}(i_d) v_{d+1}^{k_d,k} \varphi_k(y), \tag{3.1}$$

with parameters $v_1 \in \mathbb{R}^{b \times r_1}$, $v_\nu \in \mathbb{R}^{b \times r_{\nu-1} \times r_\nu}$, $2 \leq \nu \leq d$, and $v_{d+1} \in \mathbb{R}^{r_d \times \dim S}$ forming a tree tensor network

$$\mathbf{v} = (v_1, \ldots, v_{d+1}) \in \mathcal{P}_{b,d,S,\boldsymbol{r}} := \mathbb{R}^{b \times r_1} \times \mathbb{R}^{b \times r_1 \times r_2} \times \ldots \times \mathbb{R}^{b \times r_{d-1} \times r_d} \times \mathbb{R}^{r_d \times \dim S}.$$

The format $\mathcal{TT}_{\boldsymbol{r}}(\mathbf{V}_{b,d,S})$ then corresponds to the image of the space of tree tensor networks $\mathcal{P}_{b,d,S,\boldsymbol{r}}$ through the map

$$R_{b,d,S,\boldsymbol{r}} : \mathcal{P}_{b,d,S,\boldsymbol{r}} \to \mathcal{TT}_{\boldsymbol{r}}(\mathbf{V}_{b,d,S}) \subset \mathbf{V}_{b,d,S}$$

such that for $\mathbf{v} = (v_1, \ldots, v_{d+1}) \in \mathcal{P}_{b,d,S,\boldsymbol{r}}$, the tensor $\boldsymbol{f} = R_{b,d,S,\boldsymbol{r}}(\mathbf{v})$ is defined by (3.1). The set of functions $\Phi_{b,d,S,\boldsymbol{r}}$ in the tensor train format can be parametrized as follows:

$$\Phi_{b,d,S,\boldsymbol{r}} = \{\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) : \mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}\}, \quad \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) = T^{-1}_{b,d} \circ R_{b,d,S,\boldsymbol{r}}.$$

With an abuse of terminology, we call tensor networks both the set of tensors $\mathbf{v}$ and the corresponding function $\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v})$. The representation complexity of $\boldsymbol{f} = R_{b,d,S,\boldsymbol{r}}(\mathbf{v}) \in \mathcal{TT}_{\boldsymbol{r}}(V_{b,d,S})$ is

$$\mathcal{C}(b, d, S, \boldsymbol{r}) := \dim(\mathcal{P}_{b,d,S,\boldsymbol{r}}) = br_1 + b \sum_{\nu=2}^{d} r_{\nu-1} r_\nu + r_d \dim S. \tag{3.2}$$

[4]It is in fact a manifold, see [38, 23, 26, 25]. This definition is a continuous version of the QTT format.

Remark 3.2 (Re-Ordering Variables in the TT Format). *We chose in Definition 2.2 to order the input variables of the tensorized function $\boldsymbol{f}$ such that $y \in [0,1)$ is in the last position. This specific choice allows the interpretation of partial evaluations of $\{1,\ldots,\nu\}$-unfoldings as contiguous pieces of $f = T_{b,d}^{-1}(\boldsymbol{f})$ (see Lemma 2.13 and the discussion thereafter). Alternatively, we could have chosen the ordering $(y, i_1, \ldots, i_d) \mapsto \boldsymbol{f}(y, i_1, \ldots, i_d)$, and defined the TT-format and TT-ranks correspondingly. Essentially this is the same as considering a different tensor format, see discussion above. Many of the results of Sections 4 and 5 remain the same. In particular, the order of magnitude of the rank bounds and therefore the resulting direct and inverse estimates would not change. However, this re-ordering may lead to slightly smaller rank bounds as in Remark 3.9 or slightly larger rank bounds as in Remark 4.7.*

3.2. **General Approximation Spaces.** Let X be a quasi-normed linear space, $\Phi_n \subset X$ subsets of X for $n \in \mathbb{N}_0$ and $\Phi := (\Phi_n)_{n\in\mathbb{N}_0}$. Define the best approximation error

$$E_n(f) := E(f, \Phi_n) := \inf_{\varphi\in\Phi_n} \|f - \varphi\|_X \,.$$

With this we define approximation classes as

Definition 3.3 (Approximation Classes). *For any $f \in X$ and $\alpha > 0$, define the quantity*

$$\|f\|_{A^\alpha_q} := \begin{cases} \left(\sum_{n=1}^\infty [n^\alpha E_{n-1}(f)]^q \frac{1}{n}\right)^{1/q}, & 0 < q < \infty, \\ \sup_{n\geq 1}[n^\alpha E_{n-1}(f)], & q = \infty. \end{cases}$$

The approximation classes A^α_q of $\Phi = (\Phi_n)_{n\in\mathbb{N}_0}$ are defined by

$$A^\alpha_q := A^\alpha_q(X) := A^\alpha_q(X, \Phi) := \left\{ f \in X : \ \|f\|_{A^\alpha_q} < \infty \right\}.$$

These approximation classes have many useful properties if we further assume that (Φ_n) satisfy the following criteria for any $n \in \mathbb{N}_0$.

(P1) $0 \in \Phi_n$, $\Phi_0 = \{0\}$.
(P2) $\Phi_n \subset \Phi_{n+1}$.
(P3) $a\Phi_n = \Phi_n$ for any $a \in \mathbb{R} \setminus \{0\}$.
(P4) $\Phi_n + \Phi_n \subset \Phi_{cn}$ for some $c := c(\Phi)$.
(P5) $\bigcup_{n\in\mathbb{N}_0} \Phi_n$ is dense in X.
(P6) Φ_n is proximinal in X, i.e. each $f \in X$ has a best approximation in Φ_n.

Additionally, properties (P1) – (P6) will be frequently combined with the so-called *direct* or *Jackson* inequality

$$E_n(f) \leq Cn^{-k_J} |f|_Y \,, \quad \forall f \in Y, \tag{3.3}$$

for a semi-normed vector space Y and some parameter $k_J > 0$, and the *inverse* or *Bernstein* inequality

$$|\varphi|_Y \leq Cn^{k_B} \|\varphi\|_X \,, \quad \forall \varphi \in \Phi_n, \tag{3.4}$$

for some parameter $k_B > 0$.

The implications of (P1) – (P4) about the properties of A^α_q are as follows

- (P1)+(P3)+(P4) $\Rightarrow$ A^α_q is a linear space with a quasi-norm.
- (P1)+(P3)+(P4) $\Rightarrow$ A^α_q satisfies the *direct* or *Jackson* inequality

$$E_n(f) \leq Cn^{-\alpha} \|f\|_{A^\alpha_q} \,, \quad \forall f \in A^\alpha_q.$$

- (P1)+(P2)+(P3)+(P4) $\Rightarrow$ A^α_q satisfies the *inverse* or *Bernstein* inequality

$$\|\varphi\|_{A^\alpha_q} \leq Cn^\alpha \|\varphi\|_X \,, \quad \forall \varphi \in \Phi_n.$$

(P1) – (P4) together with a Jackson estimate as in (3.3) are required to prove so-called *direct embeddings*. (P1) – (P6) together with a Bernstein estimate (3.4) are required for *inverse embeddings*. We will see in Section 7 that, in general, approximation spaces of tensor networks are not embedded in smoothness (Besov) spaces. (P5) is typically true for any type of reasonable approximation tool[5]. We have the continuous embeddings

$$A_q^\alpha \hookrightarrow A_{\bar{q}}^\beta, \quad \text{if } \alpha > \beta \quad \text{or if } \alpha = \beta \text{ and } q \leq \bar{q}.$$

We will see that, while most properties are easy to satisfy, property (P4) will be the most critical one. In essence (P4) is a restriction on the nonlinearity of the sets Φ_n, with $c(\Phi) = 1$ being satisfied by linear subspaces.

3.2.1. *Necessity of (P4).* We could consider replacing n with n^2 in (P4), i.e.,

$$\Phi_n + \Phi_n \subset \Phi_{cn^2}. \tag{3.5}$$

This implies that A_q^α as defined in Definition 3.3 is no longer a vector space. The statements about Jackson and Bernstein inequalities as well as the relation to interpolation and smoothness spaces are no longer valid.

One could try to recover the linearity of A_q^α by modifying Definition 3.3. In Definition 3.3 we measure *algebraic* decay of $E_n(f)$. Algebraic decay is compatible with (P4) that in turn ensures A_q^α is a vector space. We could reverse this by asking: what type of decay behavior is "compatible" with (3.5) in the sense that the corresponding approximation class would be a linear space? We can introduce a growth function $\gamma : \mathbb{N} \to \mathbb{R}^+$ with $\lim_n \gamma(n) = \infty$ and define an approximation class A_∞^γ of $\Phi = (\Phi_n)_{n\in\mathbb{N}_0}$ as

$$A_\infty^\gamma := \left\{ f \in X : \sup_{n\geq 1} \gamma(n) E_{n-1}(f) < \infty \right\}.$$

With some elementary computations one can deduce that if the growth function is of the form

$$\gamma(n) := 1 + \ln(n),$$

then (3.5) implies A_∞^γ is closed under addition. However, functions in A_∞^γ have too slowly decaying errors for any practical purposes such that we do not intend to analyze this space further.

We could instead ask what form of (P4) would be compatible with a growth function such as

$$\gamma(n) := \exp(an^\alpha),$$

for some $a > 0$ and $\alpha > 0$, i.e., classes of functions with exponentially decaying errors. In this case we would have to require $c = 1$ in (P4), i.e.,

$$\Phi_n + \Phi_n \subset \Phi_n,$$

in other words, Φ_n is a linear space.

These considerations suggest that preserving (P4) in its original form is necessary to exploit the full potential of classical approximation theory while preserving some flexibility in defining Φ_n. Thus, we only consider definitions of approximation tools that satisfy (P4).

[5]Think of *universality theorems* for neural networks which hold for tensor networks as well as we will see shortly.

3.3. **Measures of Complexity.** We consider as an approximation tool Φ the collection of tensor networks $\Phi_{b,d,S,\boldsymbol{r}}$ associated with different levels and ranks,

$$\Phi := (\Phi_{b,d,S,\boldsymbol{r}})_{d\in\mathbb{N},\boldsymbol{r}\in\mathbb{N}^d},$$

and define the sets of functions Φ_n as

$$\Phi_n := \left\{\varphi \in \Phi_{b,d,S,\boldsymbol{r}} : d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d, \mathrm{compl}(\varphi) \leq n\right\}, \tag{3.6}$$

where $\mathrm{compl}(\varphi)$ is some measure of complexity of a function φ. The approximation classes of tensor networks depend on the chosen measure of complexity. We propose different measures of complexity and discuss the critical property (P4). We will then only retain definitions of $\mathrm{compl}(\cdot)$ such that the corresponding approximation tool satisfies (P4).

A function $\varphi \in \Phi$ may admit representations at different levels. We set

$$d(\varphi) = \min\{d : \varphi \in V_{b,d,S}\}$$

to be the minimal representation level of φ, and $\boldsymbol{r}(\varphi) = (r_\nu(\varphi))_{\nu=1}^{d(\varphi)}$ be the corresponding ranks. Measures of complexity may also be based on a measure of complexity $\mathrm{compl}(\mathbf{v})$ of tensor networks $\mathbf{v}$ such that $\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v})$. In this case, we would define

$$\Phi_n := \left\{\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) \in \Phi_{b,d,S,\boldsymbol{r}} : \mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}, d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d, \mathrm{compl}(\mathbf{v}) \leq n\right\}. \tag{3.7}$$

which is equivalent to the definition (3.6) if we let

$$\mathrm{compl}(\varphi) := \min\{\mathrm{compl}(\mathbf{v}) : \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) = \varphi, d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d\}, \tag{3.8}$$

where the minimum is taken over all possible representations of φ.

3.3.1. *Complexity Measure: Maximum Rank.* In many high-dimensional approximation problems it is common to consider the maximum rank as an indicator of complexity (see, e.g., [6]). By this analogy we consider for $\varphi \in \Phi$,

$$\mathrm{compl}(\varphi) := bdr_{\max}^2(\varphi) + r_{\max}(\varphi)\dim S, \quad r_{\max}(\varphi) = \max\left\{r_\nu(\varphi) :\ 1 \leq \nu \leq d(\varphi)\right\}. \tag{3.9}$$

This complexity measure does not satisfy (P4).

Proposition 3.4 ((P4) not satisfied by the complexity measure based on $r_{\max}$)**.** *Let S be closed under b-adic dilation and assume* $\dim S < \infty$*. Then, with Φ_n as defined in* (3.6) *with the measure of complexity* (3.9),

(i) there exists no constant $c \in \mathbb{R}$ such that $\Phi_n + \Phi_n \subset \Phi_{cn}$.
(ii) There exists a constant $c > 1$ such that $\Phi_n + \Phi_n \subset \Phi_{cn^2}$.

Proof. See Appendix C. □

3.3.2. *Complexity Measure: Sum of Ranks.* For a neural network, a natural measure of complexity is the number of neurons. By analogy, we can define a complexity measure equal to the sum of ranks

$$\mathrm{compl}_{\mathcal{N}}(\varphi) := \sum_{\nu=1}^{d(\varphi)} r_\nu(\varphi), \tag{3.10}$$

and the corresponding set

$$\Phi_n^{\mathcal{N}} := \left\{\varphi \in \Phi_{b,d,S,\boldsymbol{r}} : d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d, \mathrm{compl}_{\mathcal{N}}(\varphi) \leq n\right\}. \tag{3.11}$$

This complexity measure can be equivalently defined by (3.8) with $\mathrm{compl}_{\mathcal{N}}(\mathbf{v}) = \sum_{\nu=1}^{d} r_\nu$ for $\mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}$.

Lemma 3.5 ($\Phi_n^{\mathcal{N}}$ satisfies (P4))**.** *Let S be closed under b-adic dilation and* $\dim S < \infty$*. Then, the set $\Phi_n^{\mathcal{N}}$ as defined in* (3.11) *satisfies (P4) with $c = 2 + \dim S$.*

Proof. See Appendix C. □

3.3.3. *Complexity Measure: Representation Complexity.* A straight-forward choice for the complexity measure is the number of parameters required for representing φ as in (3.2), i.e.,

$$\operatorname{compl}_{\mathcal{C}}(\varphi) := \mathcal{C}(b, d(\varphi), S, \boldsymbol{r}(\varphi)) = br_1(\varphi) + b\sum_{k=2}^{d(\varphi)} r_{k-1}(\varphi)r_k(\varphi) + r_d(\varphi)\dim S, \tag{3.12}$$

and the corresponding set is defined as

$$\Phi_n^{\mathcal{C}} := \left\{\varphi \in \Phi_{b,d,S,\boldsymbol{r}} : d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d, \operatorname{compl}_{\mathcal{C}}(\varphi) \leq n\right\}. \tag{3.13}$$

Remark 3.6. *This complexity measure can be equivalently defined by* (3.8) *with* $\operatorname{compl}_{\mathcal{C}}(\mathbf{v}) = \mathcal{C}(b, d, S, \boldsymbol{r})$ *for $\mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}$. This means that* $\operatorname{compl}_{\mathcal{C}}(\varphi) = \operatorname{compl}_{\mathcal{C}}(\mathbf{v})$ *for a tensor network $\mathbf{v}$ corresponding to a minimal representation of φ, i.e. the sizes of tensors in the tensor network are equal to the ranks of the represented function φ.*

Remark 3.7. *When interpreting tensor networks as neural networks, the complexity measure* $\operatorname{compl}_{\mathcal{C}}$ *is equivalent to the number of weights for a fully connected neural network with r_ν neurons in layer ν.*

We can show the set $\Phi_n^{\mathcal{C}}$ satisfies (P4) with the help of Lemmas 2.12 and 2.23.

Lemma 3.8 ($\Phi_n^{\mathcal{C}}$ satisfies (P4))**.** *Let S be closed under b-adic dilation and* $\dim S < \infty$*. Then, the set $\Phi_n^{\mathcal{C}}$ as defined in* (3.13) *satisfies (P4) with $c = c(b, \dim S) > 1$.*

Proof. See Appendix C. □

Remark 3.9 (Re-Ordering Input Variables)**.** *In the proof of Lemma 3.8, we have used the property* (2.11) *from Lemma 2.12. As mentioned in Remark 3.2, we could consider a different ordering of the input variables $(y, i_1, \dots, i_d) \mapsto \boldsymbol{f}(y, i_1, \dots, i_d)$, and the corresponding TT-format. This would change* (2.11) *to*

$$r_{\nu,\bar{d}}(f) = 1, \quad d + 1 \leq \nu \leq \bar{d}.$$

We still require S to be closed under b-adic dilation to ensure $f \in V_{b,\bar{d},S}$.

Remark 3.10 (ℓ^2-norm of Ranks)**.** *One could, in principle, also consider defining the complexity measure as a ℓ^2-norm of the tuple of ranks*

$$\operatorname{compl}(\varphi) := b\sum_{k=1}^{d(\varphi)} r_k(\varphi)^2 + r_d(\varphi)\dim S.$$

This definition satisfies (P4) as well with analogous results as for the complexity measure $\operatorname{compl}_{\mathcal{C}}$ *for direct and inverse embeddings. The ℓ^2-norm of ranks is less sensitive to rank-anisotropy than the representation complexity* $\operatorname{compl}_{\mathcal{C}}(\varphi)$*. Note that both complexity measures reflect the cost of representing a function with tensor networks, not the cost of performing arithmetic operations, where frequently an additional power of r is required (e.g., $\sim r^3$ or higher).*

3.3.4. *Complexity Measure: Sparse Representation Complexity.* Finally, for a function $\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) \in \Phi_{b,d,S,\boldsymbol{r}}$, we consider a complexity measure that takes into account the sparsity of the tensors $\mathbf{v} = (v_1, \ldots, v_{d+1})$,

$$\text{compl}_{\mathcal{S}}(\mathbf{v}) := \sum_{\nu=1}^{d+1} \|v_\nu\|_{\ell_0}, \tag{3.14}$$

where $\|v_\nu\|_{\ell_0}$ is the number of non-zero entries in the tensor v_ν. By analogy with neural networks, this corresponds to the number of non-zero weights for sparsely connected neural networks. We define the corresponding set as

$$\Phi_n^{\mathcal{S}} := \left\{\varphi \in \Phi_{b,d,S,\boldsymbol{r}} : d \in \mathbb{N}, \boldsymbol{r} \in \mathbb{N}^d, \text{compl}_{\mathcal{S}}(\varphi) \le n\right\}, \tag{3.15}$$

with $\text{compl}_{\mathcal{S}}(\varphi)$ defined by (3.8). We can show the set $\Phi_n^{\mathcal{S}}$ satisfies (P4). For that, we need the following two lemmas.

Lemma 3.11. *Assume S is closed under b-adic dilation and* $\dim S < \infty$. *Let* $\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) \in \Phi_{b,d,S,\boldsymbol{r}}$ *with* $\boldsymbol{r} = (r_\nu)_{\nu=1}^d$. *For $\bar{d} > d$, there exists a representation* $\varphi = \mathcal{R}_{b,\bar{d},S,\bar{\boldsymbol{r}}}(\bar{\mathbf{v}}) \in \Phi_{b,\bar{d},S,\bar{\boldsymbol{r}}}$ *with* $\bar{\boldsymbol{r}} = (\bar{r}_\nu)_{\nu=1}^{\bar{d}}$ *such that* $\bar{r}_\nu = r_\nu$ *for* $1 \le \nu \le d$ *and* $\bar{r}_\nu \le \max\{\dim S, b\} \dim S$ *for* $d < \nu \le \bar{d}$, *and*

$$\text{compl}_{\mathcal{S}}(\bar{\mathbf{v}}) \le b\, \text{compl}_{\mathcal{S}}(\mathbf{v}) + (\bar{d} - d) b^2 (\dim S)^3.$$

Proof. See Appendix C. □

Lemma 3.12 (Sum of Sparse Representations). *Let* $\varphi_A = \mathcal{R}_{b,d,S,\boldsymbol{r}^A}(\mathbf{v}_A) \in \Phi_{b,d,S,\boldsymbol{r}^A}$ *and* $\varphi_B = \mathcal{R}_{b,d,S,\boldsymbol{r}^B}(\mathbf{v}_B) \in \Phi_{b,d,S,\boldsymbol{r}^B}$. *Then,* $\varphi_A + \varphi_B$ *admits a representation* $\varphi_A + \varphi_B = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) \in \Phi_{b,d,S,\boldsymbol{r}}$ *with* $r_\nu = r_\nu^A + r_\nu^B$ *for* $1 \le \nu \le d$, *and*

$$\text{compl}_{\mathcal{S}}(\varphi_A + \varphi_B) \le \text{compl}_{\mathcal{S}}(\mathbf{v}) \le \text{compl}_{\mathcal{S}}(\mathbf{v}_A) + \text{compl}_{\mathcal{S}}(\mathbf{v}_B).$$

Proof. See Appendix C. □

Lemma 3.13 ($\Phi_n^{\mathcal{S}}$ satisfies (P4)). *Let S be closed under b-adic dilation and* $\dim S < \infty$. *Then, the set $\Phi_n^{\mathcal{S}}$ as defined in* (3.15) *satisfies (P4) with* $c = b + 1 + b^2(\dim S)^3$.

Proof. Let $\varphi_A,\ \varphi_B \in \Phi_n^{\mathcal{S}}$ with $\varphi_A = \mathcal{R}_{b,d_A,S,\boldsymbol{r}^A}(\mathbf{v}_A) \in \Phi_{b,d_A,S,\boldsymbol{r}^A}$ and $\varphi_B = \mathcal{R}_{b,d_B,S,\boldsymbol{r}^B}(\mathbf{v}_B) \in \Phi_{b,d_B,S,\boldsymbol{r}^B}$ and w.l.o.g. $d_A \le d_B$. From Lemmas 3.11 and 3.12, we know that $\varphi_A + \varphi_B$ admits a representation $\varphi_A + \varphi_B = \mathcal{R}_{b,d_B,S,\boldsymbol{r}}(\mathbf{v})$ at level d_B with

$$\begin{aligned}\text{compl}_{\mathcal{S}}(\mathbf{v}) &\le b\, \text{compl}_{\mathcal{S}}(\mathbf{v}_A) + \text{compl}_{\mathcal{S}}(\mathbf{v}_B) + (d_B - d_A) b^2 (\dim S)^3 \\ &\le (b + 1 + b^2(\dim S)^3) n,\end{aligned}$$

which ends the proof. □

3.4. **Approximation Spaces of Tensor Trains.** We denote by $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$ the approximation set Φ_n asscociated with the measures of complexity $\text{compl}_{\mathcal{N}}$, $\text{compl}_{\mathcal{C}}$ and $\text{compl}_{\mathcal{S}}$ respectively. Then, for a quasi-normed linear space X, we define three different families of approximation classes

$$N_q^\alpha(X) := A_q^\alpha(X, (\Phi_n^{\mathcal{N}})_{n\in\mathbb{N}}), \tag{3.16}$$

$$C_q^\alpha(X) := A_q^\alpha(X, (\Phi_n^{\mathcal{C}})_{n\in\mathbb{N}}), \tag{3.17}$$

$$S_q^\alpha(X) := A_q^\alpha(X, (\Phi_n^{\mathcal{S}})_{n\in\mathbb{N}}), \tag{3.18}$$

with $\alpha > 0$ and $0 < q \le \infty$. Below, we will show that these approximation classes are in fact approximation spaces and we will then compare these spaces.

3.4.1. *Approximation Classes are Approximation Spaces.* We proceed with checking if $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$ satisfy properties (P1)–(P6). In particular, satisfying (P1)–(P4) will imply that the corresponding approximation classes are quasi-normed Banach spaces. The only property – other than (P4) – that is not obvious, is (P6). This is addressed in the following Lemma for $\Phi_n^{\mathcal{N}}$ and $\Phi_n^{\mathcal{C}}$.

Lemma 3.14 ($\Phi_n^{\mathcal{N}}$ and $\Phi_n^{\mathcal{C}}$ satisfy (P6))**.** *If $1 < p < \infty$ and S is a closed subspace of L^p, then $\Phi_n^{\mathcal{N}}$ and $\Phi_n^{\mathcal{C}}$ are proximinal in L^p for any $n \in \mathbb{N}$. Moreover, if S is finite-dimensional, the above is true for all $0 < p \leq \infty$.*

Proof. See Appendix C. □

As the following example shows, we cannot in general guarantee (P6) for $\Phi_n^{\mathcal{S}}$.

Example 3.15. *Suppose $b \geq 3$ and $\dim S \geq 3$. Take two linearly independent vectors $v, w \in \mathbb{R}^b$ and $f, g \in S$. For any $N \in \mathbb{N}$, set*

$$\varphi_N := (w + Nv) \otimes (v + \frac{1}{N} w) \otimes f + v \otimes v \otimes (g - Nf),$$

and

$$\varphi := v \otimes v \otimes g + v \otimes w \otimes f + w \otimes v \otimes f.$$

Then, we have the following (see [34, Proposition 9.10 and Remark 12.4]*).*

- *(i) For the canonical tensor rank, we have $r(\varphi_N) = 2$ for any $N \in \mathbb{N}$ and $r(\varphi) = 3$.*
- *(ii) As we will see in Lemma 3.22, $\varphi_N \in \Phi_{6b+2\dim S}^{\mathcal{S}}$ for any $N \in \mathbb{N}$ and $\varphi \in \Phi_{9b+3\dim S}^{\mathcal{S}}$. Moreover, this complexity is minimal for both functions.*
- *(iii) For $N \to \infty$, $\varphi_N \to \varphi$ in any norm*[6]*.*

In other words, $E_{6b+2\dim S}^{\mathcal{S}}(\varphi) = 0$, even though $\varphi \notin \Phi_{6b+2\dim S}^{\mathcal{S}}$.

Remark 3.16. *(P6) is required for showing that the approximation spaces A_q^α are continuously embedded into interpolation spaces, see* [19, Chapter 7, Theorem 9.3]*. As a side note, (P6) does not hold for ReLU or RePU networks as was discussed in* [32]*.*

We now derive the main result of this section.

Theorem 3.17 (Properties of $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$)**.** *Let $0 < p \leq \infty$, $S \subset L^p$ be a closed subspace that is also closed under b-adic dilation and $\dim S < \infty$. Then,*

- *(i) $\Phi_n^{\mathcal{N}}$ and $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$ satisfy (P1) – (P4).*
- *(ii) $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ additionally satisfy (P6).*
- *(iii) If $0 < p < \infty$ and if S contains the constant function one, $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$ additionally satisfy (P5).*

Proof. (P1) – (P3) are obvious and (P4) follows from Lemmas 3.5, 3.8 and 3.13, that yields (i). (iii) follows from the fact that

$$\bigcup_{n\in\mathbb{N}} \Phi_n = \bigcup_{d\in\mathbb{N}} \bigcup_{\boldsymbol{r}\in\mathbb{N}^d} \Phi_{b,d,S,\boldsymbol{r}} = \bigcup_{d\in\mathbb{N}} V_{b,d,S} = V_{b,S},$$

and from Theorem 2.21. Finally, (ii) follows from Lemma 3.14. □

Theorem 3.17(i) implies that the approximation classes $N_q^\alpha(L^p)$, $C_q^\alpha(L^p)$ and $S_q^\alpha(L^p)$ are quasi-normed vector spaces that satisfy the Jackson and Bernstein inequalities.

[6] Both φ_N and φ belong to a finite-dimensional vector space.

3.4.2. *Comparing Approximation Spaces.* For comparing approximation spaces $N^\alpha_q(L^p)$, $C^\alpha_q(L^p)$ and $S^\alpha_q(L^p)$, we first provide some relations between the sets $\Phi^{\mathcal{N}}_n$, $\Phi^{\mathcal{C}}_n$ and $\Phi^{\mathcal{S}}_n$.

Proposition 3.18. *For any $n \in \mathbb{N}$,*

$$\Phi^{\mathcal{C}}_n \subset \Phi^{\mathcal{S}}_n \subset \Phi^{\mathcal{N}}_n \subset \Phi^{\mathcal{C}}_{b \dim S + bn^2}.$$

Proof. See Appendix C. □

From Proposition 3.18 and the definition of approximation spaces, we obtain[7]

Theorem 3.19. *For any $\alpha > 0$, $0 < p \leq \infty$ and $0 < q \leq \infty$, the classes $N^\alpha_q(L^p)$, $C^\alpha_q(L^p)$ and $S^\alpha_q(L^p)$ satisfy the continuous embeddings*

$$C^\alpha_q(L^p) \hookrightarrow S^\alpha_q(L^p) \hookrightarrow N^\alpha_q(L^p) \hookrightarrow C^{\alpha/2}_q(L^p).$$

3.5. **About The Canonical Tensor Format.** We conclude by comparing tensor networks with the canonical tensor format

$$\mathcal{T}_r(V_{b,d,S}) = \{\boldsymbol{f} \in \mathbf{V}_{b,d,S} : r(\boldsymbol{f}) \leq r\},$$

which is the set of tensors that admit a representation

$$\boldsymbol{f}(i_1, \ldots, i_d, y) = \sum_{k=1}^{r} w_1^k(i_1) \ldots w_d^k(i_d) g_{d+1}^k(y), \quad g_{d+1}^k(y) = \sum_{q=1}^{\dim S} w_{d+1}^{q,k} \varphi_q(y),$$

with $w_\nu \in \mathbb{R}^{b \times r}$ for $1 \leq \nu \leq d$ and $w_{d+1} \in \mathbb{R}^{\dim S \times r}$. The canonical tensor format can be interpreted as a shallow sum-product neural network (or arithmetic circuit), see [16].

We let $R_{b,d,S,r}$ be the map from $(\mathbb{R}^{b\times r})^d \times \mathbb{R}^{\dim S \times r} := \mathcal{P}_{b,d,S,r}$ to $\mathbf{V}_{b,d,S}$ which associates to a set of tensors $(w_1, \ldots, w_{d+1})$ the tensor $\boldsymbol{f} = R_{b,d,S,r}(w_1, \ldots, w_{d+1})$ as defined above. We introduce the sets of functions

$$\Phi_{b,d,S,r} = T_{b,d}^{-1} \mathcal{T}_r(V_{b,d,S}),$$

which can be parametrized as follows:

$$\Phi_{b,d,S,r} = \{\varphi = \mathcal{R}_{b,d,S,r}(\mathbf{w}) : \mathbf{w} \in \mathcal{P}_{b,d,S,r}\}, \quad \mathcal{R}_{b,d,S,r} = T_{b,d}^{-1} \circ R_{b,d,S,r}.$$

For $\varphi \in V_{b,S}$, we let

$$r(\varphi) = \min\{r(\boldsymbol{f}) : \boldsymbol{f} \in \mathbf{V}_{b,d(\varphi),S}, T_{b,d}^{-1}(\boldsymbol{f}) = \varphi\}.$$

We introduce as a natural complexity measure the representation complexity

$$\mathrm{compl}_{\mathcal{R}}(\varphi) = bd(\varphi) r(\varphi) + r(\varphi) \dim S,$$

define the sets

$$\Phi^{\mathcal{R}}_n = \{\varphi \in \Phi_{b,d,S,r} : d \in \mathbb{N}, r \in \mathbb{N}, \mathrm{compl}_{\mathcal{R}}(\varphi) \leq n\},$$

and consider the corresponding approximation classes

$$R^\alpha_q(L^p) = A^\alpha_q(L^p, (\Phi^{\mathcal{R}}_n)_{n \in \mathbb{N}}),$$

with $\alpha > 0$ and $0 < q \leq \infty$. We start by showing that $\Phi^{\mathcal{R}}_n$ satisfies (P1)-(P3) and $(P5)$ (under some assumptions), but not (P4).

Lemma 3.20 ($\Phi^{\mathcal{R}}_n$ satisfies (P1)-(P3) and $(P5)$)**.** *Let $0 < p \leq \infty$ and $S \subset L^p$ be a finite-dimensional space. Then $\Phi^{\mathcal{R}}_n$ satisfies (P1)-(P3). Moreover, if S contains the constant function one, $\Phi^{\mathcal{R}}_n$ satisfies (P5) for $0 < p < \infty$.*

[7]Compare to similar results obtained for RePU networks in [32, Section 3.4].

Proof. (P1)-(P3) are obvious. (P5) follows from the fact that

$$\bigcup_{n\in\mathbb{N}} \Phi_n^{\mathcal{R}} = \bigcup_{d\in\mathbb{N}} \bigcup_{r\in\mathbb{N}} \Phi_{b,d,S,r} = \bigcup_{d\in\mathbb{N}} V_{b,d,S} = V_{b,S},$$

and from Theorem 2.21. □

Lemma 3.21 ($\Phi_n^{\mathcal{R}}$ does not satisfy (P4)). *Let $0 < p \leq \infty$ and $S \subset L^p$ be a finite-dimensional subspace which is closed under b-adic dilation and such that $r(T_{b,d}(\varphi)) = 1$ for any $\varphi \in S$ and $d \in \mathbb{N}$. Then, $\Phi_n^{\mathcal{R}}$ satisfies*

(i) $\Phi_n^{\mathcal{R}} + \Phi_n^{\mathcal{R}} \subset \Phi_{3n^2}^{\mathcal{R}}$,

(ii) there exists no constant $c > 1$ such that $\Phi_n^{\mathcal{R}} + \Phi_n^{\mathcal{R}} \subset \Phi_{cn}^{\mathcal{R}}$.

Proof. See Appendix C. □

Lemma 3.22. *For any $n \in \mathbb{N}$, we have $\Phi_n^{\mathcal{R}} \subset \Phi_n^{\mathcal{S}}$.*

Proof. See Appendix C. □

Corollary 3.23. *For any $\alpha > 0$ and $0 < q \leq \infty$,*

$$R_q^\alpha(L^p) \subset S_q^\alpha(L^p).$$

4. Encoding Classical Approximation Tools

In this section, we demonstrate how classical approximation tools can be represented with tensor networks and bound the complexity of such a representation. Specifically, we consider representing fixed knot splines, free knot splines and polynomials.[8] This will be the basis for Section 6 where we use these complexity estimates to prove embeddings of a scale of interpolation spaces into the approximation classes $N_q^\alpha(L^p)$, $C_q^\alpha(L^p)$ and $S_q^\alpha(L^p)$. Our background space is as before L^p, where we specify the range of p where necessary.

4.1. Polynomials.

Let us first consider the encoding of a polynomial of degree $\bar{m}$ in $V_{b,d,\bar{m}}$.

Lemma 4.1 (Ranks for Polynomials). *Let $\varphi \in \mathbb{P}_{\bar{m}}$, $\bar{m} \in \mathbb{N}$. For any $d \in \mathbb{N}$, $\varphi \in V_{b,d,\bar{m}}$ and for $1 \leq \nu \leq d$,*

$$r_{\nu,d}(\varphi) \leq \min\{\bar{m}+1, b^\nu\}.$$

Proof. Since $\mathbb{P}_{\bar{m}}$ is closed under b-adic dilation, the result simply follows from Lemma 2.23 (i). □

Now we consider the representation of a function $\varphi \in \mathbb{P}_{\bar{m}}$ as a tensor in $\mathbf{V}_{b,d,m}$ with $m \neq \bar{m}$. An exact representation is possible if $\bar{m} \leq m$ (see Proposition 2.19). Otherwise we have to settle for an approximation. In this section, we consider a particular type of approximation based on local interpolations that we will use in the next section. Let $W^{m+1,p}$ denote the Sobolev space of $m+1$-times weakly-differentiable p-integrable functions.

Definition 4.2 (Local Interpolation). *We consider an interpolation operator $\mathcal{I}_m$ from $L^p([0,1))$ to $S := \mathbb{P}_m$, $1 \leq p \leq \infty$, such that for all $v \in W^{m+1,p}$ and all $l = 0, \ldots, m+1$, we have*

$$|v - \mathcal{I}_m v|_{W^{l,p}} \leq C\, |v|_{W^{m+1,p}} \tag{4.1}$$

for some constant $C > 0$ independent of v. For the construction of this operator and a proof of the above property see, e.g., [21, Theorem 1.103]. Then, we introduce the operator $\mathcal{I}_{b,d,m} := \mathcal{I}_{b,d,S}$ from $L^p([0,1))$ to $V_{b,d,m}$ defined by (2.12) with $\mathcal{I}_S = \mathcal{I}_m$.

[8] See [3] for a representation of wavelets.

Polynomials and interpolations thereof can be tensorized in a highly efficient manner – a fact first observed in [28].

Lemma 4.3 (Ranks of Interpolants of Polynomials). *For $\varphi \in \mathbb{P}_{\bar{m}}$, $\bar{m} \in \mathbb{N}$, the interpolant satisfies $\mathcal{I}_{b,d,m}(\varphi) \in V_{b,d,m}$ and for $1 \leq \nu \leq d$,*

$$r_{\nu,d}(\mathcal{I}_{b,d,m}(\varphi)) \leq \min\left\{b^{\nu},\ (m+1)b^{d-\nu},\ \bar{m}+1\right\}. \tag{4.2}$$

Proof. Since $\mathcal{I}_{b,d,m}(\varphi) \in V_{b,d,m}$, the bound $r_{\nu,d}(\varphi) \leq \min\left\{b^{\nu},\ (m+1)b^{d-\nu}\right\}$ is obtained from Lemma 2.23 (ii). Then from Lemma 2.26, we know that $r_{\nu,d}(\mathcal{I}_{b,d,m}(\varphi)) \leq r_{\nu,d}(\varphi)$ for all $1 \leq \nu \leq d$, and we conclude by using Lemma 4.1. □

From Lemma 4.3, we easily deduce

Proposition 4.4 (Complexity for Encoding Interpolants of Polynomials). *For a polynomial $\varphi \in \mathbb{P}_{\bar{m}}$, $\bar{m} \in \mathbb{N}$, the different complexities from Section 3 for encoding the interpolant $\mathcal{I}_{b,d,m}(\varphi)$ of level d and degree $m \leq \bar{m}$ within $V_{b,m}$ are bounded as*

$$\begin{aligned}
&\mathrm{compl}_{\mathcal{N}}(\mathcal{I}_{b,d,m}(\varphi)) \leq (\bar{m}+1)d,\\
&\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,d,m}(\varphi)) \leq \mathrm{compl}_{\mathcal{C}}(\mathcal{I}_{b,d,m}(\varphi)) \leq b(\bar{m}+1)^2 d + b(m+1).
\end{aligned}$$

4.2. **Fixed Knot Splines.** Let $b, d \in \mathbb{N}$. We divide $[0,1)$ into $N = b^d$ intervals $[x_k, x_{k+1})$ with

$$x_k := kb^{-d}, \quad k = 0, \ldots, b^d.$$

Fix a polynomial of degree $m \in \mathbb{N}_0$ and a continuity index $\mathfrak{c} \in \mathbb{N}_0 \cup \{-1, \infty\}$. Define the space of fixed knot splines of degree m with $N+1$ knots and $\mathfrak{c}$ continuous derivatives as

$$\mathcal{S}_{\mathfrak{c}}^{N,m} := \left\{f : [0,1) \to \mathbb{R}\ :\ f_{|[x_k,x_{k+1})} \in \mathbb{P}_m,\ k = 0, \ldots, N-1 \text{ and } f \in C^{\mathfrak{c}}([0,1))\right\},$$

where $C^{-1}([0,1))$ stands for not necessarily continuous functions on $[0,1)$, $C^0([0,1))$ is the space $C([0,1))$ of continuous functions on $[0,1)$ and $C^k([0,1))$, $k \in \mathbb{N} \cup \{\infty\}$, is the usual space of k-times differentiable functions. The following property is apparent.

Lemma 4.5 (Dimension of Spline Space). *$\mathcal{S}_{\mathfrak{c}}^{N,m}$ is a finite-dimensional subspace of L^p with*

$$\dim \mathcal{S}_{\mathfrak{c}}^{N,m} = \begin{cases} (m+1)N - (N-1)(\mathfrak{c}+1), & -1 \leq \mathfrak{c} \leq m,\\ m+1, & m+1 \leq \mathfrak{c} \leq \infty. \end{cases}$$

With the above Lemma we immediately obtain

Lemma 4.6 (Ranks of Fixed Knot Splines). *Let $\varphi \in \mathcal{S}_{\mathfrak{c}}^{N,m}$ with $N = b^d$. Then, $\varphi \in V_{b,d,m}$ and for $1 \leq \nu \leq d$*

$$r_{\nu,d}(\varphi) \leq \begin{cases} \min\left\{(m-\mathfrak{c})b^{d-\nu} + (\mathfrak{c}+1),\ b^{\nu}\right\}, & -1 \leq \mathfrak{c} \leq m,\\ \min\left\{m+1,\ b^{\nu}\right\}, & m+1 \leq \mathfrak{c} \leq \infty. \end{cases} \tag{4.3}$$

Proof. For any $0 \leq j < b^{\nu}$, the restriction of φ to the interval $[b^{-\nu}j, b^{-\nu}(j+1))$ is a piecewise polynomial in $C^{\mathfrak{c}}([b^{-\nu}j, b^{-\nu}(j+1)))$ with $b^{d-\nu}$ pieces, so that $\varphi(b^{-\nu}(j+\cdot)) \in \mathcal{S}_{\mathfrak{c}}^{b^{d-\nu},m}$ (with knots $kb^{-\nu}$, $0 \leq k < b^{\nu}$). Corollary 2.14 then implies $r_{\nu,d}(f) \leq \dim(\mathcal{S}_{\mathfrak{c}}^{b^{d-j},m})$ and we obtain (4.3) by using Lemma 4.5 and Lemma 2.23. □

Remark 4.7 (General Tensor Formats). *We could generalize the above statement to a general tree-based tensor format. In this case, for $\beta \subset \{1, \ldots, d\}$ we would have the bound (see also Lemma 2.22)*

$$r_{\beta,d}(\varphi) \leq \min\left\{(m+1)b^{d-\#\beta},\ b^{\#\beta}\right\}.$$

Note that $(T_{b,d}\varphi)(j_\beta,\cdot)$ is not necessarily a contiguous piece of φ, even if β is a contiguous subset of $\{1,\ldots,d\}$, e.g., $\beta=\{j,j+1,\ldots,j+i\}$. Therefore additional continuity constraints on $\varphi\in\mathcal{S}_{\mathfrak{c}}^{N,m}$ would in general not affect the rank bound. Of course, for large d the rank reduction due to continuity constraints is not essential, unless $\mathfrak{c}=m$ and in this case the ranks would be bounded by $m+1$ in any format, see also Remark 3.2.

Proposition 4.8 (Complexity for Encoding Fixed Knot Splines). *For a fixed knot spline $\varphi\in\mathcal{S}_{\mathfrak{c}}^{N,m}$ with $N=b^d$, the different complexities from Section 3 for encoding within $V_{b,m}$ are bounded as*

$$\mathrm{compl}_{\mathcal{N}}(\varphi)\le C\sqrt{N},$$
$$\mathrm{compl}_{\mathcal{S}}(\varphi)\le \mathrm{compl}_{\mathcal{C}}(\varphi)\le CN,$$

with constants $C>0$ depending only on b and m.

Proof. From Lemma 4.6, we obtain

$$\mathrm{compl}_{\mathcal{N}}(\varphi)=\sum_{\nu=1}^{\lfloor d/2\rfloor}b^\nu+\sum_{\nu=\lfloor d/2\rfloor+1}^{d}b^{d-\nu}\le 2\frac{b}{b-1}b^{d/2}=2\frac{b}{b-1}\sqrt{N},$$

$$\begin{aligned}\mathrm{compl}_{\mathcal{C}}(\varphi)&=\sum_{\nu=1}^{\lfloor d/2\rfloor}b^{2\nu}+\sum_{\nu=\lfloor d/2\rfloor+1}^{d}b^{2(d-\nu+1)}+b(m+1)\\&\le\frac{2b^2}{b^2-1}b^d+b(m+1)=\max\{\frac{2b^2}{b^2-1},(m+1)\}N.\end{aligned}$$

and we conclude by noting that $\mathrm{compl}_{\mathcal{S}}(\varphi)\le\mathrm{compl}_{\mathcal{C}}(\varphi)$. ☐

Now we would like to encode splines of degree $\bar m$ in $V_{b,\bar d,m}$ with $m\ne\bar m$ and $\bar d\ge d$. An exact representation is not possible for $\bar m>m$. Then, we again consider the local interpolation operator from Definition 4.2.

Lemma 4.9 (Ranks of Interpolants of Fixed Knot Splines). *Let $\varphi\in\mathcal{S}_{\mathfrak{c}}^{N,\bar m}$ with $N=b^d$. For $\bar d\ge d$, the interpolant $\mathcal{I}_{b,\bar d,m}(\varphi)\in V_{b,\bar d,m}$ satisfies*

$$r_{\nu,\bar d}(\mathcal{I}_{b,\bar d,m}(\varphi))\le\begin{cases}\min\left\{(\bar m-\mathfrak{c})b^{d-\nu}+(\mathfrak{c}+1),\ b^\nu\right\}, & -1\le\mathfrak{c}\le\bar m,\\ \min\left\{\bar m+1,\ b^\nu\right\}, & \bar m+1\le\mathfrak{c}\le\infty.\end{cases},\quad 1\le\nu\le d,$$

$$r_{\nu,\bar d}(\mathcal{I}_{b,\bar d,m}(\varphi))\le\min\left\{(m+1)b^{\bar d-\nu},\ \bar m+1\right\},\quad d<\nu\le\bar d.$$

Proof. From Lemma 2.26, we know that $r_{\nu,\bar d}(\mathcal{I}_{b,\bar d,m}(\varphi))\le r_{\nu,\bar d}(\varphi)$ for all $1\le\nu\le\bar d$. For $\nu\le d$, we have from Corollary 2.14 that $r_{\nu,\bar d}(\varphi)=r_{\nu,d}(\varphi)$. Then, we obtain the first inequality from Lemma 4.6. Now consider the case $d<\nu\le\bar d$. The bound $r_{\nu,\bar d}(\mathcal{I}_{b,\bar d,m}(\varphi))\le(m+1)b^{\bar d-\nu}$ simply follows from the fact that $\mathcal{I}_{b,\bar d,m}(\varphi)\in V_{b,\bar d,m}$. Since $\varphi\in V_{b,d,\bar m}$ and $\mathbb{P}_{\bar m}$ is closed under dilation, we obtain from Lemma 2.23 the other bound $r_{\nu,\bar d}(\varphi)\le\bar m+1$. ☐

Proposition 4.10 (Complexity for Encoding Interpolants of Fixed Knot Splines). *For a fixed knot spline $\varphi\in\mathcal{S}_{\mathfrak{c}}^{N,\bar m}$ with $N=b^d$, the different complexities from Section 3 for encoding the interpolant $\mathcal{I}_{b,\bar d,m}(\varphi)$ of level $\bar d\ge d$ and degree $m\le\bar m$ within $V_{b,m}$ are bounded as*

$$\mathrm{compl}_{\mathcal{N}}(\mathcal{I}_{b,\bar d,m}(\varphi))\le C\sqrt{N}+C'(\bar d-d),$$
$$\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar d,m}(\varphi))\le\mathrm{compl}_{\mathcal{C}}(\mathcal{I}_{b,\bar d,m}(\varphi))\le CN+C'(\bar d-d),$$

with constants $C,C'>0$ depending only on b, m and $\bar m$.

Proof. Using Lemma 2.26 and Lemma 4.6 and following the proof of Proposition 4.8, we have

$$\operatorname{compl}_{\mathcal{N}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq \sum_{\nu=1}^{d} r_\nu(\varphi) + (\bar{d}-d)(\bar{m}+1) \leq \frac{2b}{b-1}\sqrt{N} + (\bar{d}-d)(\bar{m}+1),$$

$$\begin{aligned}\operatorname{compl}_{\mathcal{C}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) &\leq br_1(\varphi) + \sum_{\nu=1}^{d} br_{\nu-1}(\varphi)r_\nu(\varphi) + (\bar{d}-d)b(\bar{m}+1)^2 + b(m+1)\\ &\leq \max\{\frac{2b^2}{b^2-1}, (m+1)\}N + (\bar{d}-d)b(\bar{m}+1)^2,\end{aligned}$$

and we conclude by noting that $\operatorname{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq \operatorname{compl}_{\mathcal{C}}(\mathcal{I}_{b,\bar{d},m}(\varphi))$. □

4.3. **Free Knot Splines.** A free knot spline is a piece-wise polynomial function, for which only the maximum polynomial order and the number of polynomial pieces is known – not the location of said pieces. More precisely, the set of free knot splines of degree $m \in \mathbb{N}_0$ with $N \in \mathbb{N}$ pieces is defined as

$$\mathcal{S}_{\mathrm{fr}}^{N,m} := \Big\{ f : [0,1) \to \mathbb{R} : \ \exists (x_k)_{k=0}^N \subset [0,1] \text{ s.t. } 0 = x_0 < x_1 < \ldots < x_N = 1 \text{ and } f_{|[x_k,x_{k+1})} \in \mathbb{P}_m \Big\}.$$

Clearly $\mathcal{S}_{\mathrm{fr}}^{N,m}$ is not a linear subspace like $\mathcal{S}_{\mathfrak{c}}^{N,m}$. Rank bounds for free knot splines are slightly more tricky than for fixed knot splines. We proceed in three steps:

(1) Assume first the knots x_k of the free knot spline are all located on a multiple of b^{-d_k} for some $d_k \in \mathbb{N}$, i.e., only b-adic knots are allowed. Assume also the largest d_k is known.
(2) Show that restricting to b-adic knots does not affect the approximation class as compared to non-constrained free knot splines.
(3) Show that the largest d_k can be estimated using the desired approximation accuracy and excess regularity/integrability of the target function.

In this section, we only address point (1). In Section 5.2, we will address (2) and (3).

Definition 4.11 (Free b-adic Knot Splines). *We call a sequence of points $(x_k^b)_{k=0}^N \subset [0,1]$* b-adic *if*

$$x_k^b = i_k b^{-d_k},$$

for some $d_k \in \mathbb{N}$ and $0 \leq i_k \leq b^{d_k}$. We use the superscript b to indicate that a sequence is b-adic. With this we define the set of free b-adic knot splines *as*

$$\mathcal{S}_{\mathrm{fr}}^{b,N,m} := \Big\{ f : [0,1) \to \mathbb{R} : \ \exists (x_k^b)_{k=0}^N \subset [0,1] \textit{ s.t. } 0 = x_0^b < x_1^b < \ldots < x_N^b = 1 \textit{ and } f_{|[x_k^b,x_{k+1}^b)} \in \mathbb{P}_m \Big\}.$$

Lemma 4.12 (Ranks of Free b-adic Knot Splines). *Let $\varphi \in \mathcal{S}_{\mathrm{fr}}^{b,N,m}$ with $(x_k^b)_{k=0}^N$ being the b-adic knot sequence corresponding to φ. Let $d := \max\{d_k : \ 1 \leq k \leq N-1\}$. Then, $\varphi \in V_{b,d,m}$ and*

$$r_{\nu,d}(\varphi) \leq \min\{b^\nu,\, (m+1)b^{d-\nu},\, m+N\} \tag{4.4}$$

for $1 \leq \nu \leq d$.

Proof. For any $0 \leq j < b^\nu$, the restriction of φ to the interval $[b^{-\nu}j, b^{-\nu}(j+1))$ is either a polynomial or a piece-wise polynomial where the number of such piece-wise polynomials is at most $N-1$, since there are at most $N-1$ discontinuities in $(0,1)$. Hence, Corollary 2.14 implies that $r_\nu(\varphi) \leq m+N$ for all $1 \leq \nu \leq d$, and we obtain the other bound $r_{\nu,d}(\varphi) \leq \min\left\{b^\nu,\, (m+1)b^{d-\nu}\right\}$ from Lemma 2.23 with $\dim(S) = m+1$. □

Proposition 4.13 (Complexity for Encoding Free b-adic Knot Splines)**.** *For a free knot spline $\varphi \in \mathcal{S}_{\mathrm{fr}}^{b,N,m}$ with $d := \max\{d_k : 1 \leq k \leq N-1\}$, the different complexities from Section 3 for encoding within $V_{b,m}$ are bounded as*

$$\mathrm{compl}_{\mathcal{N}}(\varphi) \leq CdN,$$
$$\mathrm{compl}_{\mathcal{C}}(\varphi) \leq CdN^2,$$
$$\mathrm{compl}_{\mathcal{S}}(\varphi) \leq Cd^2N,$$

with constants $C > 0$ depending only on b and m.

Proof. Follows from Lemma 4.12, cf. also Proposition 4.8. See Appendix D for a detailed proof. □

Lemma 4.14 (Ranks of Interpolants of Free b-adic Knot Splines)**.** *Let $\varphi \in \mathcal{S}_{\mathrm{fr}}^{b,N,\bar{m}}$, $\bar{m} \geq m$, and $(x_k^b)_{k=0}^N$ being the b-adic knot sequence corresponding to φ. Let $d :=$ $\max\{d_k : 1 \leq k \leq N-1\}$. For $\bar{d} \geq d$, the interpolant $\mathcal{I}_{b,\bar{d},m}(\varphi) \in V_{b,\bar{d},m}$ satisfies*

$$r_{\nu,\bar{d}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq \min\left\{b^\nu,\, (\bar{m}+1)b^{d-\nu},\, \bar{m}+N\right\}, \quad 1 \leq \nu \leq d,$$
$$r_{\nu,\bar{d}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq \min\left\{(m+1)b^{\bar{d}-\nu},\, \bar{m}+1\right\}, \quad d < \nu \leq \bar{d}.$$

Proof. From Lemma 2.26, we know that $r_{\nu,\bar{d}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq r_{\nu,\bar{d}}(\varphi)$ for all $1 \leq \nu \leq \bar{d}$. For $\nu \leq d$, we have from Corollary 2.14 that $r_{\nu,\bar{d}}(\varphi) = r_{\nu,d}(\varphi)$. Then, we obtain the first inequality from Lemma 4.12. Now consider the case $d < \nu \leq \bar{d}$. The bound $r_{\nu,\bar{d}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq (m+1)b^{\bar{d}-\nu}$ simply follows from the fact that $\mathcal{I}_{b,\bar{d},m}(\varphi) \in V_{b,\bar{d},m}$. Since $\varphi \in V_{b,d,\bar{m}}$ and $\mathbb{P}_{\bar{m}}$ is closed under dilation, we obtain from Lemma 2.23 the other bound $r_{\nu,\bar{d}}(\varphi) \leq \bar{m}+1$. □

Proposition 4.15 (Complexity for Encoding Interpolants of Free b-adic Knot Splines)**.** *For a free knot spline $\varphi \in \mathcal{S}_{\mathrm{fr}}^{b,N,\bar{m}}$ with $d := \max\{d_k : 1 \leq k \leq N-1\}$, the different complexities from Section 3 for encoding the interpolant $\mathcal{I}_{b,\bar{d},m}(\varphi)$ of level $\bar{d} \geq d$ and degree $m \leq \bar{m}$ are bounded as*

$$\mathrm{compl}_{\mathcal{N}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq CdN + C'(\bar{d}-d),$$
$$\mathrm{compl}_{\mathcal{C}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq CdN^2 + C'(\bar{d}-d),$$
$$\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq Cd^2N + C'(\bar{d}-d),$$

with constants $C, C' > 0$ depending only on b, m and $\bar{m}$.

Proof. Follows from Lemma 4.14, cf. also Proposition 4.10. See Appendix D for a detailed proof. □

Remark 4.16. *Both in Lemma 4.12 and Lemma 4.14, the rank bound is of the order N and does not assume any specific structure of the spline approximation. This is a crude estimate that could be perhaps improved if one imposes additional restrictions, such as a tree-like support structure of the approximating splines.*

5. Approximation Rates

In this section, we discuss approximation rates for functions with different types of smoothness.

5.1. **Sobolev Spaces.** We consider the approximation properties of tensor networks in $V_{b,m}$ for a fixed $m \in \mathbb{N}_0$. Recall the definition of three different complexity measures $\mathrm{compl}_{\mathcal{N}}$, $\mathrm{compl}_{\mathcal{C}}$ and $\mathrm{compl}_{\mathcal{S}}$ from Section 3 and the resulting approximating sets $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$. The best approximation error for $0 < p \leq \infty$ is defined accordingly as

$$
\begin{aligned}
E_n^{\mathcal{N}}(f)_p &:= \inf_{\varphi \in \Phi_n^{\mathcal{N}}} \|f - \varphi\|_p\,, \\
E_n^{\mathcal{C}}(f)_p &:= \inf_{\varphi \in \Phi_n^{\mathcal{C}}} \|f - \varphi\|_p\,, \\
E_n^{\mathcal{S}}(f)_p &:= \inf_{\varphi \in \Phi_n^{\mathcal{S}}} \|f - \varphi\|_p\,,
\end{aligned}
\tag{5.1}
$$

and the corresponding approximation classes N_q^α, C_q^α and S_q^α as in Section 3.4.

We will apply local interpolation from Definition 4.2 to approximate functions in Sobolev spaces $W^{k,p}$ for any $k \in \mathbb{N}$. These embeddings essentially correspond to embeddings of Besov spaces $B_{p,p}^\alpha$ into approximation spaces $N_\infty^\alpha(L^p)$, $C_\infty^\alpha(L^p)$ and $S_\infty^\alpha(L^p)$: i.e., the approximation error is measured in the same norm as smoothness[9]. To this end, we require

Lemma 5.1 (Re-Interpolation)**.** *Let $f \in W^{\bar{m}+1,p}$, $1 \leq p \leq \infty$ and $\bar{m} \geq m$. For any $d \in \mathbb{N}_0$, $\mathcal{I}_{b,d,\bar{m}} f$ is a fixed knot spline in $\mathcal{S}_{-1}^{N,\bar{m}}$, $N = b^d$, and*

$$
\|f - \mathcal{I}_{b,d,\bar{m}} f\|_p \leq C b^{-d(\bar{m}+1)} |f|_{W^{\bar{m}+1,p}}
$$

where C is a constant depending only on $\bar{m}$ and p. Furthermore, for $\bar{d} \geq d$,

$$
\|\mathcal{I}_{b,d,\bar{m}} f - \mathcal{I}_{b,\bar{d},m} \mathcal{I}_{b,d,\bar{m}} f\|_p \leq C' \left(b^{-\bar{d}(m+1)} |f|_{W^{m+1,p}} + b^{-(\bar{d}-d)(m+1) - d(\bar{m}+1)} |f|_{W^{\bar{m}+1,p}} \right)
$$

where C' is a constant depending only on $\bar{m}$, m and p.

Proof. See Appendix E. □

With this we can show the direct estimate

Theorem 5.2 (Jackson Inequality for Sobolev Spaces)**.** *Let $1 \leq p \leq \infty$ and $k \in \mathbb{N}$. For any $f \in W^{k,p}$ we have*

$$
\begin{aligned}
E_n^{\mathcal{N}}(f)_p &\leq C n^{-2k} \|f\|_{W^{k,p}}\,, \\
E_n^{\mathcal{S}}(f)_p \leq E_n^{\mathcal{C}}(f)_p &\leq C n^{-k} \|f\|_{W^{k,p}}\,,
\end{aligned}
\tag{5.2}
$$

with constants C depending on k, m, b.

Proof. Let $N := b^d$ and $k := \bar{m} + 1 > m + 1$ and fix some $f \in W^{k,p}$. The case $k \leq m + 1$ can be handled similarly with fewer steps. Let $s := \mathcal{I}_{b,d,\bar{m}} f$ and $\tilde{s} := \mathcal{I}_{b,\bar{d},m} s \in V_{b,\bar{d},m}$ with a $\bar{d} \geq d$ to be specified later.

From Lemma 5.1 we have

$$
\|f - \tilde{s}\|_p \leq C_1 \|f\|_{W^{k,p}} \left(b^{-dk} + b^{-\bar{d}(m+1)} \right)
\tag{5.3}
$$

for a constant C_1 depending only on r, m and p. Thus, we set

$$
\bar{d} := \left\lceil \frac{dk}{m+1} \right\rceil,
\tag{5.4}
$$

which yields

$$
\|f - \tilde{s}\|_p \leq 2C_1 \|f\|_{W^{r,p}} N^{-k}.
\tag{5.5}
$$

[9]Compare to the embeddings for RePU networks in [32].

From Proposition 4.10 and (5.4), we can estimate the complexity of $\tilde{s}$ as

$$n := \operatorname{compl}_{\mathcal{N}}(\tilde{s}) \leq \bar{C}_2(\sqrt{N} + \log_b(N)) \leq C_2\sqrt{N},$$
$$n := \operatorname{compl}_{\mathcal{S}}(\tilde{s}) \leq \operatorname{compl}_{\mathcal{C}}(\tilde{s}) \leq \bar{C}_2(N + \log_b(N)) \leq C_2 N$$

with a constant C_2 depending on b, m and $\bar{m}$. Thus, inserting into (5.5), we obtain (5.2). For the case $\bar{m} \leq m$, the proof simplifies since we can represent s exactly and use Proposition 4.8. □

Remark 5.3. *One could extend the statement of Theorem 5.2 to the range $0 < p < 1$ by considering the Besov spaces $B^{\alpha}_{p,p}$. For $r \leq m + 1$, this can be done by using the characterization of Besov spaces $B^{\alpha}_{p,p}$ for $0 < p \leq \infty$ by dyadic splines from* [20], *as was done in* [32, Theorem 5.5] *for RePU networks. For $k > m + 1$, one would have to additionally replace the interpolation operator of Definition 4.2 with the quasi-interpolation operator from* [20].

5.2. **Besov Spaces.** The key to proving direct estimates for Besov smoothness are the estimates of Proposition 4.13 and Proposition 4.15 for free knot splines. However, there are two issues with encoding free knot splines as tensorized polynomials. First, free knot splines are not restricted to b-adic knots and thus cannot be represented exactly within $V_{b,m}$. Second, even if all knots of a spline s are b-adic, the complexity of encoding s as an element of $V_{b,m}$ depends on the minimal level $d \in \mathbb{N}$ such that $s \in V_{b,d,m}$, and this level is not known in general. We address these issues with the following two lemmas.

Lemma 5.4 (b-adic Free Knot Splines)**.** *Let $0 < p < \infty$, $0 < \alpha < \bar{m} + 1$ and let $\mathcal{S}^{b,N,\bar{m}}_{\mathrm{fr}}$ denote the set of free knot splines of order $\bar{m} + 1$ with $N + 1$ knots restricted to b-adic points of the form*

$$x_k := i_k b^{-d_k}, \quad 0 \leq k \leq N,$$

for some $d_k \in \mathbb{N}$ and $i_k \in \{0, \ldots, d_k\}$. For $\tau := (\alpha + 1/p)^{-1}$ being the Sobolev embedding number and $f \in B^{\alpha}_{\tau,\tau}$, we have

$$\inf_{s \in \mathcal{S}^{b,N,\bar{m}}_{\mathrm{fr}}} \|f - s\|_p \leq C N^{-\alpha} |f|_{B^{\alpha}_{\tau,\tau}}. \tag{5.6}$$

Proof. See Appendix E. □

Remark 5.5. *In principle, Lemma 5.4 can be extended to the case $p = \infty$, $f \in C^0$ and the Besov space $B^{\alpha}_{\tau,\tau}$ replaced by the space of functions of bounded variation. However, the following Lemma 5.6 does not hold for $p = \infty$, such that overall we can show the direct estimate of Theorem 5.8 only for $p < \infty$.*

Lemma 5.6 (Smallest Interval Free Knot Splines)**.** *Let $\delta > 1$, $1 \leq p < \infty$ and $f \in L^{p\delta}$. Let $q = q(\delta) > 1$ be the conjugate of δ defined by*

$$\frac{1}{\delta} + \frac{1}{q} = 1.$$

For $\varepsilon > 0$, let $s = \sum_{k=1}^{N} s_k$ be a piece-wise polynomial such that

$$\|f - s\|_p \leq \varepsilon,$$

where we assume s_k is a polynomial over some interval I_k, zero otherwise and I_k, $k = 1, \ldots, N$, form a partition of $[0, 1]$.

Then, we can choose an index set $\Lambda = \Lambda(\varepsilon) \subset \{1, \ldots, N\}$ and a corresponding spline $\tilde{s} = \sum_{k \in \Lambda} \tilde{s}_k$ such that

$$\|f - \tilde{s}\|_p \leq 2^{1/p}\varepsilon \quad \text{with} \quad |I_k| > N^{-q}\|f\|^{-pq}_{p\delta}\varepsilon^{pq} =: \varrho(\varepsilon), \quad k \in \Lambda. \tag{5.7}$$

Proof. See Appendix E. □

Remark 5.7. *We can guarantee $f \in L^{p\delta}$ by assuming excess regularity and using Sobolev embeddings as follows. Let $\alpha > 0$, $0 < p < \infty$, $\delta > 1$ and $\tau := (\alpha + 1/p)^{-1}$. Defining $\alpha_\delta > \alpha$ as*

$$\alpha_\delta := \alpha + \frac{\delta - 1}{p\delta},$$

we get that the Sobolev embedding number for the combination α_δ, $p\delta$ is

$$\tau_\delta := (\alpha_\delta + 1/(p\delta))^{-1} = (\alpha + 1/p)^{-1} = \tau.$$

Then, assuming $f \in B^{\alpha_\delta}_{\tau,\tau}$ implies that $f \in L^{p\delta}$.

Theorem 5.8 (Jackson Inequality for $B^{\alpha}_{\tau,\tau}$)**.** *Let $1 \le p < \infty$, $0 < \tau < p$, $\alpha > 1/\tau - 1/p$, and assume $f \in B^{\alpha}_{\tau,\tau}$. Then, for any $\sigma > 0$, we obtain the direct estimates*

$$\begin{aligned} E_n^{\mathcal{N}}(f)_p &\le C\,|f|_{B^{\alpha}_{\tau,\tau}}\, n^{-\frac{\alpha}{1+\sigma}}, \\ E_n^{\mathcal{C}}(f)_p &\le C\,|f|_{B^{\alpha}_{\tau,\tau}}\, n^{-\frac{\alpha}{2+\sigma}}, \\ E_n^{\mathcal{S}}(f)_p &\le C\,|f|_{B^{\alpha}_{\tau,\tau}}\, n^{-\frac{\alpha}{1+\sigma}}, \end{aligned} \tag{5.8}$$

where the constants C depend on $\alpha > 0$, $\sigma > 0$, b and m. In particular, they diverge to infinity as $\sigma \to 0$ or $\alpha \to 1/\tau - 1/p$.

Proof. As in Theorem 5.2, we consider only the case $m + 1 \le \alpha$, as the case $\alpha < m + 1$ can be handled analogously with fewer steps. By Lemma 5.4, we can restrict ourselves to free knot splines with b-adic knots. By Lemma 5.6, we can bound the size of the smallest interval and thus the level d. And finally, by Lemma 4.14, we can bound the ranks of an interpolation of a free knot spline. Thus, we have all the ingredients to bound the representation complexity of a free knot spline. It remains to combine these estimates with standard results from approximation theory to arrive at (5.8).

Let $N \in \mathbb{N}$ be arbitrary. From Lemma 5.4, we know there exists a spline $s \in \mathcal{S}^{b,N,\bar{m}}_{\mathrm{fr}}$ with b-adic knots such that

$$\|f - s\|_p \le C_1 N^{-\alpha}\, \|f\|_{B^{\alpha}_{\tau,\tau}}, \tag{5.9}$$

for some constant $C_1 > 0$. Set $\varepsilon := C_1\, \|f\|_{B^{\alpha}_{\tau,\tau}} N^{-\alpha}$. Since $\alpha > 1/\tau - 1/p$, there exists a $\delta > 1$ such that $f \in L^{p\delta}$. By Lemma 5.6, we can assume w.l.o.g. that $d := d(s)$ is such that

$$b^{-d} > N^{-q}\, \|f\|_{p\delta}^{pq}\, \varepsilon^{pq},$$

or equivalently

$$d < q \log_b(\varepsilon^{-p}\, \|f\|_{p\delta}^{p}\, N) = q \log_b\left[C_1^{-p}\, \|f\|^{-p}_{B^{\alpha}_{\tau,\tau}}\, \|f\|^{p}_{p\delta}\, N^{1+\alpha p}\right] \le q \log_b\left[C_1^{-p} N^{1+\alpha p}\right],$$

where $q = \delta/(\delta - 1)$.

We use the interpolant of Definition 4.2 and set $\tilde{s} := \mathcal{I}_{b,\bar{d},m} s$ for $\bar{d} \ge d$ to be specified later. Let $s_j := \boldsymbol{s}(j_1, \ldots, j_d, \cdot)$, where $\boldsymbol{s} = T_{b,d} s$, and analogously $\tilde{s}_j$. For the re-interpolation error we can estimate similar to Lemma 5.1

$$\begin{aligned} \|s - \tilde{s}\|_p^p &= \sum_{j \in I_b^d} b^{-d}\, \|s_j - \tilde{s}_j\|_p^p \le C_2 \sum_{j \in I_b^d} b^{-d} b^{-p(\bar{d}-d)(m+1)} \left\| s_j^{(m+1)} \right\|_p^p \\ &\le C_3 \sum_{j \in I_b^d} b^{-d} b^{-(\bar{d}-d)(m+1)p}\, \|s_j\|_p^p, \end{aligned}$$

where the latter follows from [19, Theorem 2.7 of Chapter 4], since s_j is a polynomial of degree $\bar{m}$.

Since s is a quasi-interpolant of f, s_j is a dilation of a polynomial (near-)best approximation of f over the corresponding interval and thus by [19, Theorem 8.1 of Chapter 12]

$$\|s_j\|_p \leq C_4 \, |f_j|_{B^\alpha_{\tau,\tau}}$$

where $f_j := \boldsymbol{f}(j_1, \ldots, j_d, \cdot)$ and for any $j \in I_b^d$.

Since $\tau < p$, we can further estimate

$$\begin{aligned}
\left(\sum_{j \in I_b^d} b^{-d} b^{-(\bar{d}-d)(m+1)p} \|s_j\|_p^p \right)^{1/p} &\leq b^{-d/p} b^{-(\bar{d}-d)(m+1)} \left(\sum_{j \in I_b^d} |f_j|^p_{B^\alpha_{\tau,\tau}} \right)^{1/p} \\
&\leq b^{-d/p} b^{-(\bar{d}-d)(m+1)} \left(\sum_{j \in I_b^d} |f_j|^\tau_{B^\alpha_{\tau,\tau}} \right)^{1/\tau} \\
&= b^{-d/p} b^{-(\bar{d}-d)(m+1)} \left(\sum_{j \in I_b^d} |f_j|^\tau_{B^\alpha_{\tau,\tau}} \right)^{1/\tau} b^{d(\alpha-1/\tau)} b^{d(1/\tau-\alpha)}.
\end{aligned}$$

We finally estimate

$$\|s - \tilde{s}\|_p \leq C_5 b^{-(\bar{d}-d)(m+1)} b^{d(1/\tau-\alpha-1/p)} \, |f|_{B^\alpha_{\tau,\tau}} .$$

More details on the relationship between Sobolev and Besov norms for functions and their tensorizations can be found in Appendix A.

Thus, to obtain at least the same approximation order as in (5.9), we set

$$\bar{d} := \left\lceil \frac{d(m+1+1/\tau-\alpha-1/p) + \alpha \log_b(N)}{m+1} \right\rceil \leq C_6 \log_b(N),$$

so that

$$\|s - \tilde{s}\|_p \leq C_5 N^{-\alpha} \, |f|_{B^\alpha_{\tau,\tau}} . \tag{5.10}$$

From Proposition 4.15, $\tilde{s} \in V_{b,d,m}$ with

$$n := \mathrm{compl}_{\mathcal{C}}(\tilde{s}) \leq C_7 \left(N^2 \log_b(N) + \log_b(N) \right) \leq C N^{2+\sigma},$$

for any $\sigma > 0$, where $C > 0$ depends on σ. Similarly for $\mathrm{compl}_{\mathcal{S}}$ and $\mathrm{compl}_{\mathcal{N}}$, we obtain from Proposition 4.15 that

$$\mathrm{compl}_{\mathcal{S}}(\tilde{s}) \leq \bar{C}(N \log_b(N)^2 + \log_b(N)) \leq C N^{1+\sigma},$$

and

$$\mathrm{compl}_{\mathcal{N}}(\tilde{s}) \leq \bar{C}(N \log_b(N) + \log_b(N)) \leq C N^{1+\sigma},$$

for any $\sigma > 0$ and constants C depending on $\sigma > 0$. Combining (5.9) with (5.10), a triangle inequality and the above complexity bounds, we obtain the desired statement. □

Remark 5.9 (Dense vs. Sparse TT). *Note that for approximation of Sobolev (and later analytic functions), the approximation rates for dense TTs from $\Phi_n^{\mathcal{C}}$ and sparse TTs from $\Phi_n^{\mathcal{S}}$ are the same, while for functions with Besov smoothness $B^s_{\tau,\tau}$ optimal rates are only achieved with sparse TTs, and dense TTs can only achieve half the optimal rate. This distinction can be roughly understood as follows.*

Functions with Sobolev smoothness can be optimally approximated with linear approximation tools that, simply put, capture all polynomial features of a function upto some refinement level d – independent of location. In this case, a TT approximation on level d, in general, represents a spline that is active/nonzero over all $n = b^d$ subintervals of the corresponding tensorization. A sparse TT representation of such a function has at most a multiple of n terms. On the other hand, the ranks saturate at $r \lesssim b^{d/2} = \sqrt{n}$ and thus a dense TT representation also has at most a multiple of n terms.

In contrast to functions with Sobolev smoothness, $B^s_{\tau,\tau}$ requires nonlinear approximation tools that capture possibly location-dependent features of the target function to achieve optimal approximation rates. In this case, a TT approximation on level d, in general, represents a spline with much fewer active intervals than the maximal possible $b^d \gg n$. Thus, while a sparse TT representation has at most a multiple of n coefficients – where the sparsity pattern encodes the location of said active subintervals – a dense TT representation has ranks bounded as $r \lesssim n^2$ and thus at most a multiple of n^2 coefficients, resulting in half the optimal approximation rate.

5.3. **Analytic Functions.** It is well known that analytic functions can be approximated by algebraic polynomials with a rate exponential in the degree of the approximating polynomials: see, e.g., [19, Chapter 7, Theorem 8.1]. In our setting, the polynomial degree in $V_{b,m}$ is fixed. However, as before we can re-interpolate and consider the corresponding approximation rate. First, we show that polynomials can be approximated with an exponential rate.

Lemma 5.10 (Approximation Rate for Polynomials)**.** *Let $P \in \mathbb{P}_{\bar{m}}$ be an arbitrary polynomial with $\bar{m} > m$ (otherwise we have exact representation). Then, for $1 \leq p \leq \infty$*

$$E^{\mathcal{N}}_n(P)_p \leq Cb^{-\frac{m+1}{(\bar{m}+1)}n} \left\|P^{(m+1)}\right\|_p,$$
$$E^{\mathcal{S}}_n(P)_p \leq E^{\mathcal{C}}_n(P)_p \leq Cb^{-\frac{m+1}{b(\bar{m}+1)^2}n} \left\|P^{(m+1)}\right\|_p,$$

with C independent of $\bar{m}$.

Proof. See Appendix E. □

This implies analytic functions can be approximated with an error decay of exponential type. For the following statement we require the distance function

$$\operatorname{dist}(z, D) := \inf_{w \in D} |z - w|, \quad z \in \mathbb{C}, \quad D \subset \mathbb{C}.$$

Theorem 5.11 (Approximation Rate for Analytic Functions)**.** *Let $\rho > 1$ and define*

$$D_\rho := \left\{ z \in \mathbb{C} : \operatorname{dist}(z, [0,1]) < \frac{\rho - 1}{2} \right\}.$$

Let $\rho := \rho(f) > 1$ be such that $f : [0,1) \to \mathbb{R}$ has an analytic extension onto $D_\rho \subset \mathbb{C}$. Then,

$$E^{\mathcal{N}}_n(f)_\infty \leq C[\min(\rho, b^{(m+1)})]^{-n^{1/2}}, \tag{5.11}$$
$$E^{\mathcal{S}}_n(f)_\infty \leq E^{\mathcal{C}}_n(f)_\infty \leq C[\min(\rho, b^{(m+1)/b})]^{-n^{1/3}},$$

where $C = C(f, m, b, \rho)$.

Proof. See Appendix E. □

Remark 5.12. *The above estimate can be further refined in the following ways:*

- *The factor in the base of the exponent can be replaced by any number* θ
$$\min(\rho, b^{(m+1)/b}) < \theta < \max(\rho, b^{(m+1)/b}),$$
with an adjusted constant C*.*
- *The inequality* (5.11) *can be stated in the form as in* [19, Chapter 7, Theorem 8.1] *to explicitly include the case* $\rho = \infty$*.*
- *One can define classes of entire functions as in* [19, Chapter 7, Theorem 8.3] *for a finer distinction of functions that can be approximated with an exponential-type rate.*
- *One can extend the result to approximation of analytic functions with singularities applying similar ideas as in* [41]*.*

6. Direct Embeddings

In this section, we discuss direct embeddings for the approximation spaces defined in Section 3.4. Since we verified that N_q^α, C_q^α and S_q^α satisfy (P1) – (P4), we can use classical approximation theory (see [19, 17]) to show that an entire scale of interpolation and smoothness spaces is continuously embedded into these approximation classes. We begin by briefly reviewing interpolation spaces. See Appendix A.2 for a definition of Besov spaces.

6.1. **Interpolation Spaces.** We consider Peetre's K-functional real interpolation method. Let X, Y be Banach spaces with $Y \hookrightarrow X$. The K-functional on X is defined as
$$K(f,t,X,Y) := K(f,t) := \inf_{g \in Y} \{\|f-g\|_X + t\,\|g\|_Y\}, \quad t > 0.$$

Definition 6.1 (Interpolation Spaces, [9, Chapter 5]). *Define a (quasi-)norm on* X
$$\|f\|_{\theta,q} := \begin{cases} \int_0^\infty \left[t^{-\theta}K(f,t)^q \frac{\mathrm{d}t}{t}\right]^{1/q}, & 0<\theta<1,\ 0<q<\infty, \\ \sup_{t>0} t^{-\theta}K(f,t), & 0 \le \theta \le 1, \quad q = \infty. \end{cases}$$
The interpolation space $(X,Y)_{\theta,q}$ *is defined as*
$$(X,Y)_{\theta,q} := \left\{ f \in X : \ \|f\|_{\theta,q} < \infty \right\},$$
and it is a complete (quasi-)normed space.

Some basic properties of these interpolation spaces are:
- $Y \hookrightarrow (X,Y)_{\theta,q} \hookrightarrow X$;
- $(X,Y)_{\theta_1,q} \hookrightarrow (X,Y)_{\theta_2,q}$ for $\theta_1 \ge \theta_2$ and $(X,Y)_{\theta,q_1} \hookrightarrow (X,Y)_{\theta,q_2}$ for $q_1 \le q_2$;
- *re-iteration* property: let $X' := (X,Y)_{\theta_1,q_1}$, $Y' := (X,Y)_{\theta_2,q_2}$. Then, for all $0<\theta<1$ and $0<q\le\infty$, we have
$$(X',Y')_{\theta,q} = (X,Y)_{\alpha,q}, \quad \alpha := (1-\theta)\theta_1 + \theta\theta_2.$$

We cite some important results on the relationship between interpolation, approximation and smoothness spaces. To this end, an important tool are the so-called *Jackson (direct)* and *Bernstein (inverse)* inequalities from (3.3) and (3.4), respectively.

Theorem 6.2 (Interpolation and Approximation, [19, Chapter 7], [9, Chapter 5]). *If the approximation class* $A_q^\alpha(X)$ *satisfies (P1), (P3), (P4) and the space* Y *satisfies the Jackson inequality* (3.3)*, then*
$$(X,Y)_{\alpha/k_\mathrm{J},q} \hookrightarrow A_q^\alpha(X), \quad 0<\alpha<k_\mathrm{J},\ 0<q<\infty,$$
$$Y \hookrightarrow A_\infty^{k_\mathrm{J}}(X).$$

If the approximation class $A^\alpha_q(X)$ *satisfies (P1) – (P6) and the space* Y *satisfies the Bernstein inequality* (3.4)*, then*

$$\begin{aligned} A^\alpha_q(X) &\hookrightarrow (X,Y)_{\alpha/k_{\mathrm{B}},q}, \quad 0<\alpha<k_{\mathrm{B}},\\ A^{k_{\mathrm{B}}}_\infty(X) &\hookrightarrow Y. \end{aligned}$$

Theorem 6.3 (Interpolation and Smoothness, [17]). *The following identities hold:*

$$\begin{aligned} (L^p, W^{\alpha,p})_{\theta,q} &= B^{\theta\alpha}_{p,q}, \quad 0<\theta<1,\ 0<q\le\infty,\ 1\le p\le\infty\\ (B^{\alpha_1}_{p,q_1}, B^{\alpha_2}_{p,q_2})_{\theta,q} &= B^{\alpha}_{p,q}, \quad \alpha:=(1-\theta)\alpha_1+\theta\alpha_2,\ 0<p,q,q_1,q_2\le\infty\\ (L^p, B^{\alpha}_{p,\tilde q})_{\theta,q} &= B^{\theta\alpha}_{p,q}, \quad 0<\theta<1,\ 0<p,q,\tilde q\le\infty,\\ (L^p, B^{\alpha}_{\tau,\tau})_{\theta,q} &= B^{\theta\alpha}_{q,q}, \quad 0<\tau<p,\ \frac{1}{q}=\theta\alpha+\frac{1}{p}. \end{aligned}$$

6.2. **Embeddings.** Theorems 6.2 and 6.3 allow us to characterize the approximation classes introduced in Section 3.4 by classical smoothness and interpolation spaces, provided we can show for $X = L^p$ and $Y = B^\alpha_{p,q}$ the Jackson (3.3) and Bernstein (3.4) inequalities. The Jackson inequalities were shown in Section 5. We will also show later that Bernstein inequalities cannot hold. This is an expression of the fact that the spaces A^α_q are "too large" in the sense that they are not continuously embedded in any classical smoothness space.

Theorem 6.2 and Theorem 5.2 imply

Theorem 6.4 (Direct Embedding for Sobolev Spaces). *For any* $k\in\mathbb{N}$ *and* $1\le p\le\infty$*, we have*

$$W^{k,p}\hookrightarrow N^{2k}_\infty(L^p), \quad W^{k,p}\hookrightarrow C^k_\infty(L^p)\hookrightarrow S^k_\infty(L^p),$$

and for $0<q\le\infty$

$$\begin{aligned} (L^p, W^{k,p})_{\alpha/2k,q} &\hookrightarrow N^\alpha_q(L^p), && 0<\alpha<2k,\\ (L^p, W^{k,p})_{\alpha/k,q} &\hookrightarrow C^\alpha_q(L^p)\hookrightarrow S^\alpha_q(L^p), && 0<\alpha<k. \end{aligned}$$

Corollary 6.5. *Together with Theorem 6.3, this implies the statement of Result 1.7.*

Now we turn to direct embeddings of Besov spaces $B^\alpha_{\tau,\tau}$ into $N^\alpha_q(L^p)$, $C^\alpha_q(L^p)$ and $S^\alpha_q(L^p)$, where $1/\tau = \alpha + 1/p$. That is, the smoothness is measured in a weaker norm with $\tau < p$. The spaces $B^\alpha_{\tau,\tau}$ are in this sense much larger than $B^\alpha_{p,p}$.

Theorem 6.6 (Direct Embedding for $B^\alpha_{\tau,\tau}$). *Let* $1\le p<\infty$*,* $0<\tau<p$ *and* $\gamma>1/\tau-1/p$*. Then, for any* $\sigma>0$*,*

$$B^\gamma_{\tau,\tau}\hookrightarrow N^{\gamma/(1+\sigma)}_\infty(L^p), \quad B^\gamma_{\tau,\tau}\hookrightarrow C^{\gamma/(2+\sigma)}_\infty(L^p), \quad B^\gamma_{\tau,\tau}\hookrightarrow S^{\gamma/(1+\sigma)}_\infty(L^p),$$

and

$$\begin{aligned} (L^p, B^r_{\tau,\tau})_{\alpha(1+\sigma)/r,q} &\hookrightarrow N^\alpha_q(L^p), \quad 0<\alpha<r/(1+\sigma),\\ (L^p, B^r_{\tau,\tau})_{\alpha(2+\sigma)/r,q} &\hookrightarrow C^\alpha_q(L^p), \quad 0<\alpha<r/(2+\sigma),\\ (L^p, B^r_{\tau,\tau})_{\alpha(1+\sigma)/r,q} &\hookrightarrow S^\alpha_q(L^p), \quad 0<\alpha<r/(1+\sigma). \end{aligned}$$

Proof. Follows from Theorem 6.2, Remark 5.7 and Theorem 5.8. ∎

7. Inverse Embeddings

7.1. **No Inverse Embedding.** It is well known in tensor approximation of high-dimensional functions and approximation with neural networks (see [32]) that highly irregular functions can in some cases be approximated or even represented exactly with low or constant rank or complexity[10]. This fact is reflected in the lack of inverse estimates for tensorized approximation of one-dimensional functions as the next statement shows.

Theorem 7.1 (No Inverse Embedding)**.** *For any $\alpha > 0$, $0 < p, q \leq \infty$ and any $\tilde{\alpha} > 0$*

$$C_q^\alpha(L^p) \not\hookrightarrow B_{p,q}^{\tilde{\alpha}}.$$

Proof. For ease of notation we restrict ourselves to $b = 2$, but the same arguments apply for any $b \geq 2$. The proof boils down to finding a counterexample of a function that can be efficiently represented within $V_{b,m}$ but has "bad" Besov regularity. To this end, we use the *sawtooth* function, see [59] and Figure 3.

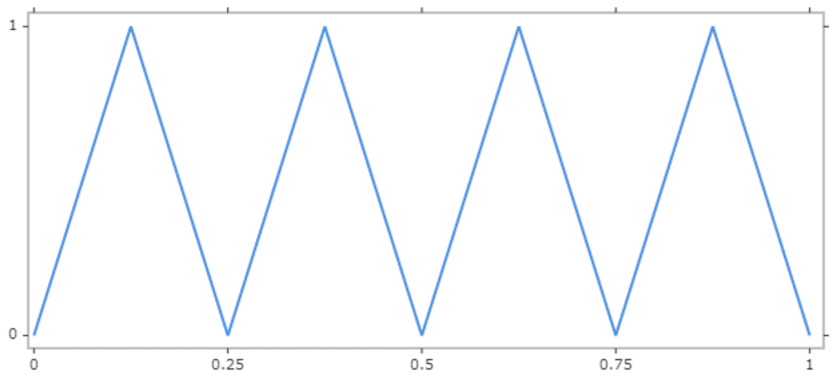


Figure 3. "Sawtooth" function.

Specifically, consider the linear functions

$$\psi_1(y) := y, \quad \psi_2(y) := 1 - y, \quad 0 \leq y < 1.$$

For arbitrary $d \in \mathbb{N}$, set

$$\boldsymbol{\varphi}_d(i_1, \ldots, i_d, y) := \delta_0(i_d)\psi_1(y) + \delta_1(i_d)\psi_2(y).$$

Then, $\varphi_d = T_{b,d}^{-1}\boldsymbol{\varphi}_d \in V_{2,d,m}$ with $r_\nu(\varphi_d) = 2$ for all $1 \leq \nu \leq d$. Thus,

$$\mathrm{compl}_{\mathcal{C}}(\varphi_d) \leq 8d + 2m + 2. \tag{7.1}$$

We can compute the L^p-norm of φ_d as

$$\|\varphi_d\|_p^p = 2^d \int_0^{b^{-d}} (2^d y)^p \, \mathrm{d}y = \frac{1}{p+1}. \tag{7.2}$$

Next, since $C_q^\alpha(L^p)$ satisfies (P1) – (P4), this implies $C_q^\alpha(L^p)$ satisfies the Bernstein inequality (see [19, Chapter 7, Theorem 9.3])

$$\|\varphi_d\|_{C_q^\alpha} \leq Cn^\alpha \|\varphi_d\|_p, \quad \forall \varphi \in \Phi_n. \tag{7.3}$$

On the other hand, by [32, Lemma 5.12],

$$\|\varphi_d\|_{B_{p,q}^{\tilde{\alpha}}} \geq c2^{\tilde{\alpha}d}, \tag{7.4}$$

for any $\tilde{\alpha} > 0$.

[10] Think of a rank-one tensor product of jump functions.

Assume the Bernstein inequality holds in $B^{\tilde{\alpha}}_{p,q}$ for some $\tilde{\alpha} > 0$. For $n \in \mathbb{N}$ large enough, let $d := \lfloor n/8 - m/4 - 1/4 \rfloor \geq 2$. Then, by (7.1), $\varphi_d \in \Phi^{\mathcal{C}}_n$. By (7.3) and (7.4),

$$Cn^{\alpha} \|\varphi_d\|_p \geq \|\varphi_d\|_{C^{\alpha}_q} \gtrsim \|\varphi_d\|_{B^{\tilde{\alpha}}_{p,q}} \gtrsim 2^{\tilde{\alpha}d} \gtrsim 2^{\frac{\tilde{\alpha}}{8}n}.$$

Together with (7.2), this is a contradiction and thus the claim follows. □

In Section 4, we demonstrated that when representing classical tools with the tensorized format we obtain a complexity that is similar (or slightly worse) than for the corresponding classical representation. This reflects the fact that these tools are tailored for approximation in classical smoothness spaces and we therefore cannot expect better "worst case" performance in these spaces. This was also observed in high-dimensional approximation, see [57, 7, 33].

On the other hand, theorem 7.1 demonstrates that tensor networks are efficient for functions that cannot be described by classical smoothness (see also [1]). The cost n in $\Phi^{\mathcal{C}}_n$ depends on both the discretization level d and the tensor ranks r_ν that, in a sense, reflect algebraic properties of the target function.

The proof of Theorem 7.1 shows that tensor networks are particularly effective in approximating functions with a high degree of self-similarity. Such functions do not have to possess any smoothness in the classical sense. The ranks reflect global algebraic features, while smoothness reflects local "rate of change" features.

7.2. **Inverse Embedding For Restricted Depth.** The above result shows that tensor networks are effective to approximate functions that do not possess Sobolev or Besov smoothness. However, one would expect that, if one enforces a full-rank structure or, equivalently, limits the depth of the corresponding tensor network, we should recover inverse estimates similar to classical tools from Section 4.

Theorem 7.2 (Inverse Embedding for Restricted $\Phi^{\mathcal{N}}_n$)**.** *Let $1 \leq p < \infty$. Define for $n \in \mathbb{N}$, $k_{\mathrm{B}} \geq 1$ and $c_{\mathrm{B}} > 0$ the restricted sets*

$$\Phi^{\mathrm{B}}_n := \{\varphi \in V_{b,m} : \mathrm{compl}_{\mathcal{N}}(\varphi) \leq n \quad \text{and} \quad d(\varphi) \leq k_{\mathrm{B}} \log_b(n) + c_{\mathrm{B}}\}, \tag{7.5}$$

where $d(\varphi)$ is the minimal possible level for a tensorized representation of φ. Then,

(i) Φ^{B}_n satisfies (P1) – (P6) and thus $A^{\alpha}_q(L^p, (\Phi^{\mathrm{B}}_n))$ are quasi-normed linear spaces satisfying direct and inverse estimates.

(ii) The following inverse estimate holds:

$$|\varphi|_{B^{m+1}_{\tau,\tau}} \leq C \|\varphi\|_p \, b^{c_{\mathrm{B}}(m+1)} n^{k_{\mathrm{B}}(m+1)},$$

for any $\varphi \in \Phi^{\mathrm{B}}_n$, where $\tau > 0$ is the Sobolev embedding number.

(iii) We have the continuous embeddings

$$A^{\alpha}_q(L^p, (\Phi^{\mathrm{B}}_n)) \hookrightarrow (L^p, B^{m+1}_{\tau,\tau})_{\frac{\alpha}{k_{\mathrm{B}}(m+1)},q}, \quad 0 < \alpha < k_{\mathrm{B}}(m+1),$$
$$A^{k_{\mathrm{B}}(m+1)}_{\infty}(L^p, (\Phi^{\mathrm{B}}_n)) \hookrightarrow B^{m+1}_{\tau,\tau}.$$

Proof. The restriction on Φ^{B}_n ensures functions such as the sawtooth function from Figure 3 are excluded.

(i) (P1) – (P3) is trivial. For (P4): since $\Phi^{\mathcal{N}}_n + \Phi^{\mathcal{N}}_n \subset \Phi^{\mathcal{N}}_{cn}$ and

$$d(\varphi_1 + \varphi_2) \leq \max(d_1, d_2) \leq k_{\mathrm{B}} \log_b(n) + c_{\mathrm{B}} \leq k_{\mathrm{B}} \log_b(cn) + c_{\mathrm{B}}$$

for $\varphi_1, \varphi_2 \in \Phi^{\mathrm{B}}_n$, then (P4) is true for Φ^{B}_n for the same c. For (P5): $\bigcup_{n=0}^{\infty} \Phi^{\mathcal{N}}_n = \bigcup_{n=0}^{\infty} \Phi^{\mathrm{B}}_n$ and thus density follows as in Theorem 2.21. Finally, (P6) follows as in Lemma 3.14.

(ii) Any $\varphi \in \Phi_n^{\mathrm{B}}$ is a spline with at most $b^{d(\varphi)} \leq b^{c_{\mathrm{B}}} n^{k_{\mathrm{B}}}$ pieces. Thus, we can use classical inverse estimates to obtain the inequality.
(iii) Follows from (ii) and Theorem 6.2.

□

Remark 7.3. *Inverse embeddings also hold for restricted sets* $\Phi_n^{\mathcal{S},B}$ *and* $\Phi_n^{\mathcal{C},B}$ *defined by* (7.5) *with complexity measure* $\mathrm{compl}_{\mathcal{N}}$ *replaced by* $\mathrm{compl}_{\mathcal{S}}$ *and* $\mathrm{compl}_{\mathcal{C}}$ *respectively. Indeed, their approximation classes* $A_q^\alpha(L^p, (\Phi_n^{\mathcal{S},B}))$ *and* $A_q^\alpha(L^p, (\Phi_n^{\mathcal{C},B}))$ *are both included in* $A_q^\alpha(L^p, (\Phi_n^B))$.

A discussion on the role of depth and the approximation power of the restricted class Φ_n^B can be found in Section 8.

8. The Roles of Depth and Sparsity

One could ask how the direct estimates would change if we replace $\Phi_n^{\mathcal{N}}$ with Φ_n^{B} from Theorem 7.2. Strictly speaking, this would require lower bounds for the complexity $n := \mathrm{compl}_{\mathcal{N}}(\varphi)$. Nonetheless, a simple example reveals some key features of Φ_n^{B}, assuming the upper bounds for n in this section are sharp to some degree.

Consider the case of Sobolev spaces $W^{k,p}$ from Theorem 5.2 with $k \leq m+1$. Then, assuming the upper bounds from Theorem 5.2 are sharp, we have

$$n \sim C_1(b, m) b^d.$$

The approximands of Theorem 5.2 satisfy $\varphi \in \Phi_n^{\mathrm{B}}$ for $k_{\mathrm{B}} = 1$ and $c_{\mathrm{B}} = c_{\mathrm{B}}(b, m)$. Hence, in this case we would indeed obtain the same approximation rate as with $\Phi_n^{\mathcal{C}}$, in addition to inverse estimates from Theorem 7.2.

Consider now $W^{k,p}$ with $k > m+1$. In this case we have $k_{\mathrm{B}} = k_{\mathrm{B}}(k) > 1$ with $k_{\mathrm{B}} \to \infty$ as $k \to \infty$. In other words, if we fix Φ_n^{B} with some $k_{\mathrm{B}} > 1$, then we would obtain direct estimates for $W^{k,p}$ as in Theorem 5.2, with $0 < k \leq \bar{k}$ for $\bar{k}$ depending on $k_{\mathrm{B}} > 1$. I.e., $\bar{k} = m+1$ for $k_{\mathrm{B}} = 1$ and $\bar{k} \to \infty$ as $k_{\mathrm{B}} \to \infty$.

Finally, consider the direct estimate for Besov spaces $B_{\tau,\tau}^{\alpha_\delta}$ from Theorem 5.8. Again, assuming the upper bounds of this lemma are sharp and $\alpha < m+1$, we would obtain

$$n \sim CNd,$$

where N is the number of knots of a corresponding free knot spline and d is the maximal level of said spline. From Lemma 5.6, we could assume $d \sim \log(N)$ and in this case

$$d \sim \log(N) \lesssim \log(N) + \log\log(N) \lesssim \log(n),$$

in which case we claim we could recover direct estimates as in Theorem 5.8. However, note that, in order to recover near to optimal rates, we would have to consider the complexity measure $\mathrm{compl}_{\mathcal{S}}$ (or $\mathrm{compl}_{\mathcal{N}}$) – i.e., we have to account for sparsity. And, as for Sobolev spaces, to capture an arbitrary regularity $\alpha \geq m+1$, Φ_n^{B} is not sufficient anymore as it requires an arbitrary depth.

Thus, when comparing approximation with tensor networks to approximation with classical tools, we see that depth can very efficiently replicate approximation with higher-order polynomials: that is, with exponential convergence. It was already noted in [28, 53] that (deep) tree tensor networks can represent polynomials with bounded rank, while the canonical (CP) tensor format, corresponding to a shallow network, can only do so approximately with ranks bounded by the desired accuracy. Moreover, similar observations about depth and polynomial degree were made about ReLU networks, see, e.g., [60, 48, 49].

On the other hand, sparsity is necessary to recover classical adaptive (free knot spline) approximation, see Theorem 6.6. In other words: sparse tensor networks can replicate

h-adaptive approximation, while deep tensor networks can replicate p-adaptive approximation, and, consequently, sparse *and* deep tensor networks can replicate hp-adaptive approximation.

For the approximation with sparse tensor networks, the development of algorithms that achieve in practice the expected rates of convergence remains mostly an open problem. In a statistical learning setting, different algorithms have been recently proposed, either based on sparsity-inducing regularization or on model selection strategies for selecting a sparsity pattern (position of non-zero entries) [30, 31, 45].

9. The Role of Tensorization

The tensorization of functions is a milestone allowing the use of tensor networks for the approximation of multivariate functions. In this section, we interpret tensorization as a non-standard and powerful featuring step. Then, we discuss the role of this particular featuring.

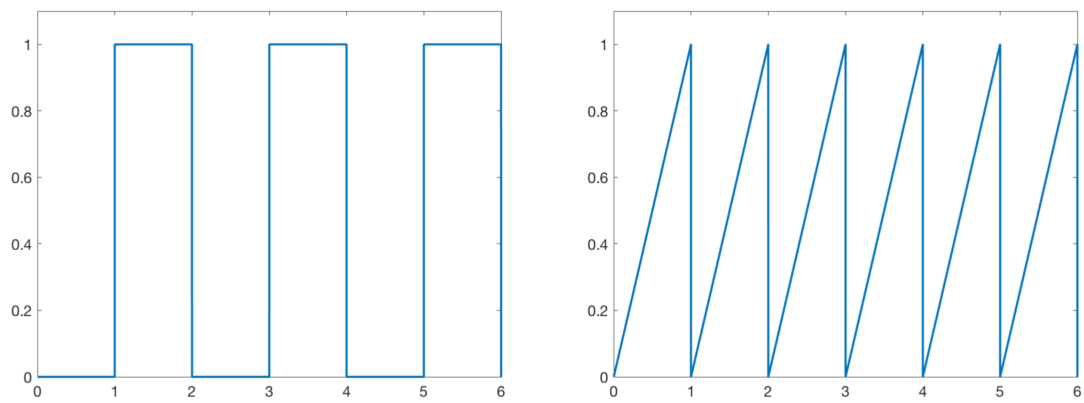
When applying $t_{b,d}^{-1}$ to the input variable x, we create $d+1$ new variables $(i_1, ..., i_d, y)$ defined by

$$i_\nu = \sigma(b^\nu x), \quad \sigma(t) = \lfloor t \rfloor \, mod \, b,$$

$1 \leq \nu \leq d$, and

$$y = \tilde{\sigma}(b^d x), \quad \tilde{\sigma}(t) = t - \lfloor t \rfloor,$$

see Figure 4 for a graphical representation of functions σ and $\tilde{\sigma}$. Then for each $1 \leq \nu \leq d$,



(A) σ (B) $\tilde{\sigma}$

Figure 4. Functions σ and $\tilde{\sigma}$

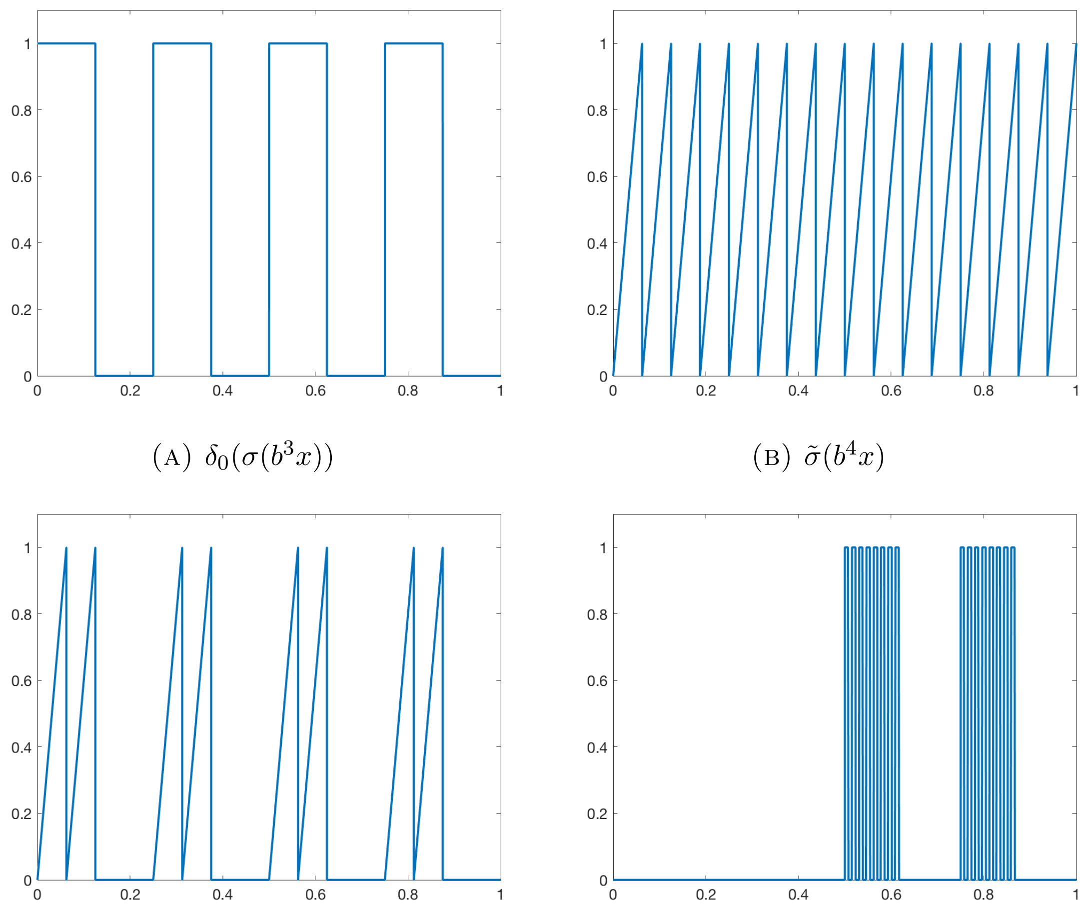
we create b features $\delta_{j_\nu}(i_\nu)$, $0 \leq j_\nu \leq b-1$, and we also create $m+1$ features $\varphi_k(y) = y^k$ from the variable y (or other features for S different from $\mathbb{P}_m$).[11] Figure 5 provides an illustration of these features and of products of these features. Finally, tensorization can be seen as a featuring step with a featuring map

$$\Phi : [0,1) \to \mathbb{R}^{b^d(m+1)}$$

which maps $x \in [0,1)$ to a $(d+1)$-order tensor

$$\Phi(x)_{j_1,...,j_{d+1}} = \delta_{j_1}(\sigma(bx)) \dots \delta_{j_d}(\sigma(b^d x)) \tilde{\sigma}(b^d x)^{j_{d+1}}.$$

[11]For $m = 0$, the extra variable y is not exploited. For $m = 1$, we only consider the variable y and for $m > 1$, we exploit more from this variable.



(A) $\delta_0(\sigma(b^3x))$ (B) $\tilde{\sigma}(b^4x)$

(C) $\delta_0(b^3x)\tilde{\sigma}(b^4x)$ (D) $\delta_1(\sigma(bx))\delta_0(\sigma(b^3x))\delta_0(b^7x)$

FIGURE 5. Representation of some features and their products for $b = 2$.

A function $\varphi \in V_{b,d,m}$ is then represented by $\varphi(x) = \sum_j \Phi(x)_j a_j$, where a is a $(d+1)$-order tensor with entries associated with the $b^d(m+1)$ features. When considering for a a full tensor (not rank-structured), it results in a linear approximation tool which is equivalent to spline approximation. Note that functions represented on Figures 5c and 5d are obtained by summing many features $\Phi(x)_{j_1,\ldots,j_{d+1}}$. However, these functions, which have rank-one tensorizations, can be represented with a rank-one tensor a in the feature tensor space, and thus can be encoded with very low complexity within our nonlinear approximation tool.

Increasing d means considering more and more features, and is equivalent to refining the discretisation. At this point, tensorization is an interpretation of a univariate function as a multivariate function, but it is also an alternative way to look at discretization.

Another featuring (which is rather straight-forward) would have consisted in taking new variables (or features) $x_{j,k} = (b^d x - j)^k \mathbb{1}_{I_j}(x) = (b^d x - j)^k \mathbb{1}_{[0,1)}(b^d x - j)$, $0 \leq k \leq m$, $0 \leq j < b^d$, where I_j is the interval $[b^d j, b^d(j+1))$. This also leads to $(m+1)b^d$ features. When considering a simple linear combination of these features, we also end up with classical fixed knot spline approximation.

Both featuring (or tensorization) methods lead to a linear feature space corresponding to classical spline approximation. One may ask what is the interest of using the very specific feature map Φ? In fact, the use of the particular feature map Φ, which is related to multi-resolution analysis, allows to further exploit sparsity or low-rankness of the tensor when approximating functions from smoothness spaces, and possibly other classes of functions such as fractals. It is well known that the approximation class of splines of degree $m \geq r-1$ is the Sobolev space $W^{r,p}$. Therefore, whatever the featuring used, the approximation class of the resulting linear approximation tool (taking linear combinations of the features) is the Sobolev space $W^{m+1,p}$. Near-optimal performance is achieved by the proposed approximation tool for a large range of smoothness spaces for any fixed m (including $m=0$), at the price of letting d grow (or equivalently the depth of the tensor networks) to capture higher regularity of functions. When working with a fixed m, exploiting low-rank structures will then be crucial.

This reveals that the power of the approximation tool considered in this work comes from the combination of a particular featuring step (the tensorization step) and the use of tensor networks.

Appendix A. Sobolev and Besov Spaces of Tensorizations

A.1. **Sobolev Spaces.** Consider functions f in the Sobolev space $W^{k,p} := W^{k,p}([0,1))$, equipped with the (quasi-)norm

$$\|f\|_{W^{k,p}} = (\|f\|_p^p + |f|_{W^{k,p}}^p)^{1/p} \quad (p < \infty), \quad \|f\|_{W^{k,\infty}} = \max\{\|f\|_p, |f|_{W^{k,\infty}}\},$$

where $|f|_{W^{k,p}}$ is a (quasi-)semi-norm defined by

$$|f|_{W^{k,p}} = \|D^k f\|_p,$$

with $D^k f := f^{(k)}$ the k-th weak derivative of f. Since f and its tensorization $\boldsymbol{f} = T_{b,d} f$ are such that $f(x) = \boldsymbol{f}(j_1, \dots, j_d, b^d x - j)$ for $x \in [b^d j, b^d(j+1))$ and $j = \sum_{k=1}^d b^{d-k} j_k$, we deduce that

$$D^k f(x) = b^{kd} \frac{\partial^k}{\partial y^k} \boldsymbol{f}(j_1, \dots, j_d, b^d x - j)$$

for $x \in [b^d j, b^d(j+1))$, that means that D^k can be identified with a rank-one operator over $\mathbf{V}_{b,d,L^p}$,

$$T_{b,d} \circ D^k \circ T_{b,d}^{-1} = id_{\{1,\dots,d\}} \otimes (b^{kd} D^k).$$

We deduce from Theorem 2.15 that for $f \in W^{k,p}$,

$$|f|_{W^{k,p}} = \|D^k f\|_p = \|T_{b,d}(D^k f)\|_p = b^{kd} \|(id_{\{1,\dots,d\}} \otimes D^k) \boldsymbol{f}\|_p,$$

with

$$(id_{\{1,\dots,d\}} \otimes D^k) \boldsymbol{f} = \sum_{j \in I_b^d} \delta_{j_1} \otimes \dots \otimes \delta_{j_d} \otimes D^k \boldsymbol{f}(j_1, \dots, j_d, \cdot).$$

Then we deduce that if $f \in W^{k,p}$, $\boldsymbol{f} = T_{b,d} f$ is in the algebraic tensor space

$$\mathbf{V}_{b,d,W^{k,p}} := (\mathbb{R}^{I_b})^{\otimes d} \otimes W^{k,p},$$

and

$$|f|_{W^{k,p}} = b^{kd} \| \sum_{j_1 \in I_b} \dots \sum_{j_d \in I_b} \delta_{j_1} \otimes \dots \otimes \delta_{j_d} \otimes D^k \boldsymbol{f}(j_1, \dots, j_d, \cdot) \|_p.$$

This implies that $T_{b,d} W^{k,p} \subset \mathbf{V}_{b,d,W^{k,p}}$ but $T_{b,d}^{-1}(\mathbf{V}_{b,d,W^{k,p}}) \not\subset W^{k,p}$. In fact, $T_{b,d}^{-1}(\mathbf{V}_{b,d,W^{k,p}}) = W^{k,p}(\mathcal{P}_{b,d})$, the broken Sobolev space associated with the partition $\mathcal{P}_{b,d} = \{[b^d j, b^d(j+1)) : 0 \le j \le b^d - 1\}$. From the above considerations, we deduce

Theorem A.1. *For any $0 < p \le \infty$ and $k \in \mathbb{N}_0$, $T_{b,d}$ is a linear isometry from the broken Sobolev space $W^{k,p}(\mathcal{P}_{b,d})$ to $\mathbf{V}_{b,d,W^{k,p}}$ equipped with the (quasi-)norm*

$$\|\boldsymbol{f}\|_{W^{k,p}} = (\|\boldsymbol{f}\|_p^p + |\boldsymbol{f}|_{W^{k,p}}^p)^{1/p} \quad (p < \infty), \quad \|\boldsymbol{f}\|_{W^{k,\infty}} = \max\{\|\boldsymbol{f}\|_\infty, |\boldsymbol{f}|_{W^{k,\infty}}\},$$

where $|\cdot|_{W^{k,p}}$ is a (quasi-)semi-norm defined by

$$|\boldsymbol{f}|_{W^{k,p}}^p := b^{d(kp-1)} \sum_{(j_1,\dots,j_d) \in I_b^d} |\boldsymbol{f}(j_1, \dots, j_d, \cdot)|_{W^{k,p}}^p$$

for $p < \infty$, and

$$|\boldsymbol{f}|_{W^{k,\infty}}^\infty := b^{dk} \max_{(j_1,\dots,j_d) \in I_b^d} |\boldsymbol{f}(j_1, \dots, j_d, \cdot)|_{W^{k,\infty}}.$$

A.2. **Besov Spaces.** Let $f \in L^p$, $0 < p \leq \infty$ and consider the difference operator

$$\Delta_h : L^p([0,1)) \to L^p([0,1-h)),$$
$$\Delta_h[f](\cdot) := f(\cdot + h) - f(\cdot).$$

For $k = 2, 3, \ldots$, the k-th difference is defined as

$$\Delta_h^k := \Delta_h \circ \Delta_h^{k-1},$$

with $\Delta_h^1 := \Delta_h$. *The k-th modulus of smoothness* is defined as

$$\omega_k(f,t)_p := \sup_{0<h\leq t} \left\| \Delta_h^k[f] \right\|_p, \quad t > 0. \tag{A.1}$$

Definition A.2 (Besov Spaces)**.** *For parameters $\alpha > 0$ and $0 < p, q \leq \infty$, define $k := \lfloor \alpha \rfloor + 1$ and the Besov (quasi-)semi-norm as*

$$|f|_{B^\alpha_{p,q}} := \begin{cases} \left(\int_0^1 [t^{-\alpha}\omega_k(f,t)_p]^q \frac{\mathrm{d}t}{t} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{0<t\leq 1} t^{-\alpha}\omega_k(f,t)_p, & q = \infty. \end{cases}$$

The Besov (quasi-)norm is defined as

$$\|f\|_{B^\alpha_{p,q}} := \|f\|_p + |f|_{B^\alpha_{p,q}}.$$

The Besov space *is defined as*

$$B^\alpha_{p,q} := \left\{ f \in L^p : \ \|f\|_{B^\alpha_{p,q}} < \infty \right\}.$$

As in Appendix A.1, we would like to compare the Besov space $B^\alpha_{p,q}$ with the algebraic tensor space

$$\mathbf{V}_{b,d,B^\alpha_{p,q}} := (\mathbb{R}^{I_b})^{\otimes d} \otimes B^\alpha_{p,q}.$$

First, we briefly elaborate how the Besov (quasi-)semi-norm scales under affine transformations of the interval. I.e., suppose we are given a function $f : [a,b) \to \mathbb{R}$ with $-\infty < a < b < \infty$ and a transformed $\bar{f}$ such that

$$\bar{f} : [\bar{a}, \bar{b}) \to \mathbb{R}, \quad \bar{x} \mapsto x := \frac{b-a}{\bar{b}-\bar{a}}(\bar{x} - \bar{a}) + a \mapsto f(x) = \bar{f}(\bar{x}),$$

for $-\infty < \bar{a} < \bar{b} < \infty$. Then,

$$\Delta_{\bar{h}}^k[\bar{f}] : [\bar{a}, \bar{b} - r\bar{h}) \to \mathbb{R},$$
$$\Delta_{\bar{h}}^k[\bar{f}](\bar{x}) = \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} \bar{f}(\bar{x} + i\bar{h}) = \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} f(x + ih) = \Delta_h^k[f](x)$$

with

$$h := \frac{b-a}{\bar{b}-\bar{a}}\bar{h}.$$

For $0 < p < \infty$, we obtain for the modulus of smoothness

$$\omega_k(\bar{f}, \bar{t})_p^p = \frac{\bar{b}-\bar{a}}{b-a}\omega_k(f,t)_p^p, \quad t := \frac{b-a}{\bar{b}-\bar{a}}\bar{t},$$

and for $p = \infty$, $\omega_k(\bar{f}, \bar{t})_\infty = \omega_k(f,t)_\infty$. Finally, for the Besov (quasi-)semi-norm this implies

$$|\bar{f}|_{B^\alpha_{p,q}} = \left(\frac{\bar{b}-\bar{a}}{b-a}\right)^{1/p-\alpha} |f|_{B^\alpha_{p,q}}, \quad 0 < q \le \infty,\ 0 < p < \infty,$$
$$|\bar{f}|_{B^\alpha_{p,q}} = \left(\frac{\bar{b}-\bar{a}}{b-a}\right)^{-\alpha} |f|_{B^\alpha_{p,q}}, \quad 0 < q \le \infty,\ p = \infty.$$

With this scaling at hand, for $p < \infty$, what remains is "adding up" Besov (quasi-)norms of partial evaluations $\boldsymbol{f}(j_1, \ldots, j_d, \cdot)$. The modulus of smoothness from (A.1) is not suitable for this task. Instead, we can use an equivalent measure of smoothness via the *averaged modulus of smoothness* (see [19, §5 of Chapter 6 and §5 of Chapter 12])

$$\mathrm{w}_k(f,t)_p^p := \frac{1}{t}\int_0^t \left\|\Delta_h^k[f]\right\|_p^p \,\mathrm{d}h, \quad 0 < p < \infty.$$

We can then define a (quasi-)semi-norm

$$|f|_{B^\alpha_{p,q}} = \left(\int_0^1 [t^{-\alpha}\mathrm{w}_k(f,t)_p]^q \frac{\mathrm{d}t}{t}\right)^{1/q}, \quad 0 < q < \infty,$$

which is equivalent to the former one from Definition A.2 and therefore results in the same Besov space $B^\alpha_{p,q}$. Expanding the right-hand-side and interchanging the order of integration allows us to add up the contributions to $|f|_{B^\alpha_{p,q}}$ over the intervals $[b^d j, b^d(j+1))$, provided that $q = p$. However, note that this is not the same as summing over $|\boldsymbol{f}(j_1, \ldots, j_d, \cdot)|_{B^\alpha_{p,q}}$, since the latter necessarily omits the contributions of $|\Delta_h^k[f]|$ across the right boundaries of the intervals $[b^d j, b^d(j+1))$.

Example A.3. *Consider the function*

$$f(x) := \begin{cases} 1, & 0 \le x \le 1/2, \\ 0, & \textit{otherwise}. \end{cases}$$

Take $0 < \alpha < 1$ and $r = 1$ in Definition A.2. The first difference is then

$$\Delta_h[f](x) = \begin{cases} 1, & 1/2 - h < x \le 1/2, \\ 0, & \textit{otherwise}. \end{cases}$$

For $0 < p < \infty$,

$$\|\Delta_h[f]\|_p^p = h,$$

and for $p = \infty$,

$$\|\Delta_h[f]\|_\infty = 1.$$

Thus, for the ordinary modulus of smoothness we obtain

$$\omega_k(f,t)_p = t^{1/p}, \quad 0 < p < \infty,$$
$$\omega_k(f,t)_\infty = 1.$$

Inserting this into Definition A.2, we see that $f \in B^\alpha_{p,q}$ if and only if $p \neq \infty$ and $0 < \alpha < 1/p$. In this case $0 < |f|_{B^\alpha_{p,q}} < \infty$.

On the other hand, for $b = 2$ and $d = 1$, the partial evaluations of the tensorization $T_{2,1}f$ are the constant functions 0 and 1. Thus, any Besov semi-norm of these partial evaluations is 0 and consequently the sum as well. We see that, unlike in Theorem A.1, even if a function f has Besov regularity, the Besov norm of f is in general not equivalent to the sum of the Besov norms of partial evaluations.

Theorem A.4. *Let $0 < p = q \leq \infty$ and $\alpha > 0$. Let $B^\alpha_{p,p}$ be equipped with the (quasi-)norm associated with the modulus of smoothness when $p = \infty$ or the averaged modulus of smoothness when $p < \infty$. Then, we equip the tensor space $\mathbf{V}_{b,d,B^\alpha_{p,p}}$ with the (quasi-)norm*

$$\|\boldsymbol{f}\|_{B^\alpha_{p,p}} := (\|\boldsymbol{f}\|^p_p + |\boldsymbol{f}|^p_{B^\alpha_{p,p}})^{1/p} \quad (p < \infty), \quad \|\boldsymbol{f}\|_{B^\alpha_{\infty,\infty}} := \max\{\|\boldsymbol{f}\|_\infty, |\boldsymbol{f}|_{B^\alpha_{\infty,\infty}}\},$$

where $|\cdot|_{B^\alpha_{p,p}}$ is a (quasi-)semi-norm defined by

$$|\boldsymbol{f}|^p_{B^\alpha_{p,p}} := b^{d(\alpha p-1)} \sum_{(j_1,\dots,j_d)\in I^d_b} |\boldsymbol{f}(j_1,\dots,j_d,\cdot)|^p_{B^\alpha_{p,p}},$$

for $p < \infty$, and

$$|\boldsymbol{f}|_{B^\alpha_{\infty,\infty}} := b^{d\alpha} \max_{(j_1,\dots,j_d)\in I^d_b} |\boldsymbol{f}(j_1,\dots,j_d,\cdot)|_{B^\alpha_{\infty,\infty}}$$

Then, $T_{b,d}(B^\alpha_{p,p}) \hookrightarrow \mathbf{V}_{b,d,B^\alpha_{p,p}}$ with

$$|f|_{B^\alpha_{p,p}} \geq |T_{b,d}(f)|_{B^\alpha_{p,p}}.$$

Appendix B. Proofs for Section 2

Proof of Lemma 2.6. We have

$$t_{b,\bar d}(i_1,\dots,i_{\bar d},y) = \sum_{k=1}^{\bar d} i_k b^{-k} + b^{-\bar d} y = \sum_{k=1}^{d} i_k b^{-k} + \sum_{k=1}^{\bar d - d} i_{k+d} b^{-k-d} + b^{-\bar d} y = \sum_{k=1}^{d} i_k b^{-k} + b^{-d} z,$$

with $z = \sum_{k=1}^{\bar d-d} i_{k+d} b^{-k} + b^{-(\bar d-d)} y = t_{b,\bar d-d}(i_{d+1},\dots,i_{\bar d},y)$, which proves the first statement. Then consider an elementary tensor $\boldsymbol{v} = v_1 \otimes \dots \otimes v_d \otimes g \in \mathbf{V}_{b,d}$, with $v_k \in \mathbb{R}^{I_b}$ and $g \in \mathbb{R}^{[0,1)}$. We have

$$\begin{aligned} T_{b,\bar d} \circ T^{-1}_{b,d} \boldsymbol{v}(i_1,\dots,i_{\bar d},y) &= \boldsymbol{v}(t^{-1}_{b,d} \circ t_{b,\bar d}(i_1,\dots,i_{\bar d},y)) \\ &= \boldsymbol{v}(i_1,\dots,i_d,t_{b,\bar d-d}(i_{d+1},\dots,i_{\bar d},y)) \\ &= v_1(i_1)\dots v_d(i_d) g(t_{b,\bar d-d}(i_{d+1},\dots,i_{\bar d},y)) \\ &= v_1(i_1)\dots v_d(i_d) T_{b,\bar d-d} g(i_{d+1},\dots,i_{\bar d},y) \\ &= (v_1 \otimes \dots \otimes v_d \otimes (T_{b,\bar d-d} g))(i_{d+1},\dots,i_{\bar d},y), \end{aligned}$$

which proves the second property. The last property simply follows from $T_{b,d} \circ T^{-1}_{b,\bar d} = (T_{b,\bar d} \circ T^{-1}_{b,d})^{-1} = (id_{\{1,\dots,d\}} \otimes T_{b,\bar d-d})^{-1} = id_{\{1,\dots,d\}} \otimes T^{-1}_{b,\bar d-d}$. □

Proof of Proposition 2.7. Subsets of the form $J \times A$, with A a Borel set of $[0,1)$ and $J = \times_{k=1}^d J_k$ with $J_k \subset I_b$, $1 \leq k \leq d$, form a generating system of the Borel σ-algebra of $I^d_b \times [0,1)$. The image of such a set $J \times A$ through $t_{b,d}$ is $\bigcup_{j\in J} A_j$, where $A_{j_1,\dots,j_d} = b^{-d}(j+A)$ with $j = \sum_{k=1}^d j_k b^{d-k}$. Then

$$\begin{aligned} \lambda(t_{b,d}(J\times A)) &= \lambda(\bigcup_{j\in J} A_j) = \#J b^{-d}\lambda(A) = \#J_1 \dots \#J_d b^{-d} \lambda(A) = \mu_b(J_1)\dots\mu_b(J_d)\lambda(A) \\ &= \mu_{b,d}(J\times A). \end{aligned}$$

Then, we conclude on $T_{b,d}$ by noting that it is a linear bijection (Proposition 2.3) which preserves measurability. To prove the second statement, we first note that we clearly have the first inclusion $\mathcal{M}(I_b)^{\otimes d} \otimes \mathcal{M}([0,1)) \subset \mathcal{M}(I^d_b \times [0,1))$. To prove the other inclusion, we note that a tensor $\boldsymbol{f} \in \mathcal{M}(I^d_b \times [0,1))$ admits a representation (2.2) with $\delta_{j_k} \in \mathcal{M}(I_b)$ and $\boldsymbol{f}(j_1,\dots,j_d,\cdot) = f(b^{-d}(j+\cdot)) \in \mathcal{M}([0,1))$ ($j = \sum_{k=1}^d j_k b^{d-k}$), where the latter identification is deduced from Lemma 2.4. The final identification simply follows from the identification $\mathcal{M}(I_b) = \mathbb{R}^{I_b}$. ☐

Proof of Lemma 2.10. $\boldsymbol{f}$ is identified with a tensor in $\mathbf{V}_\beta \otimes \mathbf{V}_{\beta^c}$ with $\mathbf{V}_\beta \in (\mathbb{R}^{I_b})^{\otimes \#\beta}$ and $\mathbf{V}_{\beta^c} = \mathbf{V}_{b,d-\#\beta}$. We have $U^{\min}_{\beta^c}(\boldsymbol{f}) = \{(\boldsymbol{\varphi}_\beta \otimes id_{\beta^c})\boldsymbol{f} : \boldsymbol{\varphi}_\beta \in (\mathbf{V}_\beta)'\}$ with $(\mathbf{V}_\beta)'$ the algebraic dual of $\mathbf{V}_\beta$ (see [22, Corollary 2.19]). Then for any basis $\{\boldsymbol{\varphi}^{j_\beta}_\beta : j_\beta \in I_b^{\#\beta}\}$ of $(\mathbf{V}_\beta)'$, we have $U^{\min}_{\beta^c}(\boldsymbol{f}) = \operatorname{span}\{(\boldsymbol{\varphi}^{j_\beta}_\beta \otimes id_{\beta^c})\boldsymbol{f} : j_\beta \in I_b^{\#\beta}\}$. We conclude by introducing the particular basis $\boldsymbol{\varphi}^{j_\beta}_\beta = \delta_{j_\beta}$, with $\delta_{j_\beta} = \otimes_{\nu\in\beta}\delta_{j_\nu} \in \mathbb{R}^{I_b^{\#\beta}}$, and by noting that $(\delta_{j_\beta} \otimes id_{\beta^c})\boldsymbol{f} = \boldsymbol{f}(j_\beta, \cdot) \in \mathbf{V}_{\beta^c}$. ☐

Proof of Lemma 2.12. For any set $\beta \subset \{1, \dots, d+1\}$ and any partition $\beta = \gamma \cup \alpha$, the minimal subspaces from Definition 2.9 satisfy the hierarchy property (see [34, Corollary 6.18]) $U^{\min}_\beta(\boldsymbol{f}) \subset U^{\min}_\gamma(\boldsymbol{f}) \otimes U^{\min}_\alpha(\boldsymbol{f})$, from which we deduce that $r_\beta(\boldsymbol{f}) \le r_\gamma(\boldsymbol{f}) r_\alpha(\boldsymbol{f})$. Then for $1 \le \nu \le d-1$, by considering $\gamma = \{1, \dots, \nu\}$ and $\alpha = \{\nu+1\}$, we obtain $r_{\nu+1}(\boldsymbol{f}) \le r_\nu(\boldsymbol{f}) r_{\{\nu+1\}}(\boldsymbol{f})$, where $r_{\{\nu+1\}}(\boldsymbol{f}) = \dim U^{\min}_{\{\nu+1\}}(\boldsymbol{f}) \le b$, which yields the first inequality. By considering $\gamma = \{\nu+1\}$ and $\alpha = \{\nu+2, \dots, d+1\}$, we obtain $r_\nu(\boldsymbol{f}) = r_{\{\nu+1,\dots,d+1\}}(\boldsymbol{f}) \le r_{\{\nu+1\}}(\boldsymbol{f}) r_{\{\nu+2,\dots,d+1\}}(\boldsymbol{f}) = r_{\{\nu+1\}}(\boldsymbol{f}) r_{\nu+1}(\boldsymbol{f}) \le b r_{\nu+1}(\boldsymbol{f})$, that is the second inequality. ☐

Proof of Lemma 2.13. We have from Lemma 2.10 that

$$T^{-1}_{b,d-\nu}(U^{\min}_{\{\nu+1,\dots,d+1\}}(\boldsymbol{f}^d)) = \operatorname{span}\{T^{-1}_{b,d-\nu}(\boldsymbol{f}^d(j_1, \dots, j_\nu, \cdot)) : (j_1, \dots, j_\nu) \in I_b^\nu\},$$

where $\boldsymbol{f}^d(j_1, \dots, j_\nu, \cdot) \in \mathbf{V}_{b,d-\nu}$ is a partial evaluation of $\boldsymbol{f}^d$ along the first ν dimensions. We note that

$$\begin{aligned} T^{-1}_{b,d-\nu}(\boldsymbol{f}^d(j_1, \dots, j_\nu, \cdot)) &= ((id_{\{1,\dots,\nu\}} \otimes T^{-1}_{b,d-\nu})\boldsymbol{f}^d)(j_1, \dots, j_\nu, \cdot) \\ &= (T_{b,\nu} \circ T^{-1}_{b,d}\boldsymbol{f}^d)(j_1, \dots, j_\nu, \cdot) = \boldsymbol{f}^\nu(j_1, \dots, j_\nu, \cdot), \end{aligned}$$

where the second equality results from Lemma 2.6. The result then follows from Lemma 2.10 again. ☐

Lemma B.1. *Let S be a closed subspace of L^p, $0 < p \le \infty$. The norm $\|\cdot\|_p$ is a reasonable crossnorm on $(\ell^p(I_b))^{\otimes d} \otimes S$.*

Proof. Let $v_k \in \ell^p(I_b)$, $1 \le k \le d$, and $g \in S$. For $p < \infty$, we have

$$\begin{aligned} \|v_1 \otimes \dots \otimes v_d \otimes g\|_p^p &= \sum_{i_1 \in I_b} \dots \sum_{i_d \in I_b} |v_1(i_1)|^p \dots |v_d(i_d)|^p b^{-d} \int_0^1 |g(y)|^p dy \\ &= \|v_1\|^p_{\ell^p} \dots \|v_d\|^p_{\ell^p} \|g\|_p^p, \end{aligned}$$

and for $p = \infty$,

$$\|v_1 \otimes \dots \otimes v_d \otimes g\|_\infty = \max_{i_1 \in I_b} |v_1(i_1)| \dots \max_{i_d \in I_b} |v_d(i_d)| \operatorname*{ess\,sup}_y |g(y)| = \|v_1\|_{\ell^\infty} \dots \|v_d\|_{\ell^\infty} \|g\|_\infty,$$

which proves that $\|\cdot\|_p$ is a crossnorm. Now consider $p \ge 1$. Then, consider the dual norm

$$\|\varphi\|_p^* = \sup_{\|\boldsymbol{f}\|_p \le 1} |\varphi(\boldsymbol{f})|$$

over the algebraic tensor space $(\ell^p(I_b)^*)^{\otimes d} \otimes S^*$, where V^* stands for the continuous dual of a space V. For $(v, \psi) \in \ell^p(I_b) \times \ell^p(I_b)^*$, we consider the duality pairing $\psi(v) = b^{-1} \sum_{k=0}^{b-1} \psi_k v_k$, such that $\ell^p(I_b)^* = \ell^q(I_b)$ with $1/p + 1/q = 1$. Consider $\phi \in S^*$ and $\varphi_\nu \in \ell^p(I_b)^*$, $1 \le \nu \le d$. To prove that $\|\cdot\|_p$ is a reasonable crossnorm, we have to prove that

$$\|\varphi_1 \otimes \dots \otimes \varphi_d \otimes \phi\|_p^* \le \|\varphi_1\|_{\ell^q} \dots \|\varphi_d\|_{\ell^q} \|\phi\|_p^*,$$

with $\|\phi\|_p^* = \sup_{f\in S, \|f\|_p\le 1} |\phi(f)|$. Let $\varphi = \varphi_1 \otimes \ldots \otimes \varphi_d \in (\ell^p(I_b)^*)^{\otimes d} = \ell^q(I_b^d)$. For $j \in I_b^d$, we let $\delta_j = \delta_{j_1} \otimes \ldots \otimes \delta_{j_d} \in \ell^p(I_b^d)$. Any $\boldsymbol{f} \in V_{b,d,S}$ admits a representation $\boldsymbol{f} = \sum_{j\in I_b^d} \delta_j \otimes g_j$ where $g_j = \boldsymbol{f}(j_1, \ldots, j_d, \cdot) \in L^p$, and

$$|(\varphi_1 \otimes \ldots \otimes \varphi_d \otimes \phi)(\boldsymbol{f})| = |\varphi(\sum_{j\in I_b^d} \delta_j \phi(g_j)| = |\varphi(\mathbf{v})|$$

where $\mathbf{v} \in \ell^p(I_b^d)$ is a tensor with entries $\mathbf{v}(j) = \phi(g_j)$. Also,

$$|\varphi(\mathbf{v})| \le \|\varphi\|_{\ell^p}^* \|\mathbf{v}\|_{\ell^p} \le \|\varphi\|_{\ell^q} \|\phi\|_p^* \|\mathbf{w}\|_{\ell^p},$$

where $\mathbf{w} \in \ell^p(I_b^d)$ is a tensor with entries $\mathbf{w}(j) = \|g_j\|_p = \|\boldsymbol{f}(j_1, \ldots, j_d, \cdot)\|_p$. From Theorem 2.15, we have $\|\mathbf{w}\|_{\ell^\infty} = \max_{j\in I_b^d} \|\boldsymbol{f}(j_1, \ldots, j_d, \cdot)\|_\infty = \|\boldsymbol{f}\|_\infty$, and for $p < \infty$

$$\|\mathbf{w}\|_{\ell^p}^p = b^{-d} \sum_{j\in I_b^d} |\mathbf{w}(j)|^p = b^{-d} \sum_{j\in I_b^d} \|\boldsymbol{f}(j_1, \ldots, j_d, \cdot)\|_p^p = \|\boldsymbol{f}\|_p^p.$$

Therefore, $|(\varphi_1 \otimes \ldots \otimes \varphi_d \otimes \phi)(\boldsymbol{f})| \le \|\varphi\|_{\ell^q} \|\phi\|_p^* \|\boldsymbol{f}\|_p$. We conclude by noting that $\|\cdot\|_{\ell^q}$ is a crossnorm on $\ell^q(I_b^d) = \ell^q(I_b)^{\otimes d}$, so that $\|\varphi\|_{\ell^q(I_b^d)} = \|\varphi_1\|_{\ell^q} \ldots \|\varphi_d\|_{\ell^q}$. □

Proof of Lemma 2.18. By definition, the result is true for $d = 1$. The result is then proved by induction. Assume that for all $f \in S$, $f(b^{-d}(\cdot + k)) \in S$ for all $k \in \{0, \ldots, b^d - 1\}$. Then for $f \in S$, consider the function $f(b^{-d-1}(\cdot + k))$ with $k \in \{0, \ldots, b^{d+1} - 1\}$. We can write $k = bk'' + k'$ for some $k' \in \{0, \ldots, b-1\}$ and $k'' \in \{0, \ldots, b^d - 1\}$. Then for any $x \in [0, 1)$,

$$f(b^{-d-1}(x + k)) = f(b^{-d}(b^{-1}(x + k') + k'')) = g(b^{-1}(x + k'))$$

for some $g \in S$, and $g(b^{-1}(x + k')) = h(x)$ for some $h \in S$. Therefore $f(b^{-d-1}(\cdot + k)) = h(\cdot) \in S$, which ends the proof. □

Proof of Proposition 2.19. For $f \in S$, we have $(T_{b,1}f)(i_1, \cdot) = f(b^{-1}(\cdot + i_1))$. Then from Lemma 2.18, we have $(T_{b,1}f)(i_1, \cdot) \in S$, which implies $f \in V_{b,1,S}$. Now assume $f \in V_{b,d,S}$ for $d \in \mathbb{N}$, i.e. $T_{b,d}f = \boldsymbol{f}^d \in \mathbf{V}_{b,d,S}$. Then $\boldsymbol{f}^d(i_1, \ldots, i_d, \cdot) \in S$ and from Lemma 2.18, $\boldsymbol{f}^d(i_1, \ldots, i_d, b^{-1}(i_{d+1} + \cdot)) \in S$. Then using Lemma 2.6, we have that

$$\boldsymbol{f}^d(i_1, \ldots, i_d, b^{-1}(i_{d+1} + \cdot)) = f \circ t_{b,d}(i_1, \ldots, i_d, t_{b,1}(i_{d+1}, \cdot)) = f \circ t_{b,d+1}(i_1, \ldots, i_{d+1}, \cdot)$$

, which implies that $(T_{b,d+1}f)(i_1, \ldots, i_{d+1}, \cdot) \in S$, and therefore $f \in V_{b,d+1,S}$. □

Proof of Proposition 2.20. Since $0 \in V_{b,d,S}$ for any d, we have $0 \in V_{b,S}$. For $f_1, f_2 \in V_{b,S}$, there exists $d_1, d_2 \in \mathbb{N}$ such that $f_1 \in V_{b,d_1,S}$ and $f_2 \in V_{b,d_2,S}$. Letting $d = \max\{d_1, d_2\}$, we have from Proposition 2.19 that $f_1, f_2 \in V_{b,d,S}$, and therefore $cf_1 + f_2 \in V_{b,d,S} \subset V_{b,S}$ for all $c \in \mathbb{R}$, which ends the proof. □

Lemma B.2 (Density of Step Functions in $L^p([0, 1))$)**.** *For any $0 < p < \infty$, the set of step functions*[12] *is dense in $L^p([0, 1))$.*

Proof. The set of simple functions is dense in $L^p([0, 1))$ for $0 < p < \infty$ (see, e.g., [62, Theorem 18.3]). Then it remains to show that the set of step functions is dense in the set of simple functions. For that, it is sufficient to show that the indicator function 1_A of any measurable set in $[0, 1)$ (hence with finite measure) is the limit of a sequence of step functions. For all $\epsilon > 0$, since the Lebesgue measure λ is outer regular, there exists an open set O containing A and such that $\lambda(O \setminus A) < \epsilon$. The open set O is the union of a countable set of disjoint intervals $(I_k)_{k\in\mathbb{N}}$, $O = \bigcup_{k>0} I_k$. Let n sufficiently large such that $\lambda(\cup_{k>n} I_k) < \epsilon$, and consider the step function $f = \sum_{k=1}^n 1_{I_k}$. For $p \ge 1$,

[12] A step function is a finite linear combination of indicator functions of intervals.

$\|1_A - f\|_{L^p} \leq \|1_{O\setminus A}\|_{L^p} + \|1_O - f\|_{L^p} < 2\epsilon^{1/p}$. For $p < 1$, $\int |1_A - f|^p \leq \int |1_{O\setminus A}|^p + |1_O - f|^p = \int |1_{O\setminus A}| + |1_0 - f| < 2\epsilon$, and thus $\|1_A - f\|_{L^p} < 2^{1/p}\epsilon$. This shows the desired result for any $0 < p < \infty$. Note that this result only exploits the fact that the Lebesgue measure is outer regular, and it can be extended to any space $L^p(\mathbb{R}, \mathcal{B}_\mathbb{R}, m)$ equipped with an outer regular measure. □

Proof of Theorem 2.21. By Lemma B.2, it is sufficient to prove that $V_{b,S}$ is dense in the set of step functions over $[0,1)$. Consider a step function $f = \sum_{i=0}^{n-1} a_i \mathbb{1}_{[x_i,x_{i+1})} \neq 0$, with $0 = x_0 < x_1 < \ldots < x_n = 1$, and $\|f\|_p^p = \sum_{i=0}^{n-1} |a_i|^p (x_{i+1} - x_i)$. Let $x_i^d = b^{-d}\lfloor b^d x_i \rfloor$, $0 \leq i \leq n$, and consider the function $f_d = \sum_{i=0}^{n-1} a_i \mathbb{1}_{[x_i^d, x_{i+1}^d)}$ which is such that $f_d \in V_{b,d,S}$. Then, noting that $x_0^d = x_0 = 0$ and $x_n^d = x_n = 1$, we have

$$
\begin{aligned}
f - f_d &= \sum_{i=0}^{n-1} a_i (\mathbb{1}_{[x_i,x_{i+1})} - \mathbb{1}_{[x_i^d,x_{i+1}^d)}) = \sum_{i=0}^{n-1} a_i (\mathbb{1}_{[x_{i+1}^d,x_{i+1})} - \mathbb{1}_{[x_i^d,x_i)}) \\
&= \sum_{i=0}^{n-2} (a_i - a_{i+1}) \mathbb{1}_{[x_{i+1}^d, x_{i+1})}.
\end{aligned}
$$

Then, noting that $0 \leq x_i - x_i^d \leq b^{-d}$ for all $0 < i < n$, we have

$$
\begin{aligned}
\|f - f_d\|_p^p &= \sum_{i=0}^{n-2} |a_i - a_{i+1}|^p (x_{i+1} - x_{i+1}^d) \leq 2^p b^{-d} \sum_{i=0}^{n-1} |a_i|^p \\
&\leq 2^p b^{-d} \|f\|_p^p \big(\min_{0 \leq i \leq n-1} (x_{i+1} - x_i) \big)^{-1},
\end{aligned}
$$

so that $\|f - f_d\|_p \to 0$ as $d \to \infty$, which ends the proof. □

Proof of Lemma 2.23. (i) If $f \in S$, from Proposition 2.19, we have $f \in V_{b,\nu,S}$ for any ν, so that $r_{\nu,\nu}(f) \leq \dim S$. Then, using Corollary 2.14, we have $r_{\nu,d}(f) = r_{\nu,\nu}(f) \leq \dim S$. The other bound $r_{\nu,d}(f) \leq b^\nu$ results from Lemma 2.22.
(ii) The fact that $f \in V_{b,\bar{d},S}$ follows from Proposition 2.19. Then, from Corollary 2.14, we have that $r_{\nu,\bar{d}}(f) = r_{\nu,\nu}(f)$ for all $1 \leq \nu \leq \bar{d}$. For $1 \leq \nu \leq d$, Corollary 2.14 also implies $r_{\nu,\nu}(f) = r_{\nu,d}(f)$ and we obtain the desired inequality from Lemma 2.22. For $\nu > d$, we note that $r_{\nu,\nu}(f) = \dim U^{\min}_{\{\nu+1\}}(T_{b,\nu} f)$. From Proposition 2.19, we know that $T_{b,\nu} f \in V_{b,\nu,S}$ for $\nu \geq d$, so that $U^{\min}_{\{\nu+1\}}(T_{b,\nu} f) \subset S$ and $r_{\nu,\nu}(f) \leq \dim(S)$. The other bound $r_{\nu,\bar{d}}(f) \leq b^\nu$ results from Lemma 2.22. □

Proof of Lemma 2.25. Let $\{\phi_l\}_{1 \leq l \leq \dim S}$ be a basis of S, such that for $g \in L^p$, $\mathcal{I}_S(g) = \sum_{l=1}^{\dim S} \phi_l \sigma_l(g)$, with σ_l a linear map from L^p to $\mathbb{R}$. For $f \in L^p$, and $x \in [b^{-d}j, b^{-d}(j+1))$,

$$
\mathcal{I}_{b,d,S} f(x) = \sum_{l=1}^{\dim S} \phi_l(b^d x - j) \sigma_l(f(b^{-d}(j + \cdot))).
$$

We have $f(b^{-d}(j + \cdot)) = \boldsymbol{f}(j_1, \ldots, j_d, \cdot)$, with $j = \sum_{k=1}^d b^{d-k} j_k$ and $\boldsymbol{f} = T_{b,d} f$, so that

$$
T_{b,d}(\mathcal{I}_{b,d,S} f)(j_1, \ldots, j_d, y) = \sum_{l=1}^{\dim S} \phi_l(y) \sigma_l(\boldsymbol{f}(j_1, \ldots, j_d, \cdot)).
$$

For $\boldsymbol{f} = \varphi_1 \otimes \dots \varphi_d \otimes g$, using the linearity of σ_l, we then have

$$\begin{aligned} T_{b,d}(\mathcal{I}_{b,d,S}(T_{b,d}^{-1}\boldsymbol{f}))(j_1,\dots,j_d,y) &= \varphi_1(j_1)\dots\varphi_d(j_d)(\sum_{l=1}^{\dim S}\phi_l(y)\sigma_l(g)) \\ &= \varphi_1(j_1)\dots\varphi_d(j_d)\mathcal{I}_S(g)(y), \end{aligned}$$

which proves (2.13). □

Appendix C. Proofs for Section 3

Proof of Proposition 3.4. (ii). Let $\varphi_A,\ \varphi_B \in \Phi_n$ with $\varphi_A \in V_{b,d_A,S}$, $\varphi_B \in V_{b,d_B,S}$ and w.l.o.g. $d_B \geq d_A$. Set $r_A := r_{\max}(\varphi_A)$ and $r_B := r_{\max}(\varphi_B)$. Then,

$$\begin{aligned} \operatorname{compl}(\varphi_A+\varphi_B) &\leq bd_A(\max(r_A,\dim S)+r_B)^2 + (\max(r_A,\dim S)+r_B)\dim S \\ &\leq 2bd_Ar_B^2 + 4bd_Ar_A^2 + 4bd_A(\dim S)^2 + r_A\dim S + r_B\dim S + (\dim S)^2 \\ &\leq [4+4(\dim S)^2+\dim S]n + 4n^2 \leq [8+4(\dim S)^2+\dim S]n^2. \end{aligned}$$

(i). Let W_0 denote the principal Branch of the Lambert W function. Recall, that the multi-valued Lambert function W is the solution to the equation

$$\omega e^{\omega} = z$$

for complex valued ω and z, with countably many solutions W_k, where W_0 denotes the principal branch by convention.

Take $n \in \mathbb{N}$ large enough such that

$$d_A := \left\lfloor \frac{1}{\ln(b)} W_0\left[n\frac{\ln(b)}{2\max\{b,\dim S\}}\right]\right\rfloor \geq 2, \quad d_A := \left\lfloor \frac{n-\dim S}{b}\right\rfloor \geq 2$$

Pick a full-rank function $\varphi_A \in V_{b,d_A,S}$ such that $r_A^2 := r_{\max}^2(\varphi_A) = b^{2\left\lfloor \frac{d_A}{2}\right\rfloor}$. Then,

$$\operatorname{compl}(\varphi_A) \leq bd_Ab^{d_A} + b^{\frac{d_A}{2}}\dim S \leq 2\max\{b,\dim S\}\, d_Ab^{d_A} \leq n,$$

by the choice of d_A and the properties of the Lambert W function.

Pick any $\varphi_B \in V_{b,d_B,S}$ with $r_B := r_{\max}(\varphi_B) = 1$ and $d_B = d_A$, so that $\operatorname{compl}(\varphi_B) = bd_B + \dim S \leq n$. Then, $\varphi_A,\ \varphi_B \in \Phi_n$. On the other hand, $r_A \geq r_B$ and from [39] we can estimate the Lambert W function from below as

$$W_0\left[n\frac{\ln(b)}{2\max\{b,\dim S\}}\right] \geq \ln\left[n\frac{\ln(b)}{2\max\{b,\dim S\}}\right] - \ln\ln\left[n\frac{\ln(b)}{2\max\{b,\dim S\}}\right]$$

Then

$$\begin{aligned} \operatorname{compl}(\varphi_A+\varphi_B) &\geq bd_Ar_A^2 + r_A\dim S \\ &\geq b\left(\frac{n-\dim S}{b}-1\right)\left(b^{\frac{1}{\ln(b)}W_0\left[n\frac{\ln(b)}{2\max\{b,\dim S\}}\right]-1}\right) \\ &= \left(\frac{n-\dim S}{b}-1\right)\left(n\frac{\ln(b)}{2\max\{b,\dim S\}}\right)\left[\ln\left(n\frac{\ln(b)}{2\max\{b,\dim S\}}\right)\right]^{-1}, \end{aligned}$$

The leading term in the latter expression is

$$\frac{\ln(b)}{2b\max\{b,\dim S\}}n^2\left[\ln\left(n\frac{\ln(b)}{2\max\{b,\dim S\}}\right)\right]^{-1}.$$

This cannot be bounded by cn for any $c > 0$ and thus (i) follows. □

Proof of Lemma 3.5. Let φ_A, $\varphi_B \in \Phi_n^{\mathcal{N}}$ with $d_A := d(\varphi_A)$, $d_B := d(\varphi_B)$, $\boldsymbol{r}^A := \boldsymbol{r}^A(\varphi_A)$, $\boldsymbol{r}^B := \boldsymbol{r}^B(\varphi_B)$ and w.l.o.g. $d_A \leq d_B$. Then using Lemma 2.23,

$$\begin{aligned}\operatorname{compl}_{\mathcal{N}}(\varphi_A + \varphi_B) &\leq \sum_{\nu=1}^{d_B} (r_\nu^A + r_\nu^B) \leq \sum_{\nu=1}^{d_A} r_\nu^A + (d_B - d_A)\dim S + \sum_{\nu=1}^{d_B} r_\nu^B \\ &\leq \operatorname{compl}_{\mathcal{N}}(\varphi_A) + \operatorname{compl}_{\mathcal{N}}(\varphi_B)(1+\dim S) \leq (2+\dim S)n.\end{aligned}$$

□

Proof of Lemma 3.8. Let φ_A, $\varphi_B \in \Phi_n^{\mathcal{C}}$ with $d_A := d(\varphi_A)$, $d_B := d(\varphi_B)$, $\boldsymbol{r}^A := \boldsymbol{r}^A(\varphi_A)$, $\boldsymbol{r}^B := \boldsymbol{r}^B(\varphi_B)$ and w.l.o.g. $d_A \leq d_B$. Then

$$\begin{aligned}\operatorname{compl}_{\mathcal{C}}(\varphi_A + \varphi_B) &\leq b(r_1^A + r_1^B) + \sum_{k=2}^{d_B} b(r_{k-1}^A + r_{k-1}^B)(r_k^A + r_k^B) + (r_{d_B}^A + r_{d_B}^B)\dim S \\ &= \underbrace{br_1^A + \sum_{k=2}^{d_A} br_{k-1}^A r_k^A + r_{d_A}^A \dim S}_{N_1} + \underbrace{br_1^B + \sum_{k=2}^{d_B} br_{k-1}^B r_k^B + br_{d_B}^B \dim S}_{N_2} \\ &+ \underbrace{\sum_{k=1}^{d_A} br_{k-1}^A r_k^B + br_{k-1}^B r_k^A}_{N_3} \\ &+ \underbrace{\sum_{k=d_A+1}^{d_B} br_{k-1}^A r_k^A}_{N_4} + \underbrace{(r_{d_B}^A - r_{d_A}^A)\dim S}_{N_5} + \underbrace{\sum_{k=d_A+1}^{d_B} br_{k-1}^A r_k^B + br_{k-1}^B r_k^A}_{N_6}.\end{aligned}$$

Since φ_A, $\varphi_B \in \Phi_n^{\mathcal{C}}$, we have $N_1 = \operatorname{compl}_{\mathcal{C}}(\varphi_A) \leq n$ and $N_2 = \operatorname{compl}_{\mathcal{C}}(\varphi_B) \leq n$. Then, using Lemma 2.12, we have

$$\begin{aligned}N_3 \leq& b\Big(\sum_{k=2}^{d_A}(r_{k-1}^A)^2\Big)^{1/2}\Big(\sum_{k=2}^{d_A}(r_k^B)^2\Big)^{1/2} + b\Big(\sum_{k=2}^{d_A}(r_{k-1}^B)^2\Big)^{1/2}\Big(\sum_{k=2}^{d_A}(r_k^A)^2\Big)^{1/2} \\ \leq& b\Big(\sum_{k=2}^{d_A} br_{k-1}^A r_k^A\Big)^{1/2}\Big(\sum_{k=2}^{d_A} br_{k-1}^B r_k^B)^2\Big)^{1/2} + b\Big(\sum_{k=2}^{d_A} br_{k-1}^B r_k^B\Big)^{1/2}\Big(\sum_{k=2}^{d_A} br_{k-1}^A r_k^A\Big)^{1/2} \\ \leq& 2b\operatorname{compl}_{\mathcal{C}}(\varphi_A)^{1/2}\operatorname{compl}_{\mathcal{C}}(\varphi_B)^{1/2} \leq 2bn.\end{aligned}$$

If $d_A = d_B$, we have $N_4 = N_5 = N_6 = 0$. If $d_A < d_B$, using Lemma 2.23, we have

$$\begin{aligned}N_4 &\leq (\dim S)^2 b(d_B - d_A) \leq (\dim S)^2 \operatorname{compl}_{\mathcal{C}}(\varphi_B) \leq n(\dim S)^2, \\ N_5 &\leq (\dim S)^2 \leq (\dim S)\operatorname{compl}_{\mathcal{C}}(\varphi_A) \leq n\dim S, \\ N_6 &\leq (\dim S)\Big(\sum_{k=d_A+1}^{d_B} br_k^B + br_{k-1}^B\Big) \leq 2(\dim S)\operatorname{compl}_{\mathcal{C}}(\varphi_B) \leq 2n\dim S.\end{aligned}$$

Thus, putting all together

$$\operatorname{compl}_{\mathcal{C}}(\varphi_A + \varphi_B) \leq [(\dim S)^2 + 3\dim S + 2b + 2]n,$$

and (P4) is satisfied with $c := (\dim S)^2 + 3\dim S + 2b + 2$. □

Proof of Lemma 3.11. We have the representation

$$T_{b,d}(\varphi)(i_1,\dots,i_d,y) = \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\dim S} v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)v_{d+1}^{k_d,q}\varphi_q(y).$$

Then, from Lemma 2.6,

$$\begin{aligned}
&T_{b,\bar d}(\varphi)(i_1,\dots,i_{\bar d},y)\\
&= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\dim S} v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)v_{d+1}^{k_d,q}T_{b,\bar d-d}(\varphi_q)(i_{d+1},\dots,i_{\bar d},y)\\
&= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\dim S}\sum_{j_{d+1}=1}^{b}\\
&\quad v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)\underbrace{v_{d+1}^{k_d,q}\delta_{j_{d+1}}(i_{d+1})}_{\bar v_{d+1}^{k_d,(q,j_{d+1})}(i_{d+1})}T_{b,\bar d-d}(\varphi_q)(j_{d+1},i_{d+2},\dots,i_{\bar d},y),
\end{aligned}$$

where $\bar v_{d+1}\in\mathbb{R}^{b\times r_d\times(b\dim S)}$. Since S is closed under b-adic dilation, we know from Lemma 2.23 that $r_\nu(\varphi_q)\le\dim S$ for all $\nu\in\mathbb{N}$. Let $l=\bar d-d$ and first assume $l\ge 2$. Then, $T_{b,\bar d-d}(\varphi_q)$ admits a representation

$$\begin{aligned}
&T_{b,\bar d-d}(\varphi_q)(j_{d+1},i_{d+2},\dots,i_{\bar d},y)\\
&= \sum_{\alpha_1=1}^{\dim S}\dots\sum_{\alpha_l=1}^{\dim S}\sum_{p=1}^{\dim S} w_1^{q,\alpha_1}(j_{d+1})w_2^{q,\alpha_1,\alpha_2}(i_{d+2})\dots w_l^{q,\alpha_{l-1},\alpha_l}(i_{\bar d})w_{l+1}^{q,\alpha_l,p}\varphi_p(y)\\
&= \sum_{\alpha_2=1}^{\dim S}\dots\sum_{\alpha_l=1}^{\dim S}\sum_{p=1}^{\dim S} w_{1,2}^{q,\alpha_2}(j_{d+1},i_{d+2})\dots w_l^{q,\alpha_{l-1},\alpha_l}(i_{\bar d})w_{l+1}^{q,\alpha_l,p}\varphi_p(y)
\end{aligned}$$

with $w_{1,2}^{q,\alpha_2}(j_{d+1},i_{d+2})=\sum_{\alpha_1=1}^{\dim S}w_1^{q,\alpha_1}(j_{d+1})w_2^{q,\alpha_1,\alpha_2}(i_{d+2})$. Then,

$$\begin{aligned}
&T_{b,\bar d-d}(\varphi_q)(j_{d+1},i_{d+2},\dots,i_{\bar d},y)\\
&= \sum_{\alpha_2,q_2=1}^{\dim S}\dots\sum_{\alpha_l,q_l=1}^{\dim S}\sum_{p=1}^{\dim S}\underbrace{\delta_{q,q_2}w_{1,2}^{q,\alpha_2}(j_{d+1},i_{d+2})}_{\bar v_{d+2}^{(q,j_{d+1}),(q_2,\alpha_2)}(i_{d+2})}\dots\underbrace{\delta_{q_{l-1},q_l}w_l^{q_{l-1},\alpha_{l-1},\alpha_l}(i_{\bar d})}_{\bar v_{\bar d}^{(q_{l-1},\alpha_{l-1}),(q_l,\alpha_l)}(i_{\bar d})}\underbrace{w_{l+1}^{q_l,\alpha_l,p}}_{\bar v_{\bar d+1}^{(q_l,\alpha_l),p}}\varphi_p(y)
\end{aligned}$$

with $\bar v_{d+2}\in\mathbb{R}^{b\times(b\dim S)\times(\dim S)^2}$, $\bar v_\nu\in\mathbb{R}^{b\times(\dim S)^2\times(\dim S)^2}$ for $d+3\le\nu\le\bar d$, and $\bar v_{\bar d+1}\in\mathbb{R}^{(\dim S)^2\times\dim S}$. Then, we have $\varphi\in\mathcal{R}_{b,\bar d,S,\bar{\boldsymbol r}}(\bar{\mathbf v})$ with $\bar{\mathbf v}=(\bar v_1,\dots,v_d,\bar v_{d+1},\dots,\bar v_{\bar d+1})$, with $\bar v_\nu$ defined above for $\nu>d$, and $\bar r_\nu=r_\nu$ for $\nu\le d$, $\bar r_{d+1}=b\dim S$, and $\bar r_\nu=(\dim S)^2$ for $d+1<\nu\le\bar d$. From the definition of $\bar v_\nu$, we easily deduce that $\|\bar v_{d+1}\|_{\ell_0}=b\|v_{d+1}\|_{\ell_0}$, $\|\bar v_{d+2}\|_{\ell^0}\le b^2(\dim S)^2$, $\|\bar v_\nu\|_{\ell_0}\le b(\dim S)^3$ for $d+3\le\nu\le\bar d$, and $\|\bar v_{\bar d+1}\|_{\ell_0}\le(\dim S)^3$. Then, for $l=\bar d-d\ge 2$, we obtain

$$\begin{aligned}
\mathrm{compl}_{\mathcal S}(\bar{\mathbf v}) &= \sum_{\nu=1}^{d+1}\|\bar v_\nu\|_{\ell_0}+\sum_{\nu=d+2}^{\bar d+1}\|\bar v_\nu\|_{\ell_0}\\
&\le b\,\mathrm{compl}_{\mathcal S}(\mathbf v)+b^2(\dim S)^2+b(\dim S)^3(\bar d-d-2)+(\dim S)^3.
\end{aligned}$$

For $l=1$, we have a representation

$$T_{b,\bar d-d}(\varphi_q)(j_{d+1},y)=\sum_{p=1}^{\dim S}\bar v_{d+2}^{(q,j_{d+1}),p}\varphi_p(y)$$

with $\bar{v}_{d+2} \in \mathbb{R}^{(b \dim S) \times \dim S}$ such that $\bar{v}_{d+2}^{(q,j_{d+1}),p} = \sum_{\alpha_1=1}^{\dim S} w_1^{q,\alpha_1}(j_{d+1}) \varphi_2^{q,\alpha_1,p}$. Then, for $l = 1$, $\varphi \in \mathcal{R}_{b,\bar{d},S,\overline{\boldsymbol{r}}}(\overline{\mathbf{v}})$ with $\overline{\mathbf{v}} = (\bar{v}_1, \ldots, v_d, \bar{v}_{d+1}, \bar{v}_{\bar{d}+2})$, and

$$\mathrm{compl}_{\mathcal{S}}(\overline{\mathbf{v}}) \leq b\, \mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b(\dim S)^2 = b\, \mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b(\dim S)^2(\bar{d} - d).$$

For any $l \geq 1$, we then deduce

$$\mathrm{compl}_{\mathcal{S}}(\overline{\mathbf{v}}) \leq b\, \mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b^2(\dim S)^3(\bar{d} - d).$$

□

Proof of Lemma 3.12. φ_A and φ_B admit representations

$$T_{b,d}(\varphi_C)(i_1, \ldots, i_d, y) = \sum_{k_1=1}^{r_1^C} \cdots \sum_{k_d=1}^{r_d^C} \sum_{q=1}^{\dim S} v_1^{C,k_1}(i_1) \cdots v_d^{C,k_{d-1},k_d}(i_d) v_{d+1}^{C,k_d,q} \varphi_q(y),$$

with $C = A$ or B. Then, $\varphi_A + \varphi_B$ admit the representation

$$T_{b,d}(\varphi_A + \varphi_B)(i_1, \ldots, i_d, y) = \sum_{k_1=1}^{r_1^A + r_1^B} \cdots \sum_{k_d=1}^{r_d^A + r_d^B} \sum_{q=1}^{\dim S} v_1^{k_1}(i_1) \cdots v_d^{k_{d-1},k_d}(i_d) v_{d+1}^{k_d,q} \varphi_q(y),$$

with $v_1^{k_1} = v_1^{A,k_1}$ if $1 \leq k_1 \leq r_1^A$ and $v_1^{k_1} = v_1^{B,k_1}$ if $r_1^A < k_1 \leq r_1^A + r_1^B$,

$$v_\nu^{k_{\nu-1},k_\nu} = \begin{cases} v^{A,k_{\nu-1},k_\nu} & \text{if } 1 \leq k_{\nu-1}, k_\nu \leq r_1^A \\ v^{B,k_{\nu-1},k_\nu} & \text{if } r_1^A < k_{\nu-1}, k_\nu \leq r_1^A + r_1^B \\ 0 & \text{elsewhere,} \end{cases}$$

and $v_{d+1}^{k_d,q} = v_{d+1}^{A,k_d,q}$ if $1 \leq k_d \leq r_1^A$ and $v_{d+1}^{k_d,q} = v_{d+1}^{B,k_d,q}$ if $r_1^A < k_d \leq r_1^A + r_1^B$. From the above, we deduce that $\|v^C\|_{\ell_0} \leq \|v^A\|_{\ell_0} + \|v^B\|_{\ell_0}$, so that

$$\mathrm{compl}_{\mathcal{S}}(\mathbf{v}) = \sum_{\nu=1}^{d+1} \|v^C\|_{\ell_0} \leq \mathrm{compl}_{\mathcal{S}}(\mathbf{v}) \leq \mathrm{compl}_{\mathcal{S}}(\mathbf{v}_A) + \mathrm{compl}_{\mathcal{S}}(\mathbf{v}_B).$$

□

Proof of Lemma 3.14. For $1 < p < \infty$, the norm defined in Theorem 2.15 is a reasonable crossnorm (see Lemma B.1) and thus, in particular, not weaker than the injective norm on $V_{b,d,S}$. Thus, by [34, Lemma 8.6] $\mathcal{TT}_{\boldsymbol{r}}(V_{b,d,S})$ for $\boldsymbol{r} \in \mathbb{N}^d$ is a weakly closed subset of L^p. Moreover, the set Φ_n, with either $\Phi_n = \Phi_n^{\mathcal{N}}$ or $\Phi_n^{\mathcal{C}}$, is a finite union of the sets $\mathcal{TT}_{\boldsymbol{r}}(V_{b,d,S})$ for different $d \in \mathbb{N}$ and $\boldsymbol{r} \in \mathbb{N}^d$. Since finite unions of closed sets (in the weak topology) are closed, it follows that Φ_n is weakly closed in L^p, and a fortiori, Φ_n is also closed in the strong topology. Since L^p is reflexive for $1 < p < \infty$ and Φ_n is weakly closed, Φ_n is proximinal in L^p (see [34, Theorem 4.28]).

Now consider that S is finite-dimensional and $0 < p \leq \infty$. There exists d such that $\Phi_n \subset V_{b,d,S}$ and $V_{b,d,S}$ is finite-dimensional. Since Φ_n is a closed subset of a finite-dimensional space $V_{b,d,S}$, it is proximinal in L^p for any $0 < p \leq \infty$. □

Proof of Proposition 3.18. Consider a function $0 \neq \varphi \in V_{b,S}$ and let $d = d(\varphi)$ and $\boldsymbol{r} = \boldsymbol{r}(\varphi)$. We have

$$\mathrm{compl}_{\mathcal{N}}(\varphi) = \sum_{\nu=1}^{d} r_\nu \leq b r_1 + \sum_{r_\nu=2}^{d} b r_{\nu-1} r_\nu + b \dim S = \mathrm{compl}_{\mathcal{C}}(\varphi),$$

which implies $\Phi_n^{\mathcal{C}} \subset \Phi_n^{\mathcal{N}}$. Also

$$\mathrm{compl}_{\mathcal{C}}(\varphi) \leq br_1 + b(\sum_{\nu=1}^{d-1} r_\nu^2)^{1/2}(\sum_{\nu=2}^{d} r_\nu^2)^{1/2} + b\dim S \leq br_1 + b(\sum_{\nu=1}^{d-1} r_\nu)(\sum_{\nu=2}^{d} r_\nu) + b\dim S$$

$$\leq b(\sum_{\nu=1}^{d} r_\nu)^2 + b\dim S = b\,\mathrm{compl}_{\mathcal{N}}(\varphi)^2 + b\dim S,$$

which yields $\Phi_n^{\mathcal{N}} \subset \Phi_{b\dim S + bn^2}^{\mathcal{C}}$. Also, we clearly have $\mathrm{compl}_{\mathcal{S}}(\varphi) \leq \mathrm{compl}_{\mathcal{C}}(\varphi)$, which implies $\Phi_n^{\mathcal{C}} \subset \Phi_n^{\mathcal{S}}$. Now consider any tensor network $\mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}$ such that $\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v})$, with $d(\varphi) \leq d$ and $\boldsymbol{r}(\varphi) \leq \boldsymbol{r}$. We have that $r_1(\varphi) \leq \dim\{v_1^{k_1}(\cdot) \in \mathbb{R}^b : 1 \leq k_1 \leq r_1\} \leq \|v_1\|_{\ell_0}$ and for $2 \leq \nu \leq d$, $r_\nu(\varphi) \leq \dim\{v_\nu^{:,k_\nu}(\cdot) \in \mathbb{R}^{b\times r_{\nu-1}} : 1 \leq k_\nu \leq r_\nu\} \leq \|v_\nu\|_{\ell^0}$. Therefore

$$\mathrm{compl}_{\mathcal{N}}(\varphi) = \sum_{\nu=1}^{d} r_\nu(\varphi) \leq \sum_{\nu=1}^{d} \|v_\nu\|_{\ell^0} \leq \mathrm{compl}_{\mathcal{S}}(\mathbf{v}).$$

The inequality being true for any tensor network $\mathbf{v}$ such that $\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v})$, we deduce $\mathrm{compl}_{\mathcal{N}}(\varphi) \leq \mathrm{compl}_{\mathcal{S}}(\varphi)$, which yields $\Phi_n^{\mathcal{S}} \subset \Phi_n^{\mathcal{N}}$. ☐

Proof of Lemma 3.21. (i). Consider $\varphi_A, \varphi_B \in \Phi_n^{\mathcal{R}}$, and let $d_A = d(\varphi_A)$, $d_B = d(\varphi_B)$, $r_A = r(\varphi_A)$ and $r_B = r(\varphi_B)$. Assume w.l.o.g. that $d_A \leq d_B$. The function φ_A admits a representation

$$T_{b,d_A}\varphi_A(i_1, \ldots, i_{d_A}, y) = \sum_{k=1}^{r_A} w_1^{A,k}(i_1) \ldots w_d^{A,k}(i_d) w_{d+1}^{A,k}(y),$$

and

$$T_{b,d_B}\varphi_A(i_1, \ldots, i_{d_B}, y) = \sum_{k=1}^{r_A} w_1^{A,k}(i_1) \ldots w_d^{A,k}(i_d) T_{b,d_B}(w_{d+1}^{A,k})(y).$$

From the assumption on S, we have $T_{b,d_B}(w_{d+1}^{A,k})$ of rank 1, so that $r(T_{b,d_B}\varphi_A) \leq r_A$. We easily deduce that $r(\varphi_A + \varphi_B) \leq r_A + r_B$ and $\mathrm{compl}_{\mathcal{R}}(\varphi_A + \varphi_B) \leq bd_B(r_A + r_B) + (r_A + r_B)b\dim S \leq 2n + br_A(d_B - d_A) \leq 2n + n^2 \leq 3n^2$.
(ii). The proof idea is analogous to Proposition 3.4: we take a rank-one tensor $\varphi_B \in \Phi_n^{\mathcal{R}}$ such that $d_B \sim n$ and a full-rank tensor $\varphi_A \in \Phi_n^{\mathcal{R}}$ with $d_A < d_B$ such that $r_A \sim b^{d_A} \sim n$. Then, as in Proposition 3.4, $\mathrm{compl}_{\mathcal{R}}(\varphi_A + \varphi_B) \sim n^2$. ☐

Proof of Lemma 3.22. Let $\varphi \in \Phi_n^{\mathcal{R}}$, $d = d(\varphi)$, $r = r(\varphi)$. The function φ admits a representation

$$T_{b,d}\varphi(i_1, \ldots, i_d, y) = \sum_{k=1}^{r} \sum_{q=1}^{\dim S} w_1^k(i_1) \ldots w_d^k(i_d) w_{d+1}^{q,k}\varphi_q(y).$$

Letting $v_1 = w_1$, $v_{d+1} = w_{d+1}$ and $v_\nu \in \mathbb{R}^{b\times r\times r}$ such that $v_\nu^{k_{\nu-1},k_\nu} = \delta_{k_{\nu-1},k_\nu} w_\nu^{k_\nu}$ for $2 \leq \nu \leq d$, and letting $\boldsymbol{r} = (r, \ldots, r) \in \mathbb{N}^d$, we have

$$T_{b,d}\varphi(i_1, \ldots, i_d, y) = \sum_{k_1=1}^{r} \cdots \sum_{k_d=1}^{r} \sum_{q=1}^{\dim S} v_1^k(i_1) \ldots w_d^{k_{d-1},k_d}(i_d) w_{d+1}^{q,k_d}\varphi_q(y),$$

which proves that $T_{b,d}\varphi \in \Phi_{b,d,S,\boldsymbol{r}}$ with

$$\mathrm{compl}_{\mathcal{S}}(\varphi) = \sum_{\nu=1}^{d+1} \|v_\nu\|_{\ell_0} = \sum_{\nu=1}^{d+1} \|w_\nu\|_{\ell_0} \leq brd + r\dim S = \mathrm{compl}_{\mathcal{R}}(\varphi) \leq n,$$

that is $\varphi \in \Phi_n^{\mathcal{S}}$. ☐

Appendix D. Proofs for Section 4

Proof of Proposition 4.13. The bounds for $\mathrm{compl}_{\mathcal{N}}(\varphi)$ and $\mathrm{compl}_{\mathcal{C}}(\varphi)$ directly follow from Lemma 4.12. To obtain the bound on the sparse representation complexity, we have to provide a representation of φ in a tensor format. First, we note that the interval $I_k = [x^b_{k-1}, x^b_k)$ is such that $I_k = \cup_{i=1}^{n_k} I_{k,i}$, where the $I_{k,i}$ are n_k contiguous intervals from b-adic partitions of $[0,1)$, and the minimal n_k can be bounded as $n_k \leq 2d(b-1)$. To illustrate why this bound holds, we refer to Figure 6.

If d is the maximal level, the subsequent partitioning of $[0,1)$ for levels $l = 0,1,2,\ldots,d$ can be represented as a tree, where each vertex has b sons, i.e., each interval is subsequently split into b intervals. Then, the end-points x^b_{k-1} and x^b_k of an arbitrary interval I_k correspond to two points in this interval partition tree. The task of finding a minimal sub-partitioning $I_k = \cup_{i=1}^{n_k} I_{k,i}$ is then equivalent to finding the shortest path in this tree, and $2d$ represents the longest possible path.

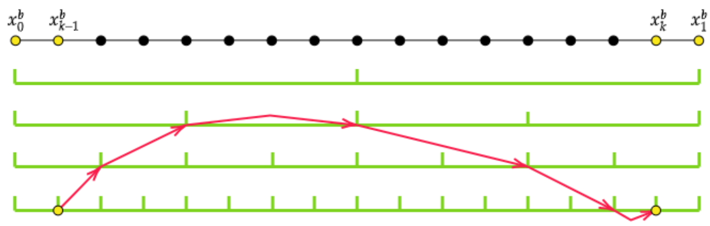


Figure 6. Visual representation of different partitioning levels of the interval $[0,1)$, with $b=2$ and $d=4$.

In Figure 6, we depict a scenario close to the "worst case". In order to reach vertex x^b_k from vertex x^b_{k-1}, at most, we would have to traverse the tree up (towards the root) and back down. On each level, we would need at most $b-1$ horizontal steps. Thus, we require at most $2d(b-1)$ steps to reach x^b_k.

Then, φ admits a representation as $\varphi = \sum_{k=1}^N \sum_{i=1}^{n_k} s_{i,k}$, with $s_{i,k}$ supported on $I_{i,k}$ and polynomial on this interval. Let $\lambda := (k,i)$. We have $I_\lambda = [b^{d_\lambda} j_\lambda, b^{d_\lambda}(j_\lambda + 1))$ for some $d_\lambda \leq d$ and $j_\lambda \in \{0,\ldots,b^{d_\lambda - 1} - 1\}$. By denoting $(j_{\lambda,1},\ldots,j_{\lambda,d_\lambda})$ the representation of j_λ in base b, s_λ admits a tensorization

$$T_{b,d_\lambda}(s_\lambda) = \delta_{j_{\lambda,1}} \otimes \ldots \otimes \delta_{j_{\lambda,d_\lambda}} \otimes p_\lambda,$$

with $p_\lambda \in \mathbb{P}_m$, so that $\mathrm{compl}_{\mathcal{S}}(s_\lambda) \leq d_\lambda + \dim S$. From Lemmas 3.11 and 3.12, we deduce that

$$\begin{aligned}\mathrm{compl}_{\mathcal{S}}(\varphi) &\leq \sum_{k=1}^N \sum_{i=1}^{n_k} b(d_{k,i} + \dim S) + b^2(\dim S)^3(d - d_{k,i}) \leq 2b^2(\dim S)^3 d \sum_{k=1}^N n_k \\ &\leq 4b^3(m+1)^3 d^2 N.\end{aligned}$$

□

Proof of Proposition 4.15. From Lemma 4.14, we have

$$\begin{aligned}\mathrm{compl}_{\mathcal{N}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) &\leq d(\bar{m} + N) + (\bar{d} - d)(\bar{m} + N) \leq (\bar{m}+1)dN + (\bar{d} - d)(\bar{m} + N), \\ \mathrm{compl}_{\mathcal{C}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) &\leq b(\bar{m} + N) + (d-1)b(\bar{m}+N)^2 + (\bar{d} - d)b(\bar{m}+1)^2 + b(m+1) \\ &\leq 2bd(\bar{m}+1)^2 dN^2 + (\bar{d} - d)b(\bar{m}+1)^2.\end{aligned}$$

Now we consider the sparse representation complexity. The function φ admits a representation $\varphi = R_{b,d,\bar{m},\boldsymbol{r}}(\mathbf{v})$ for some $\boldsymbol{r} \in \mathbb{N}^r$ and a tensor network $\mathbf{v} \in \mathcal{P}_{b,d,\bar{m},\boldsymbol{r}}$ such that $\mathrm{compl}_{\mathcal{S}}(\mathbf{v}) = \mathrm{compl}_{\mathcal{S}}(\varphi)$ and

$$T_{b,d}(\varphi)(i_1,\dots,i_d,y) = \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\bar{m}+1} v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)v_{d+1}^{k_d,q}\varphi_q^{\bar{m}+1}(y),$$

with the $\varphi_q^{\bar{m}+1}$ forming a basis of $\mathbb{P}_{\bar{m}}$. From Proposition 4.8, we know that $\mathrm{compl}_{\mathcal{S}}(\varphi) \leq C_1 N$ for some constant C_1 depending only on b and $\bar{m}$. Then from (2.13), we have that

$$\begin{aligned}
&T_{b,\bar{d}}(\mathcal{I}_{b,\bar{d},m}(\varphi))(i_1,\dots,i_{\bar{d}},y)\\
&= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\bar{m}+1} v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)v_{d+1}^{k_d,q}T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1}))(i_{d+1},\dots,i_{\bar{d}},y)\\
&= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\bar{m}+1}\sum_{j_{d+1}=1}^{b}\\
&v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)\bar{v}_{d+1}^{k_d,(q,j_{q+1})}(i_{d+1})T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1}))(j_{d+1},\dots,i_{\bar{d}},y),
\end{aligned}$$

with $\bar{v}_{d+1}^{k_d,(q,j_{q+1})}(i_{d+1}) = v_{d+1}^{k_d,q}\delta_{j_{d+1}}(i_{d+1})$ such that $\|\bar{v}_{d+1}\|_{\ell^0} = b\|v_{d+1}\|_{\ell^0}$. Noting that $r_\nu(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1})) \leq r_\nu(\varphi_q^{\bar{m}+1}) \leq \bar{m}+1$ for all $\nu \in \mathbb{N}$, and following the proof of Lemma 3.11, we can prove that for $\bar{d}-d \geq 2$, $\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1})$ admits a representation

$$\begin{aligned}
&T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q))(j_{d+1},i_{d+2},\dots,i_{\bar{d}},y)\\
&= \sum_{\alpha_2,q_2=1}^{\bar{m}+1}\cdots\sum_{\alpha_l,q_l=1}^{\bar{m}+1}\sum_{p=1}^{m+1}\bar{v}_{d+2}^{(q,j_{d+1}),(q_2,\alpha_2)}(i_{d+2})\dots\bar{v}_{\bar{d}}^{(q_{l-1},\alpha_{l-1}),(q_l,\alpha_l)}(i_{\bar{d}})\bar{v}_{\bar{d}+1}^{(q_l,\alpha_l),p}\varphi_p(y)
\end{aligned}$$

with the φ_p forming a basis of $\mathbb{P}_m$ and with $\bar{v}_{d+2} \in \mathbb{R}^{b\times(b(\bar{m}+1))\times(\bar{m}+1)^2}$, $\bar{v}_\nu \in \mathbb{R}^{b\times(\bar{m}+1)^2\times(\bar{m}+1)^2}$ for $d+3 \leq \nu \leq \bar{d}$, and $\bar{v}_{\bar{d}+1} \in \mathbb{R}^{(\bar{m}+1)^2\times(m+1)}$. Then, we have $\mathcal{I}_{b,\bar{d},m}(\varphi) = R_{b,\bar{d},m,\bar{\boldsymbol{r}}}(\overline{\mathbf{v}})$ with $\overline{\mathbf{v}} = (\bar{v}_1,\dots,v_d,\bar{v}_{d+1},\dots,\bar{v}_{\bar{d}+1})$ such that

$$\begin{aligned}
\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) &\leq b\,\mathrm{compl}_{\mathcal{S}}(\varphi) + b^2(\bar{m}+1)^3 + b(\bar{m}+1)^4(\bar{d}-d-2) + (\bar{m}+1)^2(m+1)\\
&\leq \max\{b,m+1\}(\mathrm{compl}_{\mathcal{S}}(\varphi) + b(\bar{m}+1)^3\\
&+ (\bar{m}+1)^4(\bar{d}-d-2) + (\bar{m}+1)^2)\\
&\leq \max\{b,m+1\}(\mathrm{compl}_{\mathcal{S}}(\varphi) + b(\bar{m}+1)^4(\bar{d}-d)).
\end{aligned}$$

For $\bar{d}-d = 1$, we have the representation

$$T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1}))(j_{d+1},y) = \sum_{p=1}^{m+1}\bar{v}_{d+2}^{(q,j_{d+1}),p}\varphi_p(y)$$

with some $\bar{v}_{d+2} \in \mathbb{R}^{(b(\bar{m}+1))\times(m+1)}$. Then for $\bar{d}-d = 1$, $\varphi \in R_{b,\bar{d},S,\bar{\boldsymbol{r}}}(\overline{\mathbf{v}})$ with $\overline{\mathbf{v}} = (\bar{v}_1,\dots,v_d,\bar{v}_{d+1},\bar{v}_{\bar{d}+2})$, and

$$\begin{aligned}
\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) &\leq b\,\mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b(\bar{m}+1)(m+1)\\
&\leq \max\{b,m+1\}(\mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b(\bar{m}+1)(\bar{d}-d)).
\end{aligned}$$

Finally for $\bar{d} = d$, we simply have $\mathcal{I}_{b,\bar{d}-d,m} = \mathcal{I}_m$, and we can show that $\mathcal{I}_{b,\bar{d},m}(\varphi) = R_{b,d,m,\boldsymbol{r}}(v_1,...,v_d,\bar{v}_{d+1})$ with $\|\bar{v}_{d+1}\|_{\ell^0} \leq (m+1)\|v_{d+1}\|_{\ell^0}$, so that

$$\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq (m+1)\,\mathrm{compl}_{\mathcal{S}}(\mathbf{v}).$$

Then for any $\bar{d} \geq d$, we have

$$\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq \max\{b, m+1\}(\mathrm{compl}_{\mathcal{S}}(\varphi) + b(\bar{m}+1)^4(\bar{d}-d)),$$

and we conclude by using $\mathrm{compl}_{\mathcal{S}}(\varphi) \leq C_1 N$. $\square$

Appendix E. Proofs for Section 5

Proof of Lemma 5.1. From (2.13), we know that $s := \mathcal{I}_{b,d,\bar{m}} f$ admits a tensorization $\boldsymbol{s} := T_{b,d} s = (id_{\{1,\dots,d\}} \otimes \mathcal{I}_{\bar{m}})\boldsymbol{f}$, with $\boldsymbol{f} = T_{b,d} f$ and where $id_{\{1,\dots,d\}} : (\mathbb{R}^{I_b})^{\otimes d} \to (\mathbb{R}^{I_b})^{\otimes d}$ is the identity. Then,

$$T_{b,d}(f - s) = \sum_{j \in I_b^d} \delta_{j_1} \otimes \ldots \otimes \delta_{j_d} \otimes (g_j - \mathcal{I}_{\bar{m}} g_j),$$

with $g_j = \boldsymbol{f}(j_1, \ldots, j_d, \cdot)$. Using the property (4.1) of operator $\mathcal{I}_{\bar{m}}$, with a constant C depending on $\bar{m}$ and p, we have

$$\|g_j - \mathcal{I}_{\bar{m}} g_j\|_p \leq C |g_j|_{W^{\bar{m}+1,p}} = C \|D^{\bar{m}+1} g_j\|_p.$$

Then, using Theorem 2.15, we have for $p < \infty$

$$\|f - s\|_p^p = \sum_{j \in I_b^d} b^{-d} \|g_j - \mathcal{I}_{\bar{m}} g_j\|_p^p \leq C^p \sum_{j \in I_b^d} b^{-d} \|D^{\bar{m}+1} g_j\|_p^p = C^p \|(id_{\{1,\dots,d\}} \otimes D^{\bar{m}+1})\boldsymbol{f}\|_p^p$$

and

$$\|f - s\|_\infty = \max_{j \in I_b^d} \|g_j - \mathcal{I}_{\bar{m}} g_j\|_\infty \leq C \max_{j \in I_b^d} \|D^{\bar{m}+1} g_j\|_\infty \leq C \|(id_{\{1,\dots,d\}} \otimes D^{\bar{m}+1})\boldsymbol{f}\|_\infty.$$

Then, from Theorem 2.15 we deduce

$$\|f - s\|_p \leq C \|(id_{\{1,\dots,d\}} \otimes D^{\bar{m}+1})\boldsymbol{f}\|_p = C b^{-d(\bar{m}+1)} \|f\|_{W^{\bar{m}+1,p}}.$$

For $\bar{d} \geq d$, we obtain from Lemma 2.6 and (2.13) that

$$\begin{aligned} T_{b,d} \mathcal{I}_{b,\bar{d},m} T_{b,d}^{-1} &= T_{b,d} T_{b,\bar{d}}^{-1} (id_{\{1,\dots,\bar{d}\}} \otimes \mathcal{I}_m) T_{b,\bar{d}} T_{b,d}^{-1} \\ &= id_{\{1,\dots,d\}} \otimes (T_{b,\bar{d}-d}(id_{\{1,\dots,\bar{d}-d\}} \otimes \mathcal{I}_m) T_{b,\bar{d}-d}^{-1}) = id_{\{1,\dots,d\}} \otimes \mathcal{I}_{b,\bar{d}-d,m}. \end{aligned}$$

Then $\tilde{s} := \mathcal{I}_{b,\bar{d},m} s$ admits for tensorization

$$T_{b,d} \tilde{s} = (id_{\{1,\dots,d\}} \otimes \mathcal{I}_{b,\bar{d}-d,m})\boldsymbol{s} = \sum_{j \in I_b^d} \delta_{j_1} \otimes \ldots \otimes \delta_d \otimes (\mathcal{I}_{b,\bar{d}-d,m} \mathcal{I}_{\bar{m}} g_j),$$

which yields

$$T_{b,d}(s - \tilde{s}) = \sum_{j \in I_b^d} \delta_{j_1} \otimes \ldots \otimes \delta_{j_d} \otimes (\mathcal{I}_{\bar{m}} g_j - \mathcal{I}_{b,\bar{d}-d,m} \mathcal{I}_{\bar{m}} g_j).$$

From the property (4.1) of $\mathcal{I}_m$, with a constant $\tilde{C}$ depending on m and p, and the property of $\mathcal{I}_{\bar{m}}$, we obtain

$$\begin{aligned} \|\mathcal{I}_{\bar{m}} g_j - \mathcal{I}_{b,\bar{d}-d,m} \mathcal{I}_{\bar{m}} g_j\|_p &\leq \tilde{C} b^{-(\bar{d}-d)(m+1)} |\mathcal{I}_{\bar{m}} g_j|_{W^{m+1,p}} \\ &\leq \tilde{C} b^{-(\bar{d}-d)(m+1)} \left(|g_j|_{W^{m+1,p}} + C \, |g_j|_{W^{\bar{m}+1,p}} \right). \end{aligned}$$

In the same way as above, we deduce

$$\begin{aligned} \|s - \tilde{s}\|_p &\leq \tilde{C} b^{-(\bar{d}-d)(m+1)} \left(\|(id_{\{1,\dots,d\}} \otimes D^{m+1})\boldsymbol{f}\|_p + C \|(id_{\{1,\dots,d\}} \otimes D^{\bar{m}+1})\boldsymbol{f}\|_p \right) \\ &= C' b^{-(\bar{d}-d)(m+1)} (b^{-d(m+1)} \, |f|_{W^{m+1,p}} + b^{-d(\bar{m}+1)} \, |f|_{W^{\bar{m}+1,p}}), \end{aligned}$$

with $C' = \tilde{C} \max\{1, C\}$ depending on m, $\bar{m}$ and p, which completes the proof. $\square$

Proof of Lemma 5.4. The proof is a modification of the proof of Petrushev for free knot splines (see [19, Chapter 12, Theorem 8.2]). The first step is the optimal selection of n intervals that, in a sense, balances out the Besov norm $|f|_{B^\alpha_{\tau,\tau}}$. In this step, unlike in the case of classic free knot splines, we are restricted to b-adic knots. The second step is a polynomial approximation over each interval and is essentially the same as with free knot splines. We demonstrate this step here as well for completeness.

First, we define a set function that we will use for the selection of the $n-1$ b-adic knots. Let $k := \lfloor\alpha\rfloor + 1$ and

$$M^\tau := \int_0^1 t^{-\alpha\tau-1} \mathrm{w}_k(f,t)_\tau^\tau \, \mathrm{d}t,$$

where w_k is the *averaged modulus of smoothness*, i.e.,

$$\mathrm{w}_k(f,t)_\tau^\tau := \frac{1}{t}\int_0^t \left\|\Delta_h^k[f]\right\|_\tau^\tau \, \mathrm{d}h.$$

By [19, Chapters 2 and 12], M is equivalent to $|f|_{B^\alpha_{\tau,\tau}}$.

Let

$$g(x,h,t) := \begin{cases} t^{-\alpha\tau-2}|\Delta_h^k[f](x)|^\tau & \text{if } h\in[0,t] \text{ and } x\in[0,1-kh], \\ 0 & \text{elsewhere.} \end{cases}$$

Then,

$$M^\tau = \int_0^1\int_0^\infty\int_0^1 g(x,h,t)\,\mathrm{d}x\,\mathrm{d}h\,\mathrm{d}t = \int_0^1 G(x)\,\mathrm{d}x,$$
$$G(x) := \int_0^\infty\int_0^1 g(x,h,t)\,\mathrm{d}h\,\mathrm{d}t.$$

The aforementioned set function is then defined as

$$\Omega(t) := \int_0^t G(x)\,\mathrm{d}x.$$

This function is positive, continuous and monotonically increasing with

$$\Omega(0) = 0 \quad \text{and} \quad \Omega(1) = M^\tau \sim |f|^\tau_{B^\alpha_{\tau,\tau}}.$$

Thus, we can pick N intervals I_i, $i=1,\ldots,N$, with disjoint interiors such that

$$\bigcup_{i=1}^N I_i = [0,1] \quad \text{and} \quad \int_{I_i} G(x)\,\mathrm{d}x = \frac{M^\tau}{N}.$$

This would have been the optimal knot selection for free knot splines. For our purposes we need to restrict the intervals to b-adic knots. More precisely, we show that with restricted intervals we can get arbitrarily close to the optimal choice.

Let $\varepsilon > 0$ be arbitrary. Starting with $i=1$, due to the properties of the function $\Omega(\cdot)$, we can pick a b-adic interval I_1^ε with left end point 0 such that

$$\int_{I_1^\varepsilon} G(x)\,\mathrm{d}x \le \frac{M^\tau}{N} \le \int_{I_1^\varepsilon} G(x)\,\mathrm{d}x + \frac{\varepsilon}{N}. \tag{E.1}$$

For I_2^ε, we set the left endpoint equal to the right endpoint of I_1^ε and choose the right endpoint of I_2^ε as a b-adic knot such that (E.1) is satisfied for I_2^ε. Repeating this procedure until I_{N-1}^ε we get

$$\int_{\bigcup_{i=1}^{N-1} I_k^\varepsilon} G(x)\,\mathrm{d}x \le \frac{N-1}{N}M^\tau \le \int_{\bigcup_{i=1}^{N-1} I_i^\varepsilon} G(x)\,\mathrm{d}x + \frac{N-1}{N}\varepsilon.$$

Taking I_N^ε as the remaining interval such that $\bigcup_{i=1}^N I_i^\varepsilon = [0,1]$, we have

$$\int_{\bigcup_{i=1}^N I_i^\varepsilon} G(x)\,\mathrm{d}x = M^\tau.$$

For the last interval we see that

$$\int_{I_N^\varepsilon} G(x)\,\mathrm{d}x \geq \frac{M^\tau}{N},$$

and

$$\int_{I_N^\varepsilon} G(x)\,\mathrm{d}x = M^\tau - \int_{\bigcup_{k=1}^{N-1} I_i^\varepsilon} G(x)\,\mathrm{d}x \leq M^\tau - \frac{N-1}{N}\Big(M^\tau - \varepsilon\Big) \leq \frac{1}{N}M^\tau + \varepsilon.$$

Finally, we apply polynomial approximation over each I_i^ε. There exist polynomials P_i of degree $\leq \bar{m}$ over each I_i^ε such that for $f_i := f|_{I_i^\varepsilon}$ (see [19, Chapter 12, Theorem 8.1])

$$\|f_i - P_i\|_p^\tau (I_i^\varepsilon) \leq C^\tau |f_i|_{B^\alpha_{\tau,\tau}}^\tau (I_i^\varepsilon) \leq C' \int_{I_i^\varepsilon} G(x)\,\mathrm{d}x \leq C' \begin{cases} \frac{1}{N}M^\tau, & i = 1,\dots,N-1, \\ \frac{1}{N}M^\tau + \varepsilon, & i = N, \end{cases}$$

where $\|\cdot\| (I_i^\varepsilon)$ means we take norms over I_i^ε only. Setting $s = \sum_{i=1}^N P_i \mathbb{1}_{I_i^\varepsilon}$ and since $p/\tau > 1$, we obtain

$$\begin{aligned}
\|f - s\|_p^p = \sum_{i=1}^N \|f_i - P_i\|_p^p (I_i^\varepsilon) &\leq (N-1)CM^pN^{-p/\tau} + \left(\frac{1}{N}M^\tau + \varepsilon\right)^{p/\tau} \\
&\leq (N-1)CM^pN^{-p/\tau} + 2^{p/\tau-1}\left((\frac{1}{N}M^\tau)^{p/\tau} + \varepsilon^{p/\tau}\right) \\
&\leq \max\left\{C, 2^{p/\tau-1}\right\}\left(M^pN^{1-p/\tau} + \varepsilon^{p/\tau}\right).
\end{aligned}$$

Since the constant is independent of ε and ε can be chosen arbitrarily small, we obtain (5.6). □

Proof of Lemma 5.6. Let $f_k := f\mathbb{1}_{I_k}$. By the Hölder inequality

$$\|f_k\|_p^p = \int_0^1 |f_k(x)|^p\,\mathrm{d}x \leq \left(\int_0^1 |f(x)|^{p\delta}\,\mathrm{d}x\right)^{1/\delta}\left(\int_{I_k}\mathrm{d}x\right)^{1/q}.$$

We choose

$$\Lambda := \{k = 1,\dots,N :\ |I_k| > \varrho(\varepsilon)\}.$$

Then,

$$\sum_{k\notin\Lambda} \|f_k\|_p^p \leq \|f\|_{p\delta}^p N\varrho(\varepsilon)^{1/q} \leq \varepsilon^p.$$

For $\tilde{s}$, we thus estimate

$$\|f - \tilde{s}\|_p^p = \sum_{k\in\Lambda}\|f_k - s_k\|_p^p + \sum_{k\notin\Lambda}\|f_k\|_p^p \leq 2\varepsilon^p.$$

□

Proof of Lemma 5.10. Let $P \in \mathbb{P}_{\bar{m}}$ be arbitrary and set $s := \mathcal{I}_{b,d,m}P$. From Lemma 5.1, we obtain

$$\|P - s\|_p \leq C_1 b^{-d(m+1)} \left\|P^{(m+1)}\right\|_p. \tag{E.2}$$

From Lemma 4.1 we can estimate the complexity of $s \in V_{b,d,m}$ as

$$n := \mathrm{compl}_{\mathcal{C}}(s) \leq b^2 + b(d-1)(\bar{m}+1)^2 + (m+1)^2,$$

or

$$d \geq \frac{n - b^2 + b(\bar{m}+1)^2 - (m+1)^2}{b(\bar{m}+1)^2}.$$

Inserting into (E.2)

$$\|P - s\|_p \leq C_2 b^{-\frac{m+1}{b(\bar{m}+1)^2} n} \left\| P^{(m+1)} \right\|_p.$$

Analogously for $\mathrm{compl}_{\mathcal{N}}$

$$n := \mathrm{compl}_{\mathcal{N}}(s) \leq d(\bar{m}+1),$$

and

$$\|P - s\|_p \leq C_2 b^{-\frac{m+1}{(\bar{m}+1)} n} \left\| P^{(m+1)} \right\|_p.$$

□

Proof of Theorem 5.11. Set

$$M := \sup_{z \in D_\rho} |f(z)|,$$

and $\bar{m} \in \mathbb{N}$. From [19, Chapter 7, Theorem 8.1], we know

$$\inf_{P \in \mathbb{P}_{\bar{m}}} \|f - P\|_\infty \leq \frac{2M}{\rho - 1} \rho^{-\bar{m}}. \tag{E.3}$$

We aim at approximating an arbitrary polynomial of degree $\bar{m}$ within $V_{b,m}$. W.l.o.g. we can assume $\bar{m} > m$, since otherwise $\mathbb{P}_{\bar{m}} \subset V_{b,m}$.

From (E.2) we know

$$\|P - s\|_\infty \leq C_1 b^{-d(m+1)} \left\| P^{(m+1)} \right\|_\infty, \tag{E.4}$$

for a spline $s = \mathcal{I}_{b,d,m} P$ of degree m. To estimate the derivatives $\left\| P^{(m+1)} \right\|_\infty$, we further specify P. Let P be the sum of Chebyshev polynomials from [19, Chapter 7, Theorem 8.1] used to derive (E.3). I.e., since f is assumed to be analytic, we can expand f into a series

$$f(x) = \frac{1}{2} a_0 + \sum_{k=1}^{\infty} a_k C_k(x),$$

where C_k are Chebyshev polynomials of the first kind of degree k. We set $P = P_{\mathrm{Ch}}$ with

$$P_{\mathrm{Ch}} := \frac{1}{2} a_0 + \sum_{k=1}^{\bar{m}} a_k C_k, \tag{E.5}$$

which is such that

$$\|f - P_{\mathrm{Ch}}\|_p \leq \frac{2M}{\rho - 1} \rho^{-\bar{m}}. \tag{E.6}$$

For the derivatives of $C_k^{(m+1)}$ we get by standard estimates (see, e.g., [40])

$$\left\| C_k^{(m+1)} \right\|_\infty \leq \frac{k^2 (k^2 - 1) \cdots (k^2 - m^2)}{(2(m+1) - 1)!}.$$

And thus, for any $1 < \rho_0 < \rho$,

$$\left\| P_{\mathrm{Ch}}^{(m+1)} \right\|_\infty \leq \frac{1}{(2(m+1)-1)!} \sum_{k=m+1}^{\bar{m}} |a_k| k^2(k^2-1)\cdots(k^2-m^2)$$
$$\leq \frac{2M}{(2(m+1)-1)!} \sum_{k=m+1}^{\bar{m}} \rho_0^{-k} k^2(k^2-1)\cdots(k^2-m^2).$$

For $\bar{m} \to \infty$, this series converges to a constant depending on M, m and ρ.

We can now combine both estimates for the final approximation error. We first consider the approximation error $E_n^{\mathcal{C}}(f)_\infty$. Let $n \in \mathbb{N}$ be large enough such that

$$d := \lfloor b^{-1} n^{1/3} - (m+1) n^{-2/3} \rfloor > 1,$$
$$\bar{m} := \lfloor n^{1/3} - 1 \rfloor \geq 1.$$

For this choice of d and $\bar{m}$, let $s \in V_{b,d,m}$ be the interpolant of degree m of the Chebyshev polynomial P_{Ch} from (E.5). Then from Proposition 4.4, we obtain

$$\mathrm{compl}_{\mathcal{C}}(s) \leq bd(\bar{m}+1)^2 + b(m+1) \leq n,$$

and thus $s \in \Phi_n$. Moreover, by (E.4) and (E.6),

$$E_n^{\mathcal{C}}(f)_\infty \leq \|f - P_{\mathrm{Ch}}\|_\infty + \|P_{\mathrm{Ch}} - s\|_\infty \leq \frac{2M}{\rho - 1}\rho^{-\bar{m}} + C_1' b^{-d(m+1)} \left\| P_{\mathrm{Ch}}^{(m+1)} \right\|_\infty$$
$$\leq C_2' [\min(\rho, b^{(m+1)/b})]^{-n^{1/3}}.$$

The result for $E_n^{\mathcal{S}}(f)_\infty$ follows from $\Phi_n^{\mathcal{C}} \subset \Phi_n^{\mathcal{S}}$. Now we consider the case of $E_n^{\mathcal{N}}(f)_\infty$. Let $n \in \mathbb{N}$ be large enough such that $d := \lfloor n^{1/2} \rfloor > 1$ and $\bar{m} = \lfloor n^{1/2} - 1 \rfloor \geq 1$. Then from Proposition 4.4, we obtain

$$\mathrm{compl}_{\mathcal{N}}(s) \leq d(\bar{m}+1) \leq n.$$

Moreover, by (E.4) and (E.6),

$$E_n^{\mathcal{N}}(f)_\infty \leq \frac{2M}{\rho - 1}\rho^{-\bar{m}} + C_1' b^{-d(m+1)} \left\| P_{\mathrm{Ch}}^{(m+1)} \right\|_\infty \leq C_2' [\min(\rho, b^{(m+1)})]^{-n^{1/2}}.$$

□

References

[1] Mazen Ali. Ranks of tensor networks for eigenspace projections and the curse of dimensionality. *arXiv preprint arXiv:2012.12953*, 2020.

[2] Mazen Ali and Anthony Nouy. Approximation of smoothness classes by deep rectifier networks. *SIAM Journal on Numerical Analysis*, 59(6):3032–3051, 2021.

[3] Mazen Ali and Anthony Nouy. Approximation theory of tree tensor networks: tensorized multivariate functions. *arXiv preprint arXiv:2101.11932*, 2021.

[4] Itai Arad, Alexei Kitaev, Zeph Landau, and Umesh Vazirani. An area law and sub-exponential algorithm for 1d systems. *arXiv preprint arXiv:1301.1162*, 2013.

[5] Quentin Ayoul-Guilmard, Anthony Nouy, and Christophe Binetruy. Tensor-based multiscale method for diffusion problems in quasi-periodic heterogeneous media. *ESAIM: Mathematical Modelling and Numerical Analysis*, 52(3):869–891, may 2018.

[6] Markus Bachmayr and Wolfgang Dahmen. Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. *Foundations of Computational Mathematics*, 15(4):839–898, 2015.

[7] Markus Bachmayr, Anthony Nouy, and Reinhold Schneider. Approximation by tree tensor networks in high dimensions: Sobolev and compositional functions. *arXiv preprint arXiv:2112.01474*, 2021.

[8] Markus Bachmayr, Reinhold Schneider, and André Uschmajew. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Foundations of Computational Mathematics*, 16(6):1423–1472, April 2016.

[9] Colin Bennett and Robert Sharpley. *Interpolation of Operators*. Academic Press, Boston, 1988.

[10] Cédric Bény. Deep learning and the renormalization group. *arXiv preprint arXiv:1301.3124*, 2013.

[11] Weronika Buczyńska. The Hackbusch conjecture on tensor formats — part two. *Linear Algebra and its Applications*, 584:221–232, jan 2020.

[12] Weronika Buczyńska, Jarosław Buczyński, and Mateusz Michałek. The Hackbusch conjecture on tensor formats. *Journal de Mathématiques Pures et Appliquées*, 104(4):749–761, oct 2015.

[13] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, feb 2017.

[14] Andrzej Cichocki, Namgil Lee, Ivan Oseledets, Anh-Huy Phan, Qibin Zhao, and Danilo Mandic. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends® in Machine Learning*, 9(4-5):249–429, 2016.

[15] Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan Oseledets, Masashi Sugiyama, and Danilo Mandic. Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives. *Foundations and Trends® in Machine Learning*, 9(6):431–673, 2017.

[16] Nadav Cohen, Or Sharir, and Amnon Shashua. On the expressive power of deep learning: A tensor analysis. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, volume 49 of *JMLR Workshop and Conference Proceedings*, pages 698–728, 2016.

[17] Ronald A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, jan 1998.

[18] Ronald A. DeVore, Ralph Howard, and Charles Micchelli. Optimal nonlinear approximation. *Manuscripta Mathematica*, 63(4):469–478, dec 1989.

[19] Ronald A. DeVore and George G. Lorentz. *Constructive Approximation*. Springer-Verlag Berlin Heidelberg, 1993.

[20] Ronald A. Devore and Vasil A. Popov. Interpolation of besov spaces. *Transactions of the American Mathematical Society*, 305(1):397–414, 1988.

[21] Alexandre Ern and Jean-Luc Guermond. *Theory and Practice of Finite Elements*. Springer Verlarg, 2004.

[22] Antonio Falcó and Wolfgang Hackbusch. On minimal subspaces in tensor representations. *Foundations of Computational Mathematics*, 12:765–803, 2012.

[23] Antonio Falcó, Wolfgang Hackbusch, and Anthony Nouy. Geometric structures in tensor representations (final release). *arXiv preprint arXiv:1505.03027*, 2015.

[24] Antonio Falcó, Wolfgang Hackbusch, and Anthony Nouy. Tree-based tensor formats. *SeMA Journal*, Oct 2018.

[25] Antonio Falcó, Wolfgang Hackbusch, and Anthony Nouy. On the Dirac-Frenkel variational principle on tensor banach spaces. *Foundations of Computational Mathematics*, 19(1):159–204, Feb 2019.

[26] Antonio Falcó, Wolfgang Hackbusch, and Anthony Nouy. Geometry of tree-based tensor formats in tensor banach spaces. *arXiv preprint arXiv:2011.08466*, 2020.

[27] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[28] Lars Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2029–2054, 2010.

[29] Lars Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Technical report, Institut für Geometrie und Prakitsche Mathematik, RWTH Aachen, 2010.

[30] Erwan Grelier, Anthony Nouy, and Mathilde Chevreuil. Learning with tree-based tensor formats. *arXiv e-prints, arXiv:1811.04455*, 2018.

[31] Erwan Grelier, Anthony Nouy, and Regis Lebrun. Learning high-dimensional probability distributions using tree tensor networks. *International Journal for Uncertainty Quantification*, 12(5):47–69, 2022.

[32] Rémi Gribonval, Gitta Kutyniok, Morten Nielsen, and Felix Voigtlaender. Approximation spaces of deep neural networks. *Constructive approximation*, 55(1):259–367, 2022.

[33] Michael Griebel and Helmut Harbrecht. Analysis of tensor approximation schemes for continuous functions. *Foundations of Computational Mathematics*, pages 1–22, 2021.

[34] Wolfgang Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer Berlin Heidelberg, 2012.

[35] Wolfgang Hackbusch and Stefan Kuhn. A New Scheme for the Tensor Representation. *Journal of Fourier analysis and applications*, 15(5):706–722, 2009.

[36] Matthew B. Hastings. An area law for one-dimensional quantum systems. *Journal of statistical mechanics: theory and experiment*, 2007(08):P08024, 2007.

[37] Simon Haykin. *Neural Networks and Learning Machines*. Prentice Hall, Pearson, New York, 2009.

[38] Sebastian Holtz, Thorsten Rohwedder, and Reinhold Schneider. On manifolds of tensors of fixed TT-rank. *Numerische Mathematik*, 120(4):701–731, sep 2011.

[39] Abdolhossein Hoorfar and Mehdi Hassani. Inequalities on the Lambert W function and hyperpower function. *J. Inequalities in Pure and Applied Math.*, 9(2), 2008.

[40] David C. Handscomb J. C. Mason. *Chebyshev Polynomials*. Taylor & Francis Ltd, 2002.

[41] Vladimir Kazeev and Christoph Schwab. Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions. *Numerische Mathematik*, 138(1):133–190, jul 2017.

[42] Boris N. Khoromskij. $\mathrm{O}(\mathrm{d}\log\mathrm{n})$-quantics approximation of n-d tensors in high-dimensional numerical modeling. *Constructive Approximation*, 34(2):257–280, apr 2011.

[43] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. Quantum entanglement in deep learning architectures. *Phys. Rev. Lett.*, 122:065301, Feb 2019.

[44] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. Deep learning and quantum entanglement: Fundamental connections with implications to network design. In *International Conference on Learning Representations*, 2018.

[45] Bertrand Michel and Anthony Nouy. Learning with tree tensor networks: complexity estimates and model selection. *Bernoulli*, 28(2):910–936, 2022.

[46] Anthony Nouy. *Low-rank Methods for High-dimensional Approximation and Model Order Reduction*, chapter 4. SIAM, Philadelphia, PA, 2017.

[47] Anthony Nouy. Higher-order principal component analysis for the approximation of tensors in tree-based low-rank formats. *Numerische Mathematik*, 141(3):743–789, Mar 2019.

[48] Joost A. A. Opschoor, Philipp C. Petersen, and Christoph Schwab. Deep ReLU networks and high-order finite element methods. *Analysis and Applications*, pages 1–56, feb 2020.

[49] Joost A.A. Opschoor, Christoph Schwab, and Jakob Zech. Exponential ReLU DNN expression of holomorphic maps in high dimension. Technical report, Zurich, 2019-07.

[50] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, oct 2014.

[51] Román Orús. Tensor networks for complex quantum systems. *Nature Reviews Physics*, 1(9):538–550, aug 2019.

[52] Ivan Oseledets. Approximation of matrices with logarithmic number of parameters. *Doklady Mathematics*, 80(2):653–654, oct 2009.

[53] Ivan Oseledets. Constructive representation of functions in low-rank tensor formats. *Constructive Approximation*, 37(1):1–18, dec 2012.

[54] Ivan Oseledets and Eugene E. Tyrtyshnikov. Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM Journal on Scientific Computing*, 31(5):3744–3759, 2009.

[55] Hoifung Poon and Pedro Domingos. Sum-product networks: A new deep architecture. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 689–690. IEEE, 2011.

[56] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, jan 2015.

[57] Reinhold Schneider and André Uschmajew. Approximation rates for the hierarchical tensor format in periodic sobolev spaces. *Journal of Complexity*, 30(2):56–71, apr 2014.

[58] Martin Schwarz, Olivier Buerschaper, and Jens Eisert. Approximating local observables on projected entangled pair states. *Physical Review A*, 95(6), jun 2017.

[59] Matus Telgarsky. Representation benefits of deep feedforward networks. *arXiv preprint arXiv:1509.08101*, 2015.

[60] Dmitry Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103–114, oct 2017.

[61] Dmitry Yarotsky. Optimal approximation of continuous functions by very deep relu networks. In *Conference on learning theory*, pages 639–649. PMLR, 2018.

[62] James Yeh. *Real Analysis: Theory Of Measure And Integration (3rd Edition)*. World Scientific Publishing Company, 2014.