APPROXIMATION WITH TENSOR NETWORKS.
PART II: APPROXIMATION RATES FOR SMOOTHNESS CLASSES.

MAZEN ALI AND ANTHONY NOUY

Centrale Nantes, LMJL UMR CNRS 6629, France

Abstract. We study the approximation by tensor networks (TNs) of functions from classical smoothness classes. The considered approximation tool combines a tensorization of functions in $L^p([0,1))$, which allows to identify a univariate function with a multivariate function (or tensor), and the use of tree tensor networks (the tensor train format) for exploiting low-rank structures of multivariate functions. The resulting tool can be interpreted as a feed-forward neural network, with first layers implementing the tensorization, interpreted as a particular featuring step, followed by a sum-product network with sparse architecture.

In part I of this work, we presented several approximation classes associated with different measures of complexity of tensor networks and studied their properties.

In this work (part II), we show how classical approximation tools, such as polynomials or splines (with fixed or free knots), can be encoded as a tensor network with controlled complexity. We use this to derive direct (Jackson) inequalities for the approximation spaces of tensor networks. This is then utilized to show that Besov spaces are continuously embedded into these approximation spaces. In other words, we show that arbitrary Besov functions can be approximated with optimal or near to optimal rate. We also show that an arbitrary function in the approximation class possesses no Besov smoothness, unless one limits the *depth* of the tensor network.

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., [4, 9, 10, 19, 22, 26, 29, 31] and references therein for an overview.

In part I of this work (see [2]), we defined approximation classes of tensor networks and studied their properties. Our goal was to introduce a new method for analyzing the expressivity of tensor networks, as was done in [21] for rectified linear unit (ReLU) and rectified power unit (RePU) neural networks.

In this work (part II), we continue with our endeavor by studying how these approximation classes of tensor networks are related to the well-known Besov spaces (see, e.g., [34]). In particular, we will show that any Besov function can be approximated with optimal rate with a tensor network. On the other hand, an arbitrary function from the approximation class of tensor networks has no Besov smoothness, unless we restrict the *depth* of the tensor network. This illustrates the high expressivity of (deep) tensor networks. Interestingly, similar results were shown for deep ReLU and RePU networks in, e.g., [21, 27, 28, 36]. We also consider the case of analytic functions and show that they can be approximated with exponential rate.

The outline is as follows. In Section 2, we begin by recalling some notations and results from part I [2]. We then state the main results of this work in Section 3. In Section 4, we discuss how classical approximation tools can be encoded with tensor networks and estimate the resulting complexity. Among

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

Date: July 2, 2020.

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.

Acknowledgments: The authors acknowledge AIRBUS Group for the financial support with the project AtRandom.

the classical tools considered are fixed knot splines, free knot splines, polynomials (of higher order) and general multi-resolution analysis (MRA). In Section 5, utilizing results from the previous section, we show direct estimates for our approximation tool that lead to the embeddings from Main Result 3.1. We also show that inverse embeddings can only hold if we restrict the depth of the tensor networks. We conclude on Section 6 by a brief discussion of depth and sparse connectivity.

For a brief introduction to TNs and NNs, we refer to the introduction in [2].

2. Recalling Results of Part I

2.1. Tensorization. We consider one-dimensional functions on the unit interval

$$f : [0,1) \to \mathbb{R}.$$

We introduce a uniform partition of $[0,1)$ with b^d intervals $[x_i, x_{i+1})$ with $x_i = b^{-d}i$, $0 \le i \le b^d$, with base $b = 2, 3, ...$, and exponent $d \in \mathbb{N}_0$. Any $x \in [0,1)$ falls either on or in-between one of the points x_i. Thus, using a b-adic expansion of the integer i, we can define a *conversion map* $t_{b,d}$ as

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

for $y \in [0,1)$ and $i_k \in \{0, \ldots, b-1\} := I_b$.

Definition 2.1 (Tensorization Map)**.** *For a base $b \in \mathbb{N}$ ($b \ge 2$) and a level $d \in \mathbb{N}$, 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

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

For $i = \sum_{k=1}^{d} i_k b^{d-k}$, the partial evaluation $\boldsymbol{f}(i_1, \ldots, i_d, \cdot)$ of $\boldsymbol{f}$ is equal to the function $f(b^{-d}(i+\cdot)) \in \mathbb{R}^{[0,1)}$, which is the restriction of f to the interval $[b^{-d}i, b^{-d}(i+1))$ rescaled to $[0,1)$.

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 $g^k \in \mathbb{R}^{[0,1)}$, $1 \le k \le r$, $1 \le \nu \le d$.

We showed that L^p functions can be isometrically identified with tensors.

Theorem 2.2 (Tensorization is an L^p-Isometry)**.** *For any $0 < p \le \infty$, define*

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

Then, $T_{b,d}$ 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

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

For an interpretation of the tensorization as a particular featuring step and its implementation as a specific neural network, we refer to [2, Section 4].

2.2. **Ranks.** The minimal $r \in \mathbb{N}$ such that $\boldsymbol{f}$ admits a representation as in (2.1) is referred to as the *canonical tensor rank* of $\boldsymbol{f}$. Crucial for this work is the following notion of multi-linear rank.

Definition 2.3 (β-rank)**.** *For $\beta \subset \{1,\dots,d+1\}$, 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*

$$\boldsymbol{f} = \sum_{k=1}^{r_\beta(\boldsymbol{f})} \boldsymbol{v}_\beta^k \otimes \boldsymbol{v}_{\beta^c}^k, \tag{2.2}$$

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

Since a function $f \in \mathbb{R}^{[0,1)}$ admits a representation as a tensor for different $d \in \mathbb{N}$, we require the following notion of ranks for a function.

Definition 2.4 ((β, d)-rank)**.** *For a function $f \in \mathbb{R}^{[0,1)}$, $d \in \mathbb{N}$ and $\beta \subset \{1,\dots,d+1\}$, we define the (β,d)-rank of f, denoted $r_{\beta,d}(f)$, as the rank of its tensorization in $\mathbf{V}_{b,d}$,*

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

We mostly consider subsets β of the form $\{1,\dots,\nu\}$ or $\{\nu+1,\dots,d+1\}$ for some $\nu \in \{1,\dots,d\}$, and thus we will use the shorthand notations

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

The notion of partial evaluations will be useful for estimating the ranks of a function f.

Lemma 2.5. *Let $f \in \mathbb{R}^{[0,1)}$ and $d \in \mathbb{N}$. For any $1 \le \nu \le d$,*

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

and

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

2.3. **Tensor Subspaces.** For a subspace $S \subset \mathbb{R}^{[0,1)}$, 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 subspace of functions,

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

We frequently use $S = \mathbb{P}_m$ where $\mathbb{P}_m$ is the space of polynomials of degree up to $m \in \mathbb{N}_0$, with the shorthand notation

$$V_{b,d,m} := V_{b,d,\mathbb{P}_m}. \tag{2.3}$$

It will be important in the sequel that S satisfies the following property, analogous to properties satisfied by spaces generated by multi-resolution analysis (MRA).

Definition 2.6 (Closed under Dilation)**.** *We say that a linear space S is closed under b-adic dilation if for any $f \in S$ and any $k \in \{0,\dots,b-1\}$,*

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

If S is closed under b-adic dilation, we can show that

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

and that

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

is a linear subspace of $\mathbb{R}^{[0,1)}$. Moreover, we have

Lemma 2.7. *Let S be closed under b-adic dilation.*

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

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

(ii) If $f \in V_{b,d,S}$, then for any $\bar{d} \geq d$, $f \in V_{b,\bar{d},S}$ satisfies

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

$$r_{\nu,\bar{d}}(f) \leq \min\left\{b^{\nu}, \dim S\right\}, \quad d < \nu \leq \bar{d}. \tag{2.4}$$

Letting $\mathcal{I}_S$ be a linear projection operator from $L^p([0,1))$ to a finite-dimensional space S, we define a linear operator $\mathcal{I}_{b,d,S}$ from $L^p([0,1))$ to $V_{b,d,S}$ by

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

or [2, Lemma 2.29]

$$\mathcal{I}_{b,d,S} = T_{b,d}^{-1} \circ (id_{\{1,\ldots,d\}} \otimes \mathcal{I}_S) \circ T_{b,d}. \tag{2.5}$$

We can bound the ranks of local projections onto S in the following sense.

Lemma 2.8 (Local projection ranks)**.** *For any $f \in L^p$, $\mathcal{I}_{b,d,S}f \in V_{b,d,S}$ satisfies*

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

2.4. **Tensor Train Format and Corresponding Approximation Tool.** In this work, we will be using tree tensor networks, and more precisely the *tensor train format*, to define our approximation tool.

Definition 2.9 (Tensor Train Format)**.** *The set 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*

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

This defines a set of univariate functions

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

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

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

$$\boldsymbol{f}(i_1,\ldots,i_d,y) = \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{k_{d+1}}^{\dim S} v_1^{k_1}(i_1)v_2^{k_1,k_2}(i_2)v_3^{k_2,k_3}(i_3)\cdots v_d^{k_{d-1},k_d}(i_d)v_{d+1}^{k_d,k_{d+1}}\varphi_k(y), \tag{2.6}$$

where the parameters $\mathbf{v} := (v_1,\ldots,v_{d+1})$ form a tensor network with

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

Denoting by $\mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v})$ the map which associates to a tensor network $\mathbf{v}$ the function $f = T_{b,d}\boldsymbol{f}$ with $\boldsymbol{f}$ defined by (2.6), we have

$$\Phi_{b,d,S,\boldsymbol{r}} = \{\varphi = \mathcal{R}_{b,d,S,\boldsymbol{r}}(\mathbf{v}) : \mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}\}.$$

We introduced three different measures of complexity of a tensor network $\mathbf{v} \in \mathcal{P}_{b,d,S,\boldsymbol{r}}$:

$$\mathrm{compl}_{\mathcal{N}}(\mathbf{v}) := \sum_{\nu=1}^{d} r_\nu, \tag{2.7}$$

$$\mathrm{compl}_{\mathcal{C}}(\mathbf{v}) := br_1 + b\sum_{k=2}^{d} r_{k-1}r_k + r_d \dim S,$$

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

where $\|v_\nu\|_{\ell_0}$ is the number of non-zero entries in the tensor v_ν. The function $\mathrm{compl}_{\mathcal{C}}(\mathbf{v})$ is a natural measure of complexity equal to the number of parameters. The function $\mathrm{compl}_{\mathcal{S}}$ is also a natural measure of complexity taking account the sparsity of tensors in the tensor network, and equal to the number of non-zero parameters. When interpreting a tensor network $\mathbf{v}$ as a sum-product neural network, $\mathrm{compl}_{\mathcal{N}}(\mathbf{v})$ corresponds to the number of neurons, $\mathrm{compl}_{\mathcal{C}}(\mathbf{v})$ to the number of weights, and $\mathrm{compl}_{\mathcal{S}}(\mathbf{v})$ the number of non-zero weights (taking into account sparsity in the connectivity of the network).

Our approximation tool for univariate functions is defined as

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

where $\mathrm{compl}(\varphi)$ is a measure of complexity of a function φ, defined as

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

where the infimum is taken over all tensor networks $\mathbf{v}$ whose realization is the function φ. The three complexity measures define three types of subsets

$$\Phi_n^{\mathcal{N}} := \{\varphi \in \Phi : \mathrm{compl}_{\mathcal{N}}(\varphi) \le n\},$$
$$\Phi_n^{\mathcal{C}} := \{\varphi \in \Phi : \mathrm{compl}_{\mathcal{C}}(\varphi) \le n\},$$
$$\Phi_n^{\mathcal{S}} := \{\varphi \in \Phi : \mathrm{compl}_{\mathcal{S}}(\varphi) \le n\}.$$

For the implementation of the resulting approximation tool as a particular feed-forward neural network, we refer to [2, Section 4].

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

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

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

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

for $\alpha > 0$ and $0 < q \le \infty$, with

$$\|f\|_{A_q^\alpha} := \begin{cases} \left(\sum_{n=1}^\infty [n^\alpha E_{n-1}(f)]^q \frac{1}{n}\right)^{1/q}, & 0 < q < \infty, \\ \sup_{n\ge1}[n^\alpha E_{n-1}(f)], & q = \infty. \end{cases}$$

Considering $\Phi_n \in \{\Phi_n^{\mathcal{N}}, \Phi_n^{\mathcal{C}}, \Phi_n^{\mathcal{S}}\}$, we obtain families of approximation classes of tensor networks

$$\begin{aligned} N_q^\alpha(X) &:= A_q^\alpha(X, (\Phi_n^{\mathcal{N}})_{n\in\mathbb{N}}), \\ C_q^\alpha(X) &:= A_q^\alpha(X, (\Phi_n^{\mathcal{C}})_{n\in\mathbb{N}}), \\ S_q^\alpha(X) &:= A_q^\alpha(X, (\Phi_n^{\mathcal{S}})_{n\in\mathbb{N}}). \end{aligned} \tag{2.8}$$

There are a few properties of Φ that turn out to be crucial, if one is to obtain strong statements about the properties of the associated approximation classes.

(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.

In particular, (P1)–(P4) together with a Jackson inequality imply that the approximation classes are quasi-normed linear spaces. A main result of Part I [2] is

Theorem 2.10. *For any $\alpha > 0$, $0 < p \le \infty$ and $0 < q \le \infty$, the approximation classes $N_q^\alpha(L^p)$, $C_q^\alpha(X)$ and $S_q^\alpha(X)$ are quasi-normed linear spaces satisfying the continuous embeddings*

$$C_q^\alpha(L^p) \hookrightarrow S_q^\alpha(L^p) \hookrightarrow N_q^\alpha(L^p) \hookrightarrow C_q^{\alpha/2}(L^p).$$

3. Main Results of This Work

The main results of this work are Theorem 5.9, Theorem 5.16, Theorem 5.18 and Theorem 5.20, and can be summarized as follows.

Main Result 3.1 (Direct Embedding for Sobolev Spaces)**.** *Let $W^{r,p}$ denote the Sobolev space of $r \in \mathbb{N}$ times weakly differentiable, p-integrable functions and $B_{p,q}^\alpha$ 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 $1 \le p \le \infty$ and any $r \in \mathbb{N}$*

$$W^{r,p} \hookrightarrow N_\infty^{2r}(L^p), \quad W^{r,p} \hookrightarrow C_\infty^r(L^p) \hookrightarrow S_\infty^r(L^p),$$

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

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

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 5.8.*

Main Result 3.2 (Direct Embedding for Besov Spaces). *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* $r > 1/\tau - 1/p$ *and any* $0 < \bar{r} < r$,

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

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

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

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

Remark 3.3. *Note that both Main Result 3.1 and Main Result 3.2 apply to Besov spaces. The Besov spaces in Main Result 3.1 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, see also Figure 1 of Section 5. These Besov spaces can only be captured by non-linear approximation. Our results require* $\alpha > 1/\tau - 1/p$, *i.e., Besov spaces that are strictly above the critical line.*

Main Result 3.4 (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* $\tilde{\alpha} > 0$,

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

Main Result 3.5 (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* $[0, 1]$,

$$E^{\mathcal{N}}_n(f)_\infty \leq C\rho^{-n^{1/2}},$$

$$E^{\mathcal{S}}_n(f)_\infty \leq E^{\mathcal{C}}_n(f)_\infty \leq C\rho^{-n^{1/3}},$$

for constants $C, \rho > 1$. *This can be extended to analytic functions with singularities using ideas from [24].*

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 $W^{r,p}$ of *any* order $r > 0$.
- 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 tensor network. It does not reflect the representation or computational complexity. Moreover, from [14] we know that an optimal approximation tool with *continuous* parametrization for the Sobolev space $W^{r,p}$ cannot exceed the rate r, see also [36].
- For the approximation tools $\Phi^{\mathcal{N}}_n$ and $\Phi^{\mathcal{S}}_n$, we obtain *near to* optimal rates[1] for the Besov space $B^r_{\tau,\tau}$, for any order $r > 0$. For $\Phi^{\mathcal{C}}_n$, the approximation rate is near to half the optimal rate.
- Particularly the tool $\Phi^{\mathcal{S}}_n$ 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[2].
- Finally, an arbitrary function from any of the three approximation classes possesses no Besov smoothness. We will also see in Proposition 5.21 that this can be mainly attributed to the depth of the tensor network.

[1] I.e., the approximation rates are arbitrary close to optimal.

[2] Even though the underlying polynomial degree of the tensor network remains fixed.

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 avoid the technical difficulties of general multi-dimensional domains. However, the ideas can be in principle extended to any dimension and we intend to do so in a part III of this work.

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 approximation with tensor formats, see, e.g., [7, 8]. In the multi-dimensional case ranks remain low if the format "fits" the problem at hand, in a certain sense. 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 [3, 20, 30].

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.

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, polynomials and multi-resolution analysis (MRAs). This will be the basis for Section 5 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 2.7 (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 only if $\bar{m} \leq m$. 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 used in the next section.

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 $l = 0, \ldots, m+1$, it holds*

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

for some constant $C > 0$ independent of v. For the construction of this operator and a proof of the above property see, e.g., [18, 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 (2.3) with $\mathcal{I}_S = \mathcal{I}_m$.

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 2.7 (ii). Then from Lemma 2.8, 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 (2.7) 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 $\varphi \in \mathcal{S}_{\mathfrak{c}}^{N,m}$. 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 piece-wise 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 $kb^{-\nu}$, $0 \leq k < b^{\nu}$). Lemma 2.5 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 2.7. ☐

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 [2, Lemma 2.26])*

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

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 [2, Remark 3.3].

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 (2.7) for encoding within $V_{b,m}$ are bounded as*

$$\begin{aligned}\mathrm{compl}_{\mathcal{N}}(\varphi) &\leq C\sqrt{N},\\ \mathrm{compl}_{\mathcal{S}}(\varphi) \leq \mathrm{compl}_{\mathcal{C}}(\varphi) &\leq CN,\end{aligned}$$

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

Proof. From Lemma 4.6, we obtain

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

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

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

Now we would like to encode splines of degree $\bar{m}$ in $V_{b,\bar{d},m}$ with $m \neq \bar{m}$ and $\bar{d} \geq d$. 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 $\bar{d} \geq d$, the interpolant $\mathcal{I}_{b,\bar{d},m}(\varphi) \in V_{b,\bar{d},m}$ satisfies*

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

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

Proof. From Lemma 2.8, 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 Lemma 2.5 that $r_{\nu,\bar{d}}(\varphi) = r_{\nu,d}(\varphi)$. Then, we obtain the first inequality from Lemma 4.6. Now consider the case $d < \nu \leq \bar{d}$. 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 2.7 the other bound $r_{\nu,\bar{d}}(\varphi) \leq \bar{m}+1$. □

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* (2.7) *for encoding the interpolant $\mathcal{I}_{b,\bar{d},m}(\varphi)$ of level $\bar{d} \geq d$ and degree $m \leq \bar{m}$ within $V_{b,m}$ are bounded as*

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

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

Proof. Using Lemma 2.8 and Lemma 4.6 and following the proof of Proposition 4.8, we have

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

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

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

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

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

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 5.3.2, 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*

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

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*

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

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, Lemma 2.5 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 2.7 with $\dim(S) = m+1$. □

Proposition 4.13 (Complexity for Encoding Free 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 (2.7) for encoding within $V_{b,m}$ are bounded as*

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

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

Proof. Follows from Lemma 4.12, cf. also Proposition 4.8. □

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*

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

Proof. From Lemma 2.8, 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 Lemma 2.5 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 2.7 the other bound $r_{\nu,\bar{d}}(\varphi) \leq \bar{m}+1$. □

Proposition 4.15 (Complexity for Encoding Interpolants of Free 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 (2.7) 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*

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

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

Proof. Follows from Lemma 4.14, cf. also Proposition 4.10. □

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. See also Remark 4.25.*

4.4. **Multi-Resolution Analysis.** We have already mentioned a connection between $V_{b,S}$ and MRAs: see Definition 2.6. In this subsection, we further explore if and how MRAs are intrinsically encoded within $V_{b,S}$. Specifically, we consider the following three scenarios:

- $S = \mathbb{P}_m$ and piecewise polynomial MRAs.
- S itself contains the generators of the MRA.
- We can *approximate* the generators of the MRA with functions in S or $V_{b,d,S}$ upto a fixed accuracy $\varepsilon > 0$.

For a review of MRAs we refer to, e.g., [25, Chapter 7]. An MRA of the space L^p $(1 \leq p < \infty)$ consists of a sequence of spaces

$$V_0 \subset V_1 \subset V_2 \subset \ldots \subset L^p.$$

Usually one considers MRAs on $L^p(\mathbb{R})$, but they can be adapted to bounded domains, see [12]. The sequence V_j is required to satisfy certain properties, such as invariance under b-adic dilation and shifting – the counterpart of Definition 2.6. Another property is that the sequence V_j is generated by dilating and shifting one or a finite number of so-called *generating* functions.

For a function $\psi : [0,1) \to \mathbb{R}$, we define the level $l \in \mathbb{N}_0$ b-adic dilation, shifted by $j = 0, \ldots, b^l - 1$, as

$$\psi_{b,l,j}(x) := \begin{cases} b^{l/p}\psi(b^l x - j) & \text{for} \quad x \in [b^{-l}j, b^{-l}(j+1)) \\ 0 & \text{elsewhere.} \end{cases}$$

The factor $b^{l/p}$ normalizes these functions in L^p. The purpose of this subsection is to illustrate the interplay between the spaces generated by such functions $\psi_{b,l,j}$ and $V_{b,S}$. The following result provides the tensorized representations of functions $\psi_{b,l,j}$.

Lemma 4.17 (b-adic dilations)**.** *Let $l \in \mathbb{N}$ and $j = \sum_{k=1}^{l} j_k b^{l-k}$. Then*

$$T_{b,l}\psi_{b,l,j} = e^p_{j_1} \otimes \cdots \otimes e^p_{j_l} \otimes \psi$$

with $e^p_{j_k} = b^{1/p}\delta_{j_k}$, and for all $d \geq l$,

$$T_{b,d}\psi_{b,l,j} = e^p_{j_1} \otimes \cdots \otimes e^p_{j_l} \otimes (T_{b,d-l}\psi).$$

Proof. The expression of $T_{b,l}\psi_{b,l,j}$ follows from [2, Corollary 2.5]. The second property follows from [2, Lemma 2.6]. □

Lemma 4.17 immediately gives

Corollary 4.18. *For any subspace $S \subset L^p$, $0 \leq l, d_0$ and $0 \leq j \leq b^l - 1$, we have*

$$\psi_{b,l,j} \in V_{b,d_0+l,S} \quad \Leftrightarrow \quad \psi \in V_{b,d_0,S}.$$

Moreover, if $\psi \in V_{b,d_0,S}$ satisfies

$$r_{\nu,d_0}(\psi) \leq r_{\nu,d_0}, \quad 1 \leq \nu \leq d_0,$$

then for $d \geq d_0$ and $l := d - d_0$, we have $\psi_{b,l,j} \in V_{b,d,S}$ and

$$r_{\nu,d}(\psi_{b,l,j}) = 1 \text{ for } 1 \leq \nu \leq l, \quad \text{and} \quad r_{\nu,d}(\psi_{b,l,j}) \leq r_{\nu-l,d_0} \text{ for } l < \nu \leq d.$$

4.4.1. *Piecewise Polynomial MRA.* Let $\psi : [0,1) \to \mathbb{R}$ be a piece-wise polynomial of degree at most m and let the discontinuities be located on a subset of

$$\left\{x_k := kb^{-d_0} : \; k = 0, \ldots, b^{d_0}\right\}.$$

Then, clearly $\psi \in V_{b,d_0,m}$. For $1 \leq \nu \leq d_0$, the ranks can be bounded as in Theorem 4.6

$$r_{\nu,d_0}(\psi) \leq \min\left\{(m - \mathfrak{c})b^{d_0-\nu} + (\mathfrak{c}+1), \; b^\nu\right\},$$

where $\mathfrak{c}$ is the number of continuous derivatives. Note that we can generalize this to arbitrary b-adic discontinuity knots and proceed as in Theorem 4.12, but we restrict ourselves to the above setting for the purpose of this presentation. Moreover, if ψ is a typical wavelet or scaling function, we expect d_0 and consequently $r_\nu(\psi)$ to be small.

By Lemma 2.7, for any $d \geq d_0$ we also have $\psi \in V_{b,d,m}$ with

$$
\begin{aligned}
r_{\nu,d}(\psi) &\leq \min\left\{(m-\mathfrak{c})b^{d_0-\nu} + (\mathfrak{c}+1),\ b^\nu\right\}, \quad 1 \leq \nu \leq d_0,\\
r_{\nu,d}(\psi) &\leq m+1, \quad d_0 \leq \nu \leq d.
\end{aligned}
$$

In summary, if ψ is piece-wise polynomial as above, then, since $S = \mathbb{P}_m$ is closed under b-adic dilation, we know that

a) $\psi \in V_{b,d,m}$ for any $d \geq d_0$ with ranks bounded by Lemma 2.7 and Lemma 4.6,
b) and $\psi_{b,l,j} \in V_{b,d,m}$ for any $d \geq d_0 + l$ by Corollary 4.18 with similar rank bounds.

Example 4.19 (Haar functions)**.** *The Haar mother wavelet $\psi(x) = -\mathbb{1}_{(0,1/2)}(x) + \mathbb{1}_{(1/2,1)}(x)$ is such that $\psi \in V_{2,1,0}$ ($b = 2$, $d_0 = 1$, $m = 0$). Its tensorization at level 1 is $T_{2,1}\psi(i_1, y) = -\delta_0(i_1) + \delta_1(i_1) := h(i_1)$ and its rank $r_{1,1}(\psi) = 1$. Since $T_{2,1}\psi$ does not depend on y, we have $T_{2,d}\psi(i_1, \ldots, i_d, y) = h(i_1)$ and $r_{\nu,d}(\psi) = 1$ for all $d \in \mathbb{N}$ and $1 \leq \nu \leq d$. The Haar wavelets $\psi_{2,l,j} = 2^{-l/2}\psi(2^l x - j)$ are such that for all $d \geq l+1$, $T_{b,d}\psi_{2,l,j}(i_1, \ldots, i_d, y) = e^2_{j_1}(i_1) \ldots e^2_{j_{l-1}}(i_{l-1})h(i_l)$, and $r_{\nu,d}(\psi_{2,l,j}) = 1$ for all $1 \leq \nu \leq d$.*

Example 4.20 (Hat functions)**.** *The hat function $\psi(x) = 2x\mathbb{1}_{(0,1/2)}(x) + 2(1-x)\mathbb{1}_{(1/2,1)}(x)$ is such that $\psi \in V_{2,1,1}$ ($b = 2$, $d_0 = 1$, $m = 1$). Its tensorization at level 1 is $T_{2,1}\psi(i_1, y) = \delta_1(i_1) + (\delta_0(i_1) - \delta_1(i_1))y$ and its rank $r_{1,1}(\psi) = 2$. For any $d \geq 1$, $T_{2,d}\psi(i_1, \ldots, i_d, y) = \delta_1(i_1) + (\delta_0(i_1) - \delta_1(i_1))t_{2,d-1}(i_2, \ldots, i_d, y)$, with $t_{2,d-1}(i_2, \ldots, i_d, y) = 2^{-d+1}(\sum_{k=1}^{d-1} i_{k+1}2^k + y)$, an expression from which we deduce that $r_{\beta,d}(\psi) \leq 2$ for all $\beta \in \{1, \ldots, d+1\}$. The functions $\psi_{2,l,j} = 2^{-l/2}\psi(2^l x - j)$ are such that for all $d \geq l+1$, $T_{b,d}\psi_{2,l,j}(i_1, \ldots, i_d, y) = e^2_{j_1}(i_1) \ldots e^2_{j_l}(i_l)T_{b,d-l}\psi(i_{l+1}, \ldots, i_d, y)$, and $r_{\beta,d}(\psi_{2,l,j}) \leq 2$ for all $\beta \in \{1, \ldots, d+1\}$.*

4.4.2. *S Contains MRA Generators.* Suppose ψ from above is a mother wavelet and $\psi \in S$. Then, by Corollary 4.18, $\psi_{b,l,j} \in V_{b,l,S}$. However, we do not necessarily have $\psi_{b,l,j} \in V_{b,d,S}$ for $d > l$. This presents a problem when considering two wavelets on different levels, since the sum may not belong to $V_{b,d,S}$ for any d (see also [2, Example 2.20]). Thus, as before we require S to be closed under b-adic dilation. It is not difficult to see that this is the case exactly when S includes *all* generators for the MRA, i.e., all mother scaling functions and their shifts.

Put more precisely, let φ^q, $1 \leq q \leq Q$, be mother scaling functions, and ψ^q the corresponding mother wavelets. We intentionally include the possibility of multiple scaling functions and wavelets since this is the relevant setting for wavelets on bounded domains or orthogonal multi-wavelets. The shift-invariant setting on unbounded domains with a single scaling function and infinitely many integer shifts can be handled similarly.

We assume these functions satisfy refinement relationships of the form

$$
\begin{aligned}
\varphi^q &= \sum_{q,i} a_i^q \varphi^q(b\cdot -i),\\
\psi^q &= \sum_{q,i} c_i^q \varphi^q(b\cdot -i),
\end{aligned}
\tag{4.5}
$$

where the number of non-zero $a_i^q \neq 0$, $c_i^q \neq 0$ is typically smaller than the total number of scaled and shifted $\varphi^q(b\cdot -i)$. The refinement property is standard for MRAs.

Example 4.21 (Haar functions)**.** *Consider again Example 4.19. Here $Q = 1$, with mother Haar scaling function $\varphi(x) = \mathbb{1}_{[0,1)}(x)$ such that $\varphi(x) = \varphi(2x) + \varphi(2x-1)$, and the mother Haar wavelet $\psi(x) = -\varphi(2x) + \varphi(2x-1)$.*

Proposition 4.22 (Ranks for MRA)**.** *Suppose we have*

$$S = \operatorname{span}\{\varphi^q :\ q = 1, \ldots, Q\},$$

with functions satisfying (4.5)*. Then, S is closed b-adic dilation. Moreover, by the refinement relation, $\psi^q \in V_{b,1,S}$ and therefore for the ranks we obtain*

$$\begin{aligned}
&\varphi^q_{b,l,j} \in V_{b,d,S},\ l \geq 0,\ d \geq l, && r_\nu(\varphi^q_{b,l,j}) = 1,\ 1 \leq \nu \leq l,\\
& && r_\nu(\varphi^q_{b,l,j}) \leq \dim S,\ l < \nu \leq d,\\
&\psi^q_{b,l,j} \in V_{b,d,S},\ l \geq 0,\ d \geq l+1, && r_\nu(\psi^q_{b,l,j}) = 1,\ 1 \leq \nu \leq l,\\
& && r_\nu(\psi^q_{b,l,j}) \leq \dim S,\ l < \nu \leq d.
\end{aligned}$$

Proof. An application of Corollary 4.18 and (4.5). □

4.4.3. *Approximate MRA Representations.* Let $\eta > 0$ and suppose that for some $d = d(\eta)$ and ranks $r_\nu(\eta)$, $1 \leq \nu \leq d(\eta)$, there exists a $v^\eta \in V_{b,d,S}$ such that

$$\|\psi - v^\eta\|_p \leq \eta \quad r_\nu(v^\eta) \leq r_\nu(\eta),$$

where ψ is a mother wavelet. In the following results ψ can be replaced by a scaling function φ without any changes.

For $l \geq 0$, $0 \leq j \leq b^l - 1$ and $j = \sum_{k=1}^{l} j_k b^{l-k}$, we define

$$v^\eta_{b,l,j} := T^{-1}_{b,l}(e^p_{j_1} \otimes \cdots \otimes e^p_{j_l} \otimes v^\eta), \tag{4.6}$$

where $e^p_{j_k}$ are defined as in Lemma 4.17. The ranks of $v^\eta_{b,l,j}$ are given by Corollary 4.18. Moreover, $v^\eta_{b,l,j}$ approximates $\psi_{b,l,j}$ at least as well as v^η approximates ψ.

Lemma 4.23 (Approximate Tensorized MRA)**.** *Let ψ, v^η be given as above and $1 \leq p < \infty$. Then,*

$$\|\psi_{b,l,j} - v^\eta_{b,l,j}\|_p \leq \eta.$$

Proof. From (4.6) and Lemma 4.17, we have $T_{b,d}(\psi_{b,l,j} - v^\eta_{b,l,j}) = e^p_{j_1} \otimes \cdots \otimes e^p_{j_l} \otimes (\psi - v^\eta)$. Then. Theorem 2.2 and the crossnorm property (see also [2]) imply

$$\|\psi_{b,l,j} - v^\eta_{b,l,j}\|_p = \|e^p_{j_1}\|_{\ell^p} \ldots \|e^p_{j_l}\|_{\ell^p} \|\psi - v^\eta\|_p = \|\psi - v^\eta\|_p \leq \eta.$$

□

Let $f \in L^p$ be an arbitrary function, $\varepsilon > 0$ and f_N an N-term wavelet approximation with $N = N(\varepsilon)$. I.e.,

$$\|f - f_N\|_p \leq \varepsilon, \quad f_N = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda,$$

where $\Lambda \subset \mathcal{J}$ is a generalized index set[3] with $\#\Lambda \leq N$. We assume that ψ_λ is normalized in L^p. Suppose each ψ_λ can be approximated via $v^\eta_\lambda \in V_{b,d,S}$ as in Lemma 4.23. This gives an approximation $v^\eta \in V_{b,d,S}$ to f via

$$v^\eta = \sum_{\lambda \in \Lambda} c_\lambda v^\eta_\lambda.$$

By Lemma 4.23, we can bound the error as

$$\|f_N - v^\eta\|_p \leq \sum_{\lambda \in \Lambda} |c_\lambda| \left\|\psi_\lambda - v^\eta_\lambda\right\|_p \leq \eta \sum_{\lambda \in \Lambda} |c_\lambda|. \tag{4.7}$$

A wavelet basis allows to characterize different norms of f by its coefficients in the basis, see [15] for unbounded domains and [12] for bounded domains.

In the following we assume that the wavelet basis $\Psi := \{\psi_\lambda : \lambda \in \mathcal{J}\}$ is a normalized basis of L^p. A function $f \in L^p$ admits a representation

$$f = \sum_{\lambda \in \mathcal{J}} c_\lambda \psi_\lambda.$$

Denoting by $|\lambda|$ the dilation level of ψ_λ, we can introduce an equivalent representation

$$f = \sum_{\lambda \in \mathcal{J}} c_{\lambda,p'} \psi_{\lambda,p'}, \quad c_{\lambda,p'} = b^{-|\lambda|(\frac{1}{p'} - \frac{1}{p})} c_\lambda, \quad \psi_{\lambda,p'} = b^{|\lambda|(\frac{1}{p'} - \frac{1}{p})} \psi_\lambda, \tag{4.8}$$

[3] E.g., each index is of the form $\lambda = (b, l, j)$.

where the functions $\psi_{\lambda,p'}$ are normalized in $L^{p'}$. Then, we assume that the wavelet basis has the property that for any function $f \in B^{\alpha}_{p',q}$ (see Definition 5.1 of Besov spaces), $0 < p', q \leq \infty$, $\alpha > 0$, it holds

$$|f|^q_{B^{\alpha}_{p',q}} \sim \sum_{l=0}^{\infty} b^{l\alpha q} \left(\sum_{\lambda \in \mathcal{J}_l} |c_{\lambda,p'}|^{p'} \right)^{q/p'} \tag{4.9}$$

where $\mathcal{J}_l$ is the subset of $\mathcal{J}$ consisting of only level $|\lambda| = l$ indices. The condition (4.9) is satisfied for wavelets with sufficient regularity and vanishing moments, see [12, 13, 15]. In fact, we do not need (4.9) to hold for any $0 < p, q \leq \infty$ as the next proposition shows.

Proposition 4.24 (Approximate N-Term Expansion)**.** *Let*

$$f = \sum_{\lambda \in \mathcal{J}} c_\lambda \psi_\lambda$$

be the wavelet expansion of $f \in L^p$ for $1 < p < \infty$, where the ψ_λ are normalized in L^p. Let v^η be as above, assume $f \in B^{\alpha}_{1,1}$ for $0 < \alpha < 1$ satisfying

$$\alpha + 1/p = 1, \tag{4.10}$$

and assume the characterization (4.9) *holds for $B^{\alpha}_{1,1}$.*
Then,

$$\|f_N - v^\eta\|_p \leq C\, |f|_{B^{\alpha}_{1,1}}\, \eta,$$

where $C = C(\Psi) > 0$ is a constant independent of N, ε, η or f.

Proof. Using (4.9) with $p' = q = 1$, (4.10) and then (4.8) with $p' = 1$, we obtain

$$|f|_{B^{\alpha}_{1,1}} \sim \sum_{\lambda} b^{|\lambda|\alpha} |c_{\lambda,1}| = \sum_{\lambda} b^{|\lambda|(1-\frac{1}{p})} |c_{\lambda,1}| = \sum_{\lambda} |c_\lambda|.$$

Then we conclude by using (4.7). □

The condition of excess regularity $f \in B^{\alpha}_{1,1}$ with $\alpha = 1 - 1/p$ can be replaced by excess integrability, at the cost of more complicated estimates, or excess summability of the coefficients c_λ. The ranks of v^η are bounded by $Nr_\nu(\eta)$, and $d(v)$ is the maximum level among all $d(v^\eta_\lambda)$ that depends only on η and ε. I.e., for $\eta \sim \varepsilon$, we can approximate functions $f \in L^p$ with the same precision as the wavelet system $\Psi := \{\psi_\lambda : \lambda \in \mathcal{J}\}$, where the overall order of the approximation depends on the approximation order of Ψ and how well Ψ is approximated by $V_{b,d,S}$.

Remark 4.25. *Similarly to Remark 4.16, the ranks of N-term approximations both for exact MRA and approximate MRA representations can be bounded by a multiple of N. This is a very crude estimate that assumes no specific structure of the wavelet approximation. Of course, the ranks occurring in practice may be much smaller, if we additionally impose restrictions on Λ, e.g., if we require that Λ has a tree structure (see [11]).*

5. Direct and Inverse Estimates

In this section, we discuss direct and inverse estimates for the approximation spaces defined in Section 2.5. Since we verified that N^{α}_q, C^{α}_q and S^{α}_q satisfy (P1) – (P4), we can use classical approximation theory (see [13, 16]) to show that an entire scale of interpolation and smoothness spaces is continuously embedded into these approximation classes. We begin by briefly reviewing Besov and interpolation spaces.

5.1. Besov Spaces.

Besov spaces provide a natural framework of smoothness for approximation theory since the measure of smoothness in the Besov scale is fine enough to adequately capture different approximation classes of functions. At the same time many other smoothness spaces such as Lipschitz spaces or (fractional) Sobolev spaces are special cases of Besov spaces.

In principle, there are two standard ways of measuring smoothness: by considering local rates of change of function values, or by decomposing a function into certain building blocks and considering the rates of decay of the high-frequency components. The original definition of Besov spaces [6] and the one we follow here is the former. The latter is also possible, see [34].

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

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

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

$$\omega_r(f,t)_p := \sup_{0<h\leq t} \left\|\Delta_h^r[f]\right\|_p, \quad t > 0. \tag{5.1}$$

Definition 5.1 (Besov Spaces)**.** *For parameters $\alpha > 0$ and $0 < p, q \leq \infty$, define $r := \lfloor \alpha \rfloor + 1$ and the Besov (quasi-)semi-norm as*

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

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*

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

In Definition 5.1, any r such that $r > \alpha$ defines the same space with equivalent norms. The primary parameters are α and p: α is the order of smoothness, while p indicates the measure of smoothness. The smaller p, the less restrictive the measure of smoothness. This is particularly important for free knot spline approximation[4], where frequently $p < 1$, which allows to measure smoothness for functions that would be otherwise too irregular in the scale of Sobolev spaces. The parameter q is secondary, allowing for a finer gradation of smoothness for the same primary parameters α and p.

Varying the parameters α, p and q, we have the obvious inclusions

$$\begin{aligned} B^\alpha_{p,q_1} &\subset B^\alpha_{p,q_2}, \quad q_1 \leq q_2, \\ B^\alpha_{p_1,q} &\subset B^\alpha_{p_2,q}, \quad p_1 \geq p_2, \\ B^{\alpha_1}_{p,q} &\subset B^{\alpha_2}_{p,q}, \quad \alpha_1 \geq \alpha_2. \end{aligned}$$

For $\alpha < 1$ and $q = \infty$, $B^\alpha_{p,\infty} = \mathrm{Lip}(\alpha, p)$, where the latter is the space of functions $f \in L^p$ such that

$$\|f(\cdot + h) - f\|_p \leq Ch^\alpha.$$

If α is an integer and $p \neq 2$, $B^\alpha_{p,\infty}$ is slightly larger than the *Sobolev space* $W^{\alpha,p}$. For non-integer α, the fractional Sobolev space $W^{\alpha,p}$ is the same as $B^\alpha_{p,p}$. For the special case $p = 2$ we even have $W^{\alpha,2} = B^\alpha_{2,2}$ for any $\alpha > 0$.

To visualize the relationship between the different spaces, we refer to the DeVore diagram in Figure 1. The x-axis corresponds to $1/p$, where p is the integrability parameter and primary measure of smoothness. The y-axis corresponds to the smoothness parameter α. A point on this plot represents a space. E.g., the point $(1/2, 0)$ is the space L^2, $(0,0)$ is L^∞, $(1/2, 1)$ is $W^{1,2}$ and so on. Of particular importance is the *Sobolev embedding line*: points above this line are embedded in L^p, points on this line may or may not be embedded into L^p, and points below this line are never embedded in L^p. We cite the following important result for later reference.

Theorem 5.2 ([16, Chapter 12, Theorem 8.1])**.** *Let $\alpha > 0$, $0 < p < \infty$ and define the* Sobolev embedding number τ

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

Then,

$$B^\alpha_{\tau,\tau} \hookrightarrow B^\alpha_{\tau,p} \hookrightarrow L^p.$$

[4]Which can be viewed as the theoretical idealization of adaptive approximation.

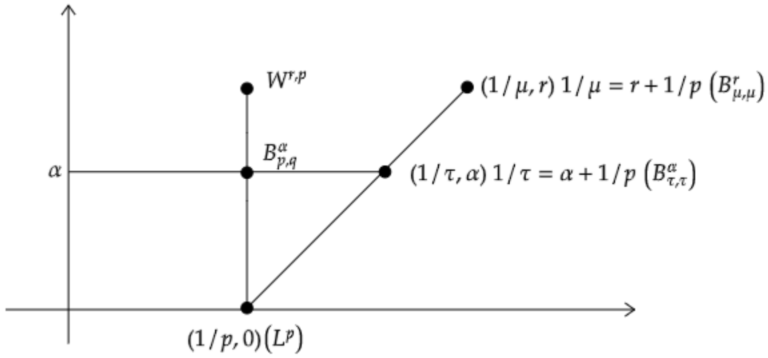

FIGURE 1. DeVore diagram of smoothness spaces [13]. The Sobolev embedding line is the diagonal with the points $(1/\tau, \alpha)$ and $(1/\mu, r)$

.

5.2. **Interpolation Spaces.** Given two linear spaces X and Y, the theory of interpolation spaces allows to define an entire scale of linear spaces that are in some sense "in-between" X and Y. Moreover, this theory provides a convenient tool for extending results derived for X and Y to spaces in-between.

To be more specific, let X and Y be normed[5] linear spaces with $Y \hookrightarrow X$. Let T be any linear operator such that $T \in \mathcal{L}(X)$ and $T \in \mathcal{L}(Y)$, i.e., T maps boundedly X and Y onto itself. If for any normed linear space Z with $Y \hookrightarrow Z \hookrightarrow X$, we have $T \in \mathcal{L}(Z)$, then such Z is called an *interpolation space.*

In this work, we specifically consider one method of constructing such interpolation spaces: Peetre's K-functional. 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 5.3 (Interpolation Spaces, [5, 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)*

$$E_n(f) \le C n^{-r_{\mathrm{J}}} |f|_Y, \quad \forall f \in Y, \tag{5.2}$$

and *Bernstein (inverse)*

$$\|\varphi\|_{A_q^\alpha} \le C n^{r_{\mathrm{B}}} \|\varphi\|_X, \quad \forall \varphi \in \Phi_n, \tag{5.3}$$

inequalities, for some $r_{\mathrm{J}} > 0$ and $r_{\mathrm{B}} > 0$.

[5]Quasi-seminormed spaces would suffice as well. We omit this for simplicity.

Theorem 5.4 (Interpolation and Approximation, [16, Chapter 7], [5, Chapter 5]). *If the approximation class $A_q^\alpha(X)$ satisfies (P1), (P3), (P4) and the space Y satisfies the Jackson inequality* (5.2)*, then*

$$\begin{aligned}(X,Y)_{\alpha/r_{\mathrm{J}},q} &\hookrightarrow A_q^\alpha(X), \quad 0<\alpha<r_{\mathrm{J}},\ 0<q<\infty,\\ Y &\hookrightarrow A_\infty^{r_{\mathrm{J}}}(X).\end{aligned}$$

If the approximation class $A_q^\alpha(X)$ satisfies (P1) – (P6) and the space Y satisfies the Bernstein inequality (5.3)*, then*

$$\begin{aligned}A_q^\alpha(X) &\hookrightarrow (X,Y)_{\alpha/r_{\mathrm{B}},q}, \quad 0<\alpha<r_{\mathrm{B}},\\ A_\infty^{r_{\mathrm{J}}}(X) &\hookrightarrow Y.\end{aligned}$$

Theorem 5.5 (Interpolation and Smoothness, [13]). *The following identities hold:*

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

5.3. **Direct Estimates.** Theorems 5.4 and 5.5 allows us to characterize the approximation classes introduced in Section 2.5 (and studied in [2]) by classical smoothness and interpolation spaces, provided we can show for $X=L^p$ and $Y=B_{p,q}^\alpha$ the Jackson (5.2) and Bernstein (5.3) inequalities. As indicated earlier, the Jackson inequalities will follow from the preparations in Section 4. We will also show that Bernstein inequalities cannot hold. This is an expression of the fact that the spaces A_q^α are "too large" in the sense that they are not continuously embedded in any classical smoothness space.

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 (2.7) and the resulting approximating sets $\Phi_n^{\mathcal{N}}$, $\Phi_n^{\mathcal{C}}$ and $\Phi_n^{\mathcal{S}}$. The best approximation error for $0<p\le\infty$ is defined accordingly as

$$\begin{aligned}E_n^{\mathcal{N}}(f)_p &:= \inf_{\varphi\in\Phi_n^{\mathcal{N}}}\|f-\varphi\|_p\,,\\ E_n^{\mathcal{C}}(f)_p &:= \inf_{\varphi\in\Phi_n^{\mathcal{C}}}\|f-\varphi\|_p\,,\\ E_n^{\mathcal{S}}(f)_p &:= \inf_{\varphi\in\Phi_n^{\mathcal{S}}}\|f-\varphi\|_p\,,\end{aligned}\tag{5.4}$$

and the corresponding approximation classes N_q^α, C_q^α and S_q^α as in (2.8).

5.3.1. *Sobolev Spaces.* We will apply local interpolation from Definition 4.2 to approximate functions in Sobolev spaces $W^{r,p}$ for any $r\in\mathbb{N}$. These embeddings essentially correspond to embeddings of Besov spaces $B_{p,p}^\alpha$ into approximation spaces $N_q^\alpha(L^p)$, $C_q^\alpha(L^p)$ and $S_q^\alpha(L^p)$: i.e., the approximation error is measured in the same norm as smoothness[6]. To this end, we require

Lemma 5.6 (Re-Interpolation). *Let $f\in W^{\bar m+1,p}$, $1\le p\le\infty$ and $\bar m\ge m$. 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*

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

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

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

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

Proof. See Appendix B. □

With this we can show the direct estimate

Lemma 5.7 (Jackson Inequality for Sobolev Spaces). *Let $1\le p\le\infty$ and $r\in\mathbb{N}$. For any $f\in W^{r,p}$ we have*

$$\begin{aligned}E_n^{\mathcal{N}}(f)_p &\le Cn^{-2r}\|f\|_{W^{r,p}}\,,\\ E_n^{\mathcal{S}}(f)_p \le E_n^{\mathcal{C}}(f)_p &\le Cn^{-r}\|f\|_{W^{r,p}}\,,\end{aligned}\tag{5.5}$$

with constants C depending on r, m, b.

[6] Compare to the embeddings for RePU networks in [21].

Proof. Let $N := b^d$ and $r := \bar{m}+1 > m+1$ and fix some $f \in W^{r,p}$. The case $r \leq m+1$ 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 5.6 we have

$$\|f - \tilde{s}\|_p \leq C_1 \|f\|_{W^{r,p}} \left(b^{-dr} + b^{-\bar{d}(m+1)}\right) \tag{5.6}$$

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

$$\bar{d} := \left\lceil \frac{dr}{m+1} \right\rceil, \tag{5.7}$$

which yields

$$\|f - \tilde{s}\|_p \leq 2C_1 \|f\|_{W^{r,p}} N^{-r}. \tag{5.8}$$

From Proposition 4.10 and (5.7), we can estimate the complexity of $\tilde{s}$ as

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

with a constant C_2 depending on b, m and $\bar{m}$. Thus, inserting into (5.8), we obtain (5.5). For the case $\bar{m} \leq m$, the proof simplifies since we can represent s exactly and use Proposition 4.8. □

Remark 5.8. *One could extend the statement of Lemma 5.7 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 [17], as was done in [21, Theorem 5.5] for RePU networks. For $r > m+1$, one would have to additionally replace the interpolation operator of Definition 4.2 with the quasi-interpolation operator from [17].*

Theorem 5.4 and Lemma 5.7 imply

Theorem 5.9 (Direct Embedding for Sobolev Spaces)**.** *For any $r \in \mathbb{N}$ and $1 \leq p \leq \infty$, we have*

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

and for $0 < q \leq \infty$

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

Corollary 5.10. *Together with Theorem 5.5, this implies the statement of Main Result 3.1.*

5.3.2. *Besov Spaces.* 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 $B^\alpha_{p,p}$.

They 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 5.11 (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

$$\inf_{s \in \mathcal{S}^{b,N,\bar{m}}_{\mathrm{fr}}} \|f - s\|_p \leq C N^{-\alpha} |f|_{B^\alpha_{\tau,\tau}}. \tag{5.9}$$

Proof. See Appendix B. □

Remark 5.12. *In principle, Lemma 5.11 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 5.13 does not hold for $p = \infty$, such that overall we can show the direct estimate of Theorem 5.15 only for $p < \infty$.*

Lemma 5.13 (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 $\tilde{s} = \sum_{k\in\Lambda}^N \tilde{s}_k$ such that

$$\|f - \tilde{s}\|_p \leq 2^{1/p}\varepsilon \quad \text{with} \quad |I_k| > N^{-q}\|f\|_{p\delta}^{-pq}\varepsilon^{pq} =: \varrho(\varepsilon), \quad k \in \Lambda. \tag{5.10}$$

Proof. See Appendix B. □

Remark 5.14. *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 $f \in L^{p\delta}$.

Lemma 5.15 (Jackson Inequality for $B^\alpha_{\tau,\tau}$)**.** *Let $1 \leq p < \infty$, $0 < \tau < p$, $\alpha > 1/\tau - 1/p$, and assume $f \in B^\alpha_{\tau,\tau}$. Then, for any $\sigma > 0$, we obtain the direct estimates*

$$\begin{aligned} E_n^{\mathcal{N}}(f)_p &\leq C\,|f|_{B^\alpha_{\tau,\tau}}\, n^{-\frac{\alpha}{1+\sigma}}, \\ E_n^{\mathcal{C}}(f)_p &\leq C\,|f|_{B^\alpha_{\tau,\tau}}\, n^{-\frac{\alpha}{2+\sigma}}, \\ E_n^{\mathcal{S}}(f)_p &\leq C\,|f|_{B^\alpha_{\tau,\tau}}\, n^{-\frac{\alpha}{1+\sigma}}, \end{aligned} \tag{5.11}$$

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. See Appendix B. □

Theorem 5.16 (Direct Embedding for $B^\alpha_{\tau,\tau}$)**.** *Let $1 \leq p < \infty$, $0 < \tau < p$ and $r > 1/\tau - 1/p$. Then, for any $\sigma > 0$,*

$$B^r_{\tau,\tau} \hookrightarrow N_\infty^{r/(1+\sigma)}(L^p), \quad B^r_{\tau,\tau} \hookrightarrow C_\infty^{r/(2+\sigma)}(L^p), \quad B^r_{\tau,\tau} \hookrightarrow S_\infty^{r/(1+\sigma)}(L^p),$$

and

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

Proof. Follows from Theorem 5.4, Remark 5.14 and Lemma 5.15. □

5.3.3. *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., [16, 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 5.17 (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 $1 \leq p \leq \infty$*

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

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

Proof. See Appendix B. □

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

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

Let $\rho := \rho(f) > 1$ be such that $f : [0,1) \to \mathbb{R}$ has an analytic extension onto $D_\rho \subset \mathbb{C}$, but not onto any $D_{\tilde{\rho}}$ for $\tilde{\rho} > \rho$. Then,

$$E_n^{\mathcal{N}}(f)_\infty \leq C[\min(\rho, b^{(m+1)})]^{-n^{1/2}}, \tag{5.12}$$
$$E_n^{\mathcal{S}}(f)_\infty \leq E_n^{\mathcal{C}}(f)_\infty \leq C[\min(\rho, b^{(m+1)/b})]^{-n^{1/3}},$$

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

Proof. See Appendix B. □

Remark 5.19. *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* (5.12) *can be stated in the form as in [16, Chapter 7, Theorem 8.1] to explicitly include the case $\rho = \infty$.*
- *One can define classes of entire functions as in [16, 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 [24].*

5.4. **Inverse Estimates.** It is well known in tensor approximation of high-dimensional functions and approximation with neural networks (see [21]) that highly irregular functions can in some cases be approximated or even represented exactly with low or constant rank or complexity[7]. This fact is reflected in the lack of inverse estimates for tensorized approximation of one-dimensional functions as the next statement shows.

Theorem 5.20 (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 [33] and Figure 2.

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

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

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

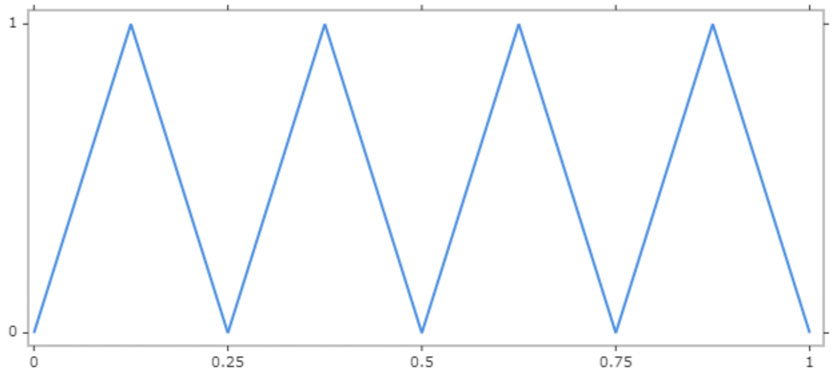

Figure 2. “Sawtooth” function.

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,

$$\mathrm{compl}_{\mathcal{C}}(\varphi_d) \leq 8d + 2m + 2. \tag{5.13}$$

We can compute the L^p-norm of φ_d as

$$\|\varphi_d\|_p^p = 2^d \int_0^{b^{-d}} (2^d y)^p \,\mathrm{d}y = \frac{1}{p+1}. \tag{5.14}$$

Next, since $C_q^\alpha(L^p)$ satisfies (P1) – (P4), this implies $C_q^\alpha(L^p)$ satisfies the Bernstein inequality (see [16, Chapter 7, Theorem 9.3])

$$\|\varphi_d\|_{C_q^\alpha} \leq C n^\alpha \|\varphi_d\|_p, \quad \forall \varphi \in \Phi_n. \tag{5.15}$$

On the other hand, by [21, Lemma 5.12],

$$\|\varphi_d\|_{B_{p,q}^{\tilde{\alpha}}} \geq c 2^{\tilde{\alpha} d}, \tag{5.16}$$

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

Assume the Bernstein inequality holds in $B_{p,q}^{\tilde{\alpha}}$ 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 (5.13), $\varphi_d \in \Phi_n^{\mathcal{C}}$. By (5.15) and (5.16),

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

Together with (5.14), 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 [32].

On the other hand, the above theorem demonstrates that tensor networks are efficient for functions that cannot be captured with classical smoothness (see also [1]). The cost n in $\Phi_n^{\mathcal{C}}$ is comprised of 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 5.20 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.

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.

Proposition 5.21 (Inverse Embedding for Restricted $\Phi_n^{\mathcal{N}}$)**.** *Let $1 \leq p < \infty$. Define for $n \in \mathbb{N}$, $r_{\mathrm{B}} > 0$ and $c_{\mathrm{B}} > 0$ the restricted sets*

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

Then,

(i) Φ_n^{B} *satisfies (P1) – (P6) and thus* $A_q^\alpha(L^p,(\Phi_n^{\mathrm{B}}))$ *are quasi-normed linear spaces satisfying direct and inverse estimates.*

(ii) The following inverse estimate holds:

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

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

(iii) We have the continuous embeddings

$$A_q^\alpha(L^p,(\Phi_n^{\mathrm{B}})) \hookrightarrow (L^p, B^{m+1}_{\tau,\tau})_{\frac{\alpha}{r_{\mathrm{B}}(m+1)},q}, \quad 0<\alpha<r_{\mathrm{B}}(m+1),$$
$$A_\infty^{r_{\mathrm{B}}(m+1)}(L^p,(\Phi_n^{\mathrm{B}})) \hookrightarrow B^{m+1}_{\tau,\tau}.$$

Proof. The restriction on Φ_n^{B} ensures functions such as the sawtooth function from Figure 2 are excluded.

(i) (P1) – (P3) is trivial. For (P4): since $\Phi_n^{\mathcal{N}} + \Phi_n^{\mathcal{N}} \subset \Phi_{cn}^{\mathcal{N}}$ and
$$d(\varphi_1+\varphi_2) \leq \max(d_1,d_2) \leq r_{\mathrm{B}}\log_b(n) \leq r_{\mathrm{B}}\log_b(cn)$$
for $\varphi_1,\varphi_2 \in \Phi_n^{\mathrm{B}}$, then (P4) holds for Φ_n^{B} for the same c. For (P5): $\bigcup_{n=0}^\infty \Phi_n^{\mathcal{N}} = \bigcup_{n=0}^\infty \Phi_n^{\mathrm{B}}$ and thus density follows as in [2, Theorem 2.25]. Finally, (P6) follows as in [2, Lemma 3.14].

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

(iii) Follows from (ii) and Theorem 5.4.

□

6. The Roles of Depth and Sparse Connectivity

One could ask how the direct estimates would change if we replace $\Phi_n^{\mathcal{N}}$ with Φ_n^{B} from Proposition 5.21. Strictly speaking, this would require lower bounds for the complexity $n := \mathrm{compl}_{\mathcal{N}}(\varphi)$. Nonetheless, a simple thought experiment 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 $W^{r,p}$ from Lemma 5.7 with $r \leq m+1$. Then, assuming the upper bounds from Lemma 5.7 are sharp, we have

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

I.e., the approximands of Lemma 5.7 satisfy $\varphi \in \Phi_n^{\mathrm{B}}$ for $r_{\mathrm{B}} = 1$ and $c_{\mathrm{B}} = c_{\mathrm{B}}(b,m)$. Hence, in this case we would indeed obtain the same approximation rate as with Φ_n, in addition to inverse estimates from Proposition 5.21.

Consider now $W^{r,p}$ with $r > m+1$. In this case we have $r_{\mathrm{B}} = r_{\mathrm{B}}(r) > 1$ with $r_{\mathrm{B}} \to \infty$ as $r \to \infty$. In other words, if we fix Φ_n^{B} with some $r_{\mathrm{B}} > 1$, then we would obtain direct estimates for $W^{r,p}$ as in Lemma 5.7, with $0 < r \leq \bar{r}$ for $\bar{r}$ depending on $r_{\mathrm{B}} > 1$. I.e., $\bar{r} = m+1$ for $r_{\mathrm{B}} = 1$ and $\bar{r} \to \infty$ as $r_{\mathrm{B}} \to \infty$.

Finally, consider the direct estimate for Besov spaces $B^{\alpha_\delta}_{\tau,\tau}$ from Lemma 5.15. 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 5.13, 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 Lemma 5.15. 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, for $\alpha \geq m+1$, Φ_n^{B} is not sufficient anymore as we require depth (and sparsity).

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 [20, 30] 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., [27, 28, 35].

On the other hand, sparse connectivity is necessary to recover classical adaptive (free knot spline) approximation, see Theorem 5.16. 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.

References

[1] M. Ali. Low-rank approximability and entropy area laws for ground states of unbounded hamiltonians, 2019.

[2] M. Ali and A. Nouy. Approximation with tensor networks. part i: Approximation spaces, 2020.

[3] Q. Ayoul-Guilmard, A. Nouy, and C. 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.

[4] M. Bachmayr, R. Schneider, and A. Uschmajew. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Foundations of Computational Mathematics*, 16(6):1423–1472, Apr. 2016.

[5] C. Bennett and R. Sharpley. *Interpolation of operators.* Academic Press, Boston, 1988.

[6] O. V. Besov. On a certain family of functional spaces. embedding and extension theorems. *Dokl. Akad. Nauk SSSR*, 126:1163–1165, 1959.

[7] W. Buczyńska. The hackbusch conjecture on tensor formats — part two. *Linear Algebra and its Applications*, 584:221–232, jan 2020.

[8] W. Buczyńska, J. Buczyński, and M. Michałek. The hackbusch conjecture on tensor formats. *Journal de Mathématiques Pures et Appliquées*, 104(4):749–761, oct 2015.

[9] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, and D. 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.

[10] A. Cichocki, A.-H. Phan, Q. Zhao, N. Lee, I. Oseledets, M. Sugiyama, and D. 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.

[11] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Tree approximation and optimal encoding. *Applied and Computational Harmonic Analysis*, 11(2):192–226, sep 2001.

[12] A. Cohen, W. Dahmen, and R. A. DeVore. Multiscale decompositions on bounded domains. *Transactions of the American Mathematical Society*, 352, 1997.

[13] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, jan 1998.

[14] R. A. DeVore, R. Howard, and C. Micchelli. Optimal nonlinear approximation. *Manuscripta Mathematica*, 63(4):469–478, dec 1989.

[15] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov. Hyperbolic wavelet approximation. *Constructive Approximation*, 14(1):1–26, jan 1997.

[16] R. A. DeVore and G. G. Lorentz. *Constructive Approximation.* Springer-Verlag Berlin Heidelberg, 1993.

[17] R. A. Devore and V. A. Popov. Interpolation of besov spaces. *Transactions of the American Mathematical Society*, 305(1):397–414, 1988.

[18] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements.* Springer Verlarg, 2004.

[19] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning.* 2016.

[20] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Technical report, Institut für Geometrie und Prakitsche Mathematik, RWTH Aachen, 2010.

[21] R. Gribonval, G. Kutyniok, M. Nielsen, and F. Voigtlaender. Approximation spaces of deep neural networks. 2019.

[22] S. Haykin. *Neural networks and learning machines.* Prentice Hall/Pearson, New York, 2009.

[23] D. C. H. J. C. Mason. *Chebyshev Polynomials.* Taylor & Francis Ltd, 2002.

[24] V. Kazeev and C. Schwab. Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions. *Numerische Mathematik*, 138(1):133–190, jul 2017.

[25] S. Mallat. *A Wavelet Tour of Signal Processing.* Elsevier LTD, Oxford, 2009.

[26] A. Nouy. *Low-Rank Methods for High-Dimensional Approximation and Model Order Reduction*, chapter 4. SIAM, Philadelphia, PA, 2017.

[27] J. A. Opschoor, C. Schwab, and J. Zech. Exponential relu dnn expression of holomorphic maps in high dimension. Technical report, Zurich, 2019-07.

[28] J. A. A. Opschoor, P. C. Petersen, and C. Schwab. Deep ReLU networks and high-order finite element methods. *Analysis and Applications*, pages 1–56, feb 2020.

[29] R. Orús. Tensor networks for complex quantum systems. *Nature Reviews Physics*, 1(9):538–550, aug 2019.

[30] I. V. Oseledets. Constructive representation of functions in low-rank tensor formats. *Constructive Approximation*, 37(1):1–18, dec 2012.

[31] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[32] R. Schneider and A. Uschmajew. Approximation rates for the hierarchical tensor format in periodic sobolev spaces. *Journal of Complexity*, 30(2):56–71, apr 2014.

[33] M. Telgarsky. Representation benefits of deep feedforward networks, 2015.

[34] H. Triebel. *Theory of Function Spaces II.* Springer Basel, 1992.

[35] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103–114, oct 2017.

[36] D. Yarotsky. Optimal approximation of continuous functions by very deep relu networks, 2018.

Appendix A. 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 3.

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.

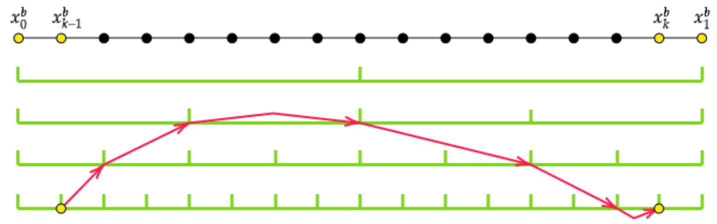

Figure 3. Visual representation of different partitioning levels of the interval $[0,1)$, with $b=2$ and $d=4$.

In Figure 3, 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 [2, Lemmas 3.11 and 3.12], we deduce that

$$\mathrm{compl}_{\mathcal{S}}(\varphi) \leq \sum_{k=1}^N \sum_{i=1}^{n_k} b(d_{k,i} + \dim S) + b^2(\dim S)^3(d - d_{k,i}) \leq 2b^2(\dim S)^3 d \sum_{k=1}^N n_k \leq 4b^3(m+1)^3 d^2 N.$$

□

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

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

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 (2.5), we have that

$$
\begin{aligned}
&T_{b,\bar{d}}(\mathcal{I}_{b,\bar{d},m}(\varphi))(i_1,\ldots,i_{\bar{d}},y)\\
&= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\bar{m}+1} v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d) v_{d+1}^{k_d,q} T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1}))(i_{d+1},\ldots,i_{\bar{d}},y)\\
&= \sum_{k_1=1}^{r_1}\cdots\sum_{k_d=1}^{r_d}\sum_{q=1}^{\bar{m}+1}\sum_{j_{d+1}=1}^{b} v_1^{k_1}(i_1)\cdots v_d^{k_{d-1},k_d}(i_d)\bar{v}_{d+1}^{k_d,(q,j_{q+1})}(i_{d+1}) T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q^{\bar{m}+1}))(j_{d+1},\ldots,i_{\bar{d}},y),
\end{aligned}
$$

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 [2, 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

$$
\begin{aligned}
&T_{b,\bar{d}-d}(\mathcal{I}_{b,\bar{d}-d,m}(\varphi_q))(j_{d+1},i_{d+2},\ldots,i_{\bar{d}},y)\\
&= \sum_{\alpha_2,q_2=1}^{\bar{m}+1}\cdots\sum_{\alpha_l,q_l=1}^{\bar{m}+1}\sum_{p=1}^{m+1} \bar{v}_{d+2}^{(q,j_{d+1}),(q_2,\alpha_2)}(i_{d+2})\ldots \bar{v}_{\bar{d}}^{(q_{l-1},\alpha_{l-1}),(q_l,\alpha_l)}(i_{\bar{d}})\bar{v}_{\bar{d}+1}^{(q_l,\alpha_l),p}\varphi_p(y)
\end{aligned}
$$

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 $\mathcal{I}_{b,\bar{d},m}(\varphi) = R_{b,\bar{d},m,\bar{\boldsymbol{r}}}(\bar{\mathbf{v}})$ with $\bar{\mathbf{v}} = (\bar{v}_1,\ldots,v_d,\bar{v}_{d+1},\ldots,\bar{v}_{\bar{d}+1})$ 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 $\bar{d}-d = 1$, $\varphi \in R_{b,\bar{d},S,\bar{\boldsymbol{r}}}(\bar{\mathbf{v}})$ with $\bar{\mathbf{v}} = (\bar{v}_1,\ldots,v_d,\bar{v}_{d+1},\bar{v}_{\bar{d}+2})$, and

$$
\mathrm{compl}_{\mathcal{S}}(\mathcal{I}_{b,\bar{d},m}(\varphi)) \leq b\,\mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b(\bar{m}+1)(m+1) \leq \max\{b,m+1\}(\mathrm{compl}_{\mathcal{S}}(\mathbf{v}) + b(\bar{m}+1)(\bar{d}-d)).
$$

Finally for $\bar{d} = d$, we simply have $\mathcal{I}_{b,\bar{d}-d,m} = \mathcal{I}_m$, and we can show that $\mathcal{I}_{b,\bar{d},m}(\varphi) = R_{b,d,m,\boldsymbol{r}}(v_1,\ldots,v_d,\bar{v}_{d+1})$ 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 $\mathrm{compl}_{\mathcal{S}}(\varphi) \leq C_1 N$. □

Appendix B. Proofs for Section 5

Proof of Lemma 5.6. From (2.5), we know that $s := \mathcal{I}_{b,d,\bar{m}}f$ admits a tensorization $\boldsymbol{s} := T_{b,d}s = (id_{\{1,\ldots,d\}} \otimes \mathcal{I}_{\bar{m}})\boldsymbol{f}$, with $\boldsymbol{f} = T_{b,d}f$. 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 2.2, we have for $p < \infty$

$$\|f - s\|_p^p = \sum_{j \in I_b^d} b^{-d} \|g_j - \mathcal{I}_{\bar{m}} g_j\|_p^p \leq C^p \sum_{j \in I_b^d} b^{-d} \|D^{\bar{m}+1} g_j\|_p^p = C^p \|(id_{\{1,\ldots,d\}} \otimes D^{\bar{m}+1}) \boldsymbol{f}\|_p^p$$

and

$$\|f - s\|_\infty = \max_{j \in I_b^d} \|g_j - \mathcal{I}_{\bar{m}} g_j\|_\infty \leq C \max_{j \in I_b^d} \|D^{\bar{m}+1} g_j\|_\infty \leq C \|(id_{\{1,\ldots,d\}} \otimes D^{\bar{m}+1}) \boldsymbol{f}\|_\infty.$$

Then, from [2, Theorem 2.15] we deduce

$$\|f - s\|_p \leq C \|(id_{\{1,\ldots,d\}} \otimes D^{\bar{m}+1}) \boldsymbol{f}\|_p = C b^{-d(\bar{m}+1)} \|f\|_{W^{\bar{m}+1,p}}.$$

For $\bar{d} \geq d$, we obtain from [2, Lemma 2.6] and (2.5) that $T_{b,d} \mathcal{I}_{b,\bar{d},m} T_{b,d}^{-1} = T_{b,d} T_{b,\bar{d}}^{-1} (id_{\{1,\ldots,\bar{d}\}} \otimes \mathcal{I}_m) T_{b,\bar{d}} T_{b,d}^{-1} = id_{\{1,\ldots,d\}} \otimes (T_{b,\bar{d}-d} (id_{\{1,\ldots,\bar{d}-d\}} \otimes \mathcal{I}_m) T_{b,\bar{d}-d}^{-1}) = id_{\{1,\ldots,d\}} \otimes \mathcal{I}_{b,\bar{d}-d,m}$. Then $\tilde{s} := \mathcal{I}_{b,\bar{d},m} s$ admits for tensorization $T_{b,d} \tilde{s} = (id_{\{1,\ldots,d\}} \otimes \mathcal{I}_{b,\bar{d}-d,m}) \boldsymbol{s} = \sum_{j \in I_b^d} \delta_{j_1} \otimes \ldots \otimes \delta_d \otimes (\mathcal{I}_{b,\bar{d}-d,m} \mathcal{I}_{\bar{m}} g_j)$, 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

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

In the same way as above, we deduce

$$\begin{aligned} \|s - \tilde{s}\|_p &\leq \tilde{C} b^{-(\bar{d}-d)(m+1)} \left(\|(id_{\{1,\ldots,d\}} \otimes D^{m+1}) \boldsymbol{f}\|_p + C \|(id_{\{1,\ldots,d\}} \otimes D^{\bar{m}+1}) \boldsymbol{f}\|_p \right) \\ &= C' b^{-(\bar{d}-d)(m+1)} (b^{-d(m+1)} \, |f|_{W^{m+1,p}} + b^{-d(\bar{m}+1)} \, |f|_{W^{\bar{m}+1,p}}), \end{aligned}$$

with $C' = \tilde{C} \max\{1, C\}$ depending on m, $\bar{m}$ and p, which completes the proof. □

Proof of Lemma 5.11. The proof is a modification of the proof of Petrushev for free knot splines (see [16, 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 $r := \lfloor \alpha \rfloor + 1$ and

$$M^\tau := \int_0^1 t^{-\alpha\tau - 1} \mathrm{w}_r(f, t)_\tau^\tau \, \mathrm{d}t,$$

where w_r is the *averaged modulus of smoothness*, i.e.,

$$\mathrm{w}_r(f, t)_\tau^\tau := \frac{1}{t} \int_0^t \|\Delta_h^r[f]\|_\tau^\tau \, \mathrm{d}h.$$

By [16, Chapters 2 and 12], M is equivalent to $|f|_{B^\alpha_{\tau,\tau}}$.

Let

$$g(x, h, t) := \begin{cases} t^{-\alpha\tau - 2} |\Delta_h^r[f](x)|^\tau & \text{if } h \in [0, t] \text{ and } x \in [0, 1 - rh], \\ 0 & \text{elsewhere.} \end{cases}$$

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_k, $k = 1, \ldots, N$, with disjoint interiors such that

$$\bigcup_{k=1}^{N} I_k = [0,1] \quad \text{and} \quad \int_{I_k} G(x)\,\mathrm{d}x = \frac{M^\tau}{N}.$$

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 $k = 1$, 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

$$\int_{I_1^\varepsilon} G(x)\,\mathrm{d}x \le \frac{M^\tau}{N} \le \int_{I_1^\varepsilon} G(x)\,\mathrm{d}x + \frac{\varepsilon}{N}. \tag{B.1}$$

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 (B.1) is satisfied for I_2^ε. Repeating this procedure until I_{N-1}^ε we get

$$\int_{\bigcup_{k=1}^{N-1} I_k^\varepsilon} G(x)\,\mathrm{d}x \le \frac{N-1}{N} M^\tau \le \int_{\bigcup_{k=1}^{N-1} I_k^\varepsilon} G(x)\,\mathrm{d}x + \frac{N-1}{N}\varepsilon.$$

Taking I_N^ε as the remaining interval such that $\bigcup_{k=1}^N I_k^\varepsilon = [0,1]$, we have

$$\int_{\bigcup_{k=1}^{N} I_k^\varepsilon} G(x)\,\mathrm{d}x = M^\tau.$$

For the last interval we see that

$$\int_{I_N^\varepsilon} G(x)\,\mathrm{d}x \ge \frac{M^\tau}{N},$$

and

$$\int_{I_N^\varepsilon} G(x)\,\mathrm{d}x = M^\tau - \int_{\bigcup_{k=1}^{N-1} I_k^\varepsilon} G(x)\,\mathrm{d}x \le M^\tau - \frac{N-1}{N}\Big(M^\tau - \varepsilon\Big) \le \frac{1}{N} M^\tau + \varepsilon.$$

Finally, we apply polynomial approximation over each I_k^ε. There exist polynomials P_k of degree $\le \bar{m}$ over each I_k^ε such that for $f_k := f|_{I_k^\varepsilon}$ (see [16, Chapter 12, Theorem 8.1])

$$\|f_k - P_k\|_p^\tau (I_k^\varepsilon) \le C^\tau |f_k|^\tau_{B^\alpha_{\tau,\tau}} (I_k^\varepsilon) \le C' \int_{I_k^\varepsilon} G(x)\,\mathrm{d}x \le C' \begin{cases} \frac{1}{N} M^\tau, & k = 1, \ldots, N-1, \\ \frac{1}{N} M^\tau + \varepsilon, & k = N, \end{cases}$$

where $\|\cdot\| (I_k^\varepsilon)$ means we take norms over I_k^ε only. Setting $s = \sum_{k=1}^N P_k \mathbb{1}_{I_k^\varepsilon}$ and since $p/\tau > 1$, we obtain

$$\begin{aligned}\|f - s\|_p^p &= \sum_{k=1}^{N} \|f_k - P_k\|_p^p (I_k^\varepsilon) \le (N-1) C M^p N^{-p/\tau} + \left(\frac{1}{N} M^\tau + \varepsilon\right)^{p/\tau} \\ &\le (N-1) C M^p N^{-p/\tau} + 2^{p/\tau - 1}\left((\frac{1}{N} M^\tau)^{p/\tau} + \varepsilon^{p/\tau}\right) \le \max\left\{C, 2^{p/\tau-1}\right\}\left(M^p N^{1-p/\tau} + \varepsilon^{p/\tau}\right).\end{aligned}$$

Since the constant is independent of ε and ε can be chosen arbitrarily small, we obtain (5.9). □

Proof of Lemma 5.13. Let $f_k := f \mathbb{1}_{I_k}$. By the Hölder inequality

$$\|f_k\|_p^p = \int_0^1 |f_k(x)|^p\,\mathrm{d}x \le \left(\int_0^1 |f(x)|^{p\delta}\,\mathrm{d}x\right)^{1/\delta} \left(\int_{I_k} \mathrm{d}x\right)^{1/q}.$$

We choose

$$\Lambda := \{k = 1, \ldots, N : \; |I_k| > \varrho(\varepsilon)\}.$$

Then,

$$\sum_{k \notin \Lambda} \|f_k\|_p^p \le \|f\|_{p\delta}^p \, N \varrho(\varepsilon)^{1/q} \le \varepsilon^p.$$

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 5.15. As in Lemma 5.7, we consider only the case $m+1 \leq \alpha$, as the case $\alpha < m+1$ can be handled analogously with fewer steps. By Lemma 5.11, we can restrict ourselves to free knot splines with b-adic knots. By Lemma 5.13, 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 (5.11).

Let $N \in \mathbb{N}$ be arbitrary. From Lemma 5.11, we know there exists a spline $s \in \mathcal{S}_{\mathrm{fr}}^{b,N,\bar{m}}$ with b-adic knots such that

$$\|f-s\|_p \leq C_1 N^{-\alpha} \|f\|_{B^\alpha_{\tau,\tau}}, \tag{B.2}$$

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 5.13, 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

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

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 $\bar{d} \geq d$ 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 5.6

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

where the latter follows from [16, 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 [16, 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 $j \in I_b^d$. Together with [2, Proposition 2.19], we finally estimate

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

Thus, to obtain at least the same approximation order as in (B.2), we set

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

so that

$$\|s-\tilde{s}\|_p \leq C_5 N^{-\alpha}|f|_{B^\alpha_{\tau,\tau}}. \tag{B.3}$$

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

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

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

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

and

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

for any $\sigma > 0$ and constants C depending on $\sigma > 0$. Combining (B.2) with (B.3), a triangle inequality and the above complexity bounds, we obtain the desired statement. □

Proof of Lemma 5.17. Let $P \in \mathbb{P}_{\bar{m}}$ be arbitrary and set $s := \mathcal{I}_{b,d,m}P$. From Lemma 5.6, we obtain

$$\|P - s\|_p \leq C_1 b^{-d(m+1)} \left\|P^{(m+1)}\right\|_p. \tag{B.4}$$

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 (B.4)

$$\|P - s\|_p \leq C_2 b^{-\frac{m+1}{b(\bar{m}+1)^2}n} \left\|P^{(m+1)}\right\|_p.$$

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 5.18. Set

$$M := \sup_{z \in D_\rho} |f(z)|,$$

and $\bar{m} \in \mathbb{N}$. From [16, Chapter 7, Theorem 8.1], we know

$$\inf_{P \in \mathbb{P}_{\bar{m}}} \|f - P\|_\infty \leq \frac{2M}{\rho - 1}\rho^{-\bar{m}}. \tag{B.5}$$

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 (B.4) we know

$$\|P - s\|_\infty \leq C_1 b^{-d(m+1)} \left\|P^{(m+1)}\right\|_\infty, \tag{B.6}$$

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 [16, Chapter 7, Theorem 8.1] used to derive (B.5). 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

$$P_{\mathrm{Ch}} := \frac{1}{2}a_0 + \sum_{k=1}^{\bar{m}} a_k C_k, \tag{B.7}$$

which is such that

$$\|f - P_{\mathrm{Ch}}\|_p \leq \frac{2M}{\rho - 1}\rho^{-\bar{m}}. \tag{B.8}$$

For the derivatives of $C_k^{(m+1)}$ we get by standard estimates (see, e.g., [23])

$$\left\|C_k^{(m+1)}\right\|_\infty \leq \frac{k^2(k^2-1)\cdots(k^2-m^2)}{(2(m+1)-1)!}.$$

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 is a sum of poly-logarithmic series, and thus it 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

$$d := \left\lfloor b^{-1} n^{1/3} - (m+1) n^{-2/3} \right\rfloor > 1,$$
$$\bar{m} := \lfloor n^{1/3} - 1 \rfloor \geq 1.$$

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 (B.7). 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 (B.6) and (B.8),

$$E_n^{\mathcal{C}}(f)_\infty \leq \|f - P_{\mathrm{Ch}}\|_\infty + \|P_{\mathrm{Ch}} - s\|_\infty \leq \frac{2M}{\rho-1}\rho^{-\bar{m}} + C_1' b^{-d(m+1)} \left\|P_{\mathrm{Ch}}^{(m+1)}\right\|_\infty$$
$$\leq C_2' [\min(\rho, b^{(m+1)/b})]^{-n^{1/3}}.$$

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 (B.6) and (B.8),

$$E_n^{\mathcal{N}}(f)_\infty \leq \frac{2M}{\rho-1}\rho^{-\bar{m}} + C_1' b^{-d(m+1)} \left\|P_{\mathrm{Ch}}^{(m+1)}\right\|_\infty \leq C_2' [\min(\rho, b^{(m+1)})]^{-n^{1/2}}.$$

□