SPECTRAL BARRON SPACE FOR DEEP NEURAL NETWORK APPROXIMATION

YULEI LIAO AND PINGBING MING

ABSTRACT. We prove the sharp embedding between the spectral Barron space and the Besov space with embedding constants independent of the input dimension. Given the spectral Barron space as the target function space, we prove a dimension-free convergence result that if the neural network contains L hidden layers with N units per layer, then the upper and lower bounds of the L^2-approximation error are $\mathcal{O}(N^{-sL})$ with $0 < sL \leq 1/2$, where $s \geq 0$ is the smoothness index of the spectral Barron space.

1. INTRODUCTION

A series of works have been devoted to studying the neural network approximation error and generalization error with the Barron class [Bar92, Bar93, Bar94, KB18] as the target function space. For f a complex-valued function and $s \geq 0$, the spectral norm $\upsilon_{f,s}$ is defined as

$$\upsilon_{f,s} := \int_{\mathbb{R}^d} |\xi|^s |\widehat{f}(\xi)| \mathrm{d}\xi,$$

where $\widehat{f}$ is the Fourier transform of f in the sense of distribution. A function f is said to belong to the Barron class if the spectral norm $\upsilon_{f,s}$ is finite, and the pointwise Fourier inversion holds true. Nonetheless, it is vital to note that this definition lacks rigor, as it fails to specify the conditions under which the pointwise Fourier inversion is valid. Addressing this issue is a nontrivial matter, as has been discussed in [PT97]. Subsequently, the authors in [LGM+17, Xu20, SX22, SX23]

assume $f \in L^1(\mathbb{R}^d)$ and $s \geq 0$, and define

(1.1) $$\widehat{\mathscr{B}}^s(\mathbb{R}^d) := \left\{ f \in L^1(\mathbb{R}^d) \mid v_{f,0} + v_{f,s} < \infty \right\}.$$

For functions in $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$, the Fourier transform and the pointwise Fourier inversion are valid. Unfortunately, we shall prove in Lemma 2.3 that $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$ equipped with the norm $v_{f,0} + v_{f,s}$ is not complete, and thus, it is not a Banach space.

To tackle this issue, an alternative function spaces has been proposed, which roots tracing back to the seminar work of Hörmander [Hör63]. It is defined as follows:

$$\mathscr{F}L_p^s(\mathbb{R}^d) := \left\{ f \in \mathscr{S}'(\mathbb{R}^d) \mid (1 + |\xi|^s)\widehat{f}(\xi) \in L^p(\mathbb{R}^d) \right\}$$

for $1 \leq p \leq \infty$ and $s \geq 0$. This space has been studied extensively and bears various names. It is sometimes called the Hörmander space, as mentioned in works such as [Hör63, MR01, MM14, IV21]; alternatively, it may be referred to the Fourier Lebesgue space, as evidenced in the works such as [GG02, PTT10, BO13, KST20]. We focus on $p = 1$ and $s \geq 0$, and denote it as the spectral Barron space:

$$\mathscr{B}^s(\mathbb{R}^d) := \left\{ f \in \mathscr{S}'(\mathbb{R}^d) \mid v_{f,0} + v_{f,s} < \infty \right\},$$

which is equipped with the norm

$$\| f \|_{\mathscr{B}^s(\mathbb{R}^d)} := v_{f,0} + v_{f,s} = \int_{\mathbb{R}^d} (1 + |\xi|^s)|\widehat{f}(\xi)|\mathrm{d}\xi.$$

We show in Lemma 2.1 that the pointwise Fourier inversion is valid for functions in $\mathscr{B}^s(\mathbb{R}^d)$ with a nonnegative s. Some authors also refer to $\mathscr{B}^s(\mathbb{R}^d)$ as the Fourier algebra or Wiener algebra, whose algebraic properties, such as the Wiener-Levy theorem [Wie32, Lév35, HKKR59], have been extensively studied in [RS00, LST12].

Another popular space for analyzing shallow neural networks is the Barron space introduced in [EMW19, EMW22], which can be viewed as shallow neural networks with infinite width. Recent works such as [EW22a, EW22b] claimed that the spectral Barron space is considerably smaller than the Barron space. However, as pointed out in [CPV23], this claim lacks accuracy because they have not discriminated the smoothness index s in $\mathscr{B}^s(\mathbb{R}^d)$. In addition, the variation space, initially introduced in [BCDD08], has been studied in relation to the spectral Barron space $\mathscr{B}^s(\mathbb{R}^d)$ and the Barron space in [SX24, SX23]. These spaces have been exploited to study the regularity of partial differential equations [CLL21, LLW21, EW22b, CLLZ23]. Recently a novel space, originating from variational spline theory, as discussed in [PN22], which is closely related to the variation space, has emerged as a target function space for neural network approximation [PN23].

The first objective of the present work is the analytical properties of $\mathscr{B}^s(\mathbb{R}^d)$. In Lemma 2.2, we show that $\mathscr{B}^s(\mathbb{R}^d)$ is complete, while Lemma 2.3 shows that $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$ is not complete. This distinction highlights a key difference between these two spaces. Furthermore, Lemma 2.5 provides an example that illustrates functions in $\mathscr{B}^s(\mathbb{R}^d)$ may decay arbitrarily slow. This elegantly constructed example, utilizing

the generalized Hypergeometric function, unveils intriguing relationships between the Fourier transform and the decay rate of the functions. Furthermore, we study the relations between $\mathscr{B}^s(\mathbb{R}^d)$ and some classical function spaces. In Theorem 2.9, we establish the connections between $\mathscr{B}^s(\mathbb{R}^d)$ and the Besov space. Moreover, in Corollary 2.13, we establish the connections between $\mathscr{B}^s(\mathbb{R}^d)$ and the Sobolev spaces.

Notably, we prove the embedding relation

$$B_{2,1}^{s+d/2}(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d) \hookrightarrow B_{\infty,1}^s(\mathbb{R}^d),$$

which is an optimal result seemingly absent from the existing literature. Roughly speaking, the optimality is understood in the sense that $B_{2,1}^{s+d/2}(\mathbb{R}^d)$ is the largest Besov space contained in $\mathscr{B}^s(\mathbb{R}^d)$, while $B_{\infty,1}^s(\mathbb{R}^d)$ is the smallest Besov space that contains $\mathscr{B}^s(\mathbb{R}^d)$. We refer to Thoerem 2.9 for a precise statement. Moreover, the embedding constants are independent of the input dimension d, which indicates that the embedding is effective in high dimension. This embedding may serve as a bridge to study how the Barron space, the variation space and the space in [PN22] are related to the classical function spaces such as the Besov space. To the best of our knowledge, only a few studies have addressed the embedding relationships between neural network-related spaces and classical function spaces, such as [GKNV22, GV23], while the scope of their investigations has not been comprehensive.

The second objective of this work is to explore the neural network approximation on a bounded domain. Building upon Barron's seminal works on approximating functions in $\mathscr{B}^1(\mathbb{R}^d)$ with L^2-norm, recent studies have extended the approximation to functions in $\mathscr{B}^{k+1}(\mathbb{R}^d)$ with H^k-norm, as demonstrated in [SX20, Xu20]. Furthermore, improved approximation rates have been achieved for functions in $\mathscr{B}^s(\mathbb{R}^d)$ with large s in works such as [BN20a, MSX22, SX22]. These advancements contribute to a deeper understanding of the approximation capabilities of neural networks.

The distinction between deep ReLU networks and shallow networks has been highlighted in the separation theorems presented in [ES16, Tel16, SES22, GIJK23]. These theorems offer examples that can be well approximated by deep networks but not by shallow networks without the curse of dimensionality. This sheds light on the differences in the expressive power between the shallow and deep neural networks. Moreover, the approximation rates for neural networks targeting mixed derivative Besov/Sobolev spaces, spectral Barron spaces, and Hölder spaces have also been investigated. These studies contribute to a broader understanding of the approximation capabilities of neural networks in various function spaces as in [MD19, BN20b, EPGB21, LSYZ21, TS21]. Additional works focusing on deep neural network approximation in novel spaces may be founded in [SH20, EW22b, GKNV22].

We focus on the L^2-approximation properties for functions in $\mathscr{B}^s(\mathbb{R}^d)$ when s is small. In Theorem 3.5, we establish that a neural network with L hidden layers and N units in each layer can approximate functions in $\mathscr{B}^s(\mathbb{R}^d)$ with a convergence rate of $\mathcal{O}(N^{-sL})$ when $0 < sL \leq 1/2$. This bound is sharp, as proved in Theorem 3.11. Our results provide optimal convergence rates compared to existing literature. For deep neural networks, a similar result has been presented in [BN20b] with a convergence rate of $\mathcal{O}(N^{-sL/2})$. For shallow neural network; i.e., $L = 1$, convergence rates of $\mathcal{O}(N^{-1/2})$ have been established in [MM22, SX22] when $s = 1/2$. However, it is worth noting that the constants in their estimates depend on the dimension at least polynomially, or even exponentially, and require other bounded norms besides $v_{f,s}$. Our results provide a significant advancement by achieving optimal convergence rates without the additional dependency on dimension or other bounded norms.

The remaining part of the paper is structured as follows. In Section 2, we demonstrate that the spectral Barron space is a Banach space and examine its relationship with other function spaces. This analysis provides a foundation for understanding the properties of the spectral Barron space. In Section 3, we delve into the error estimation for approximating functions in the spectral Barron space using deep neural networks with finite depth and infinite width. By investigating the convergence properties of these networks, we gain insights into their approximation capabilities and provide error bounds for their performance. Finally, in Section 4, we conclude our work by summarizing the key findings and contributions of this study. We also discuss potential avenues for future research and highlight the significance of our results in the broader context of function approximation using neural networks. Certain technical results that are tedious but not the main focusing have been postponed to the Appendix.

2. Completeness of $\mathscr{B}^s$ and its relation to other function spaces

This part discusses the completeness of the spectral Barron space and embedding relations to other classical function spaces. Firstly, we fix some notations. Let $\mathscr{S}$ be the Schwartz space and let $\mathscr{S}'$ be its topological dual space, i.e., the space of tempered distribution. The Gamma function

$$\Gamma(s) := \int_0^\infty t^{s-1} e^{-t} \mathrm{d}t, \qquad s > 0.$$

Denoting the surface area of the unit sphere $\mathbb{S}^{d-1}$ by $\omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$. The volume of the unit ball is $\nu_d = \omega_{d-1}/d$. The Beta function

$$B(\alpha, \beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \qquad \alpha, \beta > 0.$$

The series formulation of the first kind of Bessel function is defined as

$$J_\nu(x) := (x/2)^\nu \sum_{k=0}^\infty (-1)^k \frac{(x/2)^{2k}}{\Gamma(\nu+k+1)k!}.$$

This definition may be found in [Luk62, § 1.4.1, Eq. (1)].

For $f \in L^1(\mathbb{R}^d)$, its Fourier transform of f is defined as

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{-2\pi i x \cdot \xi} \mathrm{d}x,$$

and the inverse Fourier transform is defined as

$$f^\vee(x) := \int_{\mathbb{R}^d} f(\xi) e^{2\pi i x \cdot \xi} \mathrm{d}\xi.$$

If $f \in \mathscr{S}'(\mathbb{R}^d)$, then the Fourier transform in the sense of distribution means

$$\langle \widehat{f}, \varphi \rangle = \langle f, \widehat{\varphi} \rangle \qquad \text{for any} \quad \varphi \in \mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d).$$

We shall frequently use the following Hausdorff-Young inequality. Let $1 \le p \le 2$ and $f \in L^p(\mathbb{R}^d)$, then

$$(2.1) \qquad \|\widehat{f}\|_{L^{p'}(\mathbb{R}^d)} \le \|f\|_{L^p(\mathbb{R}^d)},$$

where p' is the conjugate exponent of p; i.e. $1/p + 1/p' = 1$.

We shall use the following pointwise Fourier inversion theorem.

Lemma 2.1. *Let* $g \in L^1(\mathbb{R}^d)$, *then* $\widehat{g^\vee} = g$ *in* $\mathscr{S}'(\mathbb{R}^d)$. *Furthermore, let* $f \in \mathscr{S}'(\mathbb{R}^d)$ *and* $\widehat{f} \in L^1(\mathbb{R}^d)$, *then* $(\widehat{f})^\vee = f$, *a.e. on* $\mathbb{R}^d$.

Proof. By definition, there holds

$$\langle \widehat{g^\vee}, \varphi \rangle = \langle g^\vee, \widehat{\varphi} \rangle = \langle g, \varphi \rangle \qquad \text{for any} \quad \varphi \in \mathscr{S}(\mathbb{R}^d).$$

Therefore, $\widehat{g^\vee} = g$ in $\mathscr{S}'(\mathbb{R}^d)$. Note that $\widehat{f} \in L^1(\mathbb{R}^d)$,

$$\langle (\widehat{f})^\vee, \varphi \rangle = \langle \widehat{f}, \varphi^\vee \rangle = \langle f, \varphi \rangle \qquad \text{for any} \quad \varphi \in \mathscr{S}(\mathbb{R}^d).$$

By the Hausdorff-Young inequality (2.1),

$$\|(\widehat{f})^\vee\|_{L^\infty(\mathbb{R}^d)} \le \|\widehat{f}\|_{L^1(\mathbb{R}^d)}.$$

Therefore, f is a linear bounded operator on $L^1(\mathbb{R}^d)$; i.e., $f \in [L^1(\mathbb{R}^d)]^* = L^\infty(\mathbb{R}^d)$ due to $\mathscr{S}(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d)$ and

$$|\langle f, \varphi \rangle| = |\langle (\widehat{f})^\vee, \varphi \rangle| \le \|(\widehat{f})^\vee\|_{L^\infty(\mathbb{R}^d)} \|\varphi\|_{L^1(\mathbb{R}^d)} \le \|\widehat{f}\|_{L^1(\mathbb{R}^d)} \|\varphi\|_{L^1(\mathbb{R}^d)}.$$

Hence, $(\widehat{f})^\vee = f$, a.e. on $\mathbb{R}^d$ because $(\widehat{f})^\vee - f \in L^\infty(\mathbb{R}^d)$ [Bre11, Corollary 4.24]. $\square$

A direct consequence of Lemma 2.1 is that the pointwise Fourier inversion is valid for functions in $\mathscr{B}^s(\mathbb{R}^d)$. We shall frequently use this fact later on.

2.1. Completeness of the spectral Barron space.

Lemma 2.2. (1) $\mathscr{B}^s(\mathbb{R}^d), s \geq 0$ *is a Banach space.*

(2) *When* $s > 0$, $\mathscr{B}^s(\mathbb{R}^d)$ *is not a Banach space if the norm* $\| f \|_{\mathscr{B}^s(\mathbb{R}^d)}$ *is replaced by* $\upsilon_{f,s}$.

Proof. We give a brief proof for the first claim for the readers' convenience, which has been stated in [Hör63, Theorem 2.2.1].

It is sufficient to check the completeness of $\mathscr{B}^s(\mathbb{R}^d)$. For any Cauchy sequence $\{f_k\}_{k=1}^\infty \subset \mathscr{B}^s(\mathbb{R}^d)$, there exists $g \in L^1(\mathbb{R}^d)$ such that $\widehat{f_k} \to g$ in $L^1(\mathbb{R}^d)$. Therefore there exists a sub-sequence of $\{f_k\}_{k=1}^\infty$ (still denoted by f_k) such that $\widehat{f_k} \to g$ a.e. on $\mathbb{R}^d$.

Define the measure μ by setting that for any measurable set $E \subset \mathbb{R}^d$,

$$\mu(E) := \int_E |\xi|^s \mathrm{d}\xi.$$

Then $\{\widehat{f_k}\}_{k=1}^\infty$ is a Cauchy sequence in $L^1(\mathbb{R}^d, \mu)$ and there exists $h \in L^1(\mathbb{R}^d, \mu)$ such that $\widehat{f_k} \to h$ in $L^1(\mathbb{R}^d, \mu)$. Therefore there exists a sub-sequence of $\{f_k\}_{k=1}^\infty$ (still denoted by f_k) such that $\widehat{f_k} \to h$ μ-a.e. on $\mathbb{R}^d$. Note that for any measurable set $E \subset \mathbb{R}^d$, $\mu(E) = 0$ is equivalent to $|E| = 0$. Therefore $\widehat{f_k} \to h$ a.e. on $\mathbb{R}^d$. By the uniqueness of limitation, $h = g$, a.e. on $\mathbb{R}^d$.

Define $f = g^\vee$. Lemma 2.1 shows that $\widehat{f} = g$ in $\mathscr{S}'(\mathbb{R}^d)$. Therefore $f \in \mathscr{B}^s(\mathbb{R}^d)$ and $f_k \to f$ in $\mathscr{B}^s(\mathbb{R}^d)$. Hence $\mathscr{B}^s$ is complete and it is a Banach space.

We prove (2) by contradiction, suppose that the assertion (2) is false, i.e., $\mathscr{B}^s(\mathbb{R}^d)$ equipped merely with the spectral norm $\upsilon_{f,s}$ is a Banach space, then there exists C depending only on s and d such that for any $f \in \mathscr{B}^s(\mathbb{R}^d)$,

$$(2.2) \qquad\qquad\qquad \upsilon_{f,0} \leq C\upsilon_{f,s}.$$

The following example suggests that this is false.

For some $\delta > -1$, let

$$f_n(x) = \left(\sum_{k=1}^n 2^{kd}(1 - 2^{2k}|\xi|^2)_+^\delta \right)^\vee (x).$$

We rewrite f_n in terms of the Bochner-Riesz multipliers that is defined by

$$\phi_R = \left(\left(1 - \frac{|\xi|^2}{R^2}\right)_+^\delta \right)^\vee, \qquad \delta > -1.$$

Therefore, $f_n = \sum_{k=1}^n \phi_{2^{-k}}(x)$. We claim that ϕ_R admits the following explicit representation

$$(2.3) \qquad\qquad \phi_R(x) = \frac{\Gamma(\delta + 1)}{\pi^\delta |x|^{\delta + d/2}} R^{-\delta + d/2} J_{\delta + d/2}(2\pi|x|R),$$

and for $s \geq 0$, the spectrum norm of ϕ_R is

$$(2.4) \qquad \upsilon_{\phi_R,s} = \frac{\omega_{d-1}}{2} B\left(\frac{s+d}{2}, \delta+1\right) R^{s+d}.$$

The proof of the above two identities is postponed to Appendix A.1. It follows from (2.3) that

$$(2.5) \qquad f_n(x) = \frac{\Gamma(\delta+1)}{\pi^\delta |x|^{\delta+d/2}} \sum_{k=1}^{n} 2^{k(\delta+d/2)} J_{\delta+d/2}(2^{1-k}\pi|x|),$$

and $f_n \in \mathscr{B}^s(\mathbb{R}^d)$ with

$$\upsilon_{f_n,s} = \sum_{k=1}^{n} 2^{kd} \upsilon_{\phi_{2^{-k}},s} = \frac{1-2^{-ns}}{2^{s+1}-2} \omega_{d-1} B\left(\frac{s+d}{2}, \delta+1\right),$$

and

$$\upsilon_{f_n,0} = \sum_{k=1}^{n} 2^{kd} \upsilon_{\phi_{2^{-k}},0} = \frac{\omega_{d-1}}{2} B\left(\frac{d}{2}, \delta+1\right) n.$$

where we have used (2.4). It is clear that

$$\frac{\omega_{d-1}}{2^{s+1}} B\left(\frac{s+d}{2}, \delta+1\right) \leq \upsilon_{f_n,s} \leq \frac{\omega_{d-1}}{2^{s+1}-2} B\left(\frac{s+d}{2}, \delta+1\right).$$

Hence $\upsilon_{f_n,0} \simeq \mathcal{O}(n)$ while $\upsilon_{f_n,s} \simeq \mathcal{O}(1)$. This shows that (2.2) is invalid for a large number n. This proves the second claim and completes the proof. $\qquad\square$

Similar to $\mathscr{B}^s(\mathbb{R}^d)$, the space $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$ defined in (1.1) has been exploited as the target function space for neural network approximation by several authors [LGM$^+$17, Xu20, SX22, SX23]. The advantage of this space is that the Fourier transform is well-defined and the pointwise Fourier inversion is true for functions belonging to $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$. Unfortunately, $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$ is not a Banach space as shown below.

Lemma 2.3. *The space $\widehat{\mathscr{B}}^s(\mathbb{R}^d), s \geq 0$ defined in (1.1) equipped with the norm $\upsilon_{f,0} + \upsilon_{f,s}$ is not a Banach space.*

To prove Lemma 2.3, we recall the Barron spectrum space introduced by MENG AND MING in [MM22]: For $s \in \mathbb{R}$ and $1 \leq p \leq 2$,

$$(2.6) \qquad \mathscr{B}_p^s(\mathbb{R}^d) := \left\{ f \in L^p(\mathbb{R}^d) \mid \|f\|_{L^p(\mathbb{R}^d)} + \upsilon_{f,s} < \infty \right\}$$

equipped with the norm $\|f\|_{\mathscr{B}_p^s(\mathbb{R}^d)} := \|f\|_{L^p(\mathbb{R}^d)} + \upsilon_{f,s}$. A useful moment inequality that compares the spectral norm of different indexes has been proved in [MM22, Lemma 2.1]: For $1 \leq p \leq 2$ and $-d/p < s_1 < s_2$, there exists C depending only on s_1, s_2, d and p such that

$$(2.7) \qquad \upsilon_{f,s_1} \leq C \|f\|_{L^p(\mathbb{R}^d)}^\gamma \upsilon_{f,s_2}^{1-\gamma},$$

where $\gamma = (s_2 - s_1)/(s_2 + d/p)$. For any $\varepsilon > 0$, denoting $f_\varepsilon := f(x/\varepsilon)$ and using

$$\upsilon_{f_\varepsilon,s} = \varepsilon^{-s} \upsilon_{f,s},$$

we observe that the inequality (2.7) is dilation invariant because it is invariant if we replace f by f_ε.

Proof of Lemma 2.3. The authors in [MM22] have proved that $\mathscr{B}_p^s(\mathbb{R}^d)$ defined above is a Banach space. For any $f \in \mathscr{B}_1^s(\mathbb{R}^d)$, taking $s_1 = 0, s_2 = s$ and $p = 1$ in (2.7), we obtain, there exists C depending only on d and s such that

$$\upsilon_{f,0} \leq C\| f \|_{L^1(\mathbb{R}^d)}^\gamma \upsilon_{f,s}^{1-\gamma} \leq C\| f \|_{\mathscr{B}_1^s(\mathbb{R}^d)},$$

where $\gamma = s/(s + d)$.

On the contrary, suppose that $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$ equipped with the norm $\upsilon_{f,0} + \upsilon_{f,s}$ is also a Banach space, then by the bounded inverse theorem and the above moment inequality (2.7), we get, there exists C depending only on s and d such that for any $f \in \mathscr{B}_1^s(\mathbb{R}^d)$,

$$(2.8) \qquad\qquad \| f \|_{L^1(\mathbb{R}^d)} \leq C(\upsilon_{f,0} + \upsilon_{f,s}).$$

We obtain a contradiction by the following example.

For some $\delta > (d - 1)/2$, we define

$$f_n(x) := \left(\sum_{k=1}^{n} (1 - 2^{2k}|\xi|^2)_+^\delta \right)^\vee (x).$$

Using (2.3) and noting $f_n = \sum_{k=1}^{n} \phi_{2^{-k}}$, we have the explicit form of f_n as

$$(2.9) \qquad f_n(x) = \frac{\Gamma(\delta + 1)}{\pi^\delta |x|^{\delta + d/2}} \sum_{k=1}^{n} 2^{k(\delta - d/2)} J_{\delta + d/2}(2^{1-k}\pi|x|).$$

Using (2.4), we get

$$\upsilon_{f_n,s} = \sum_{k=1}^{n} \upsilon_{\phi_{2^{-k}},s} = \frac{1 - 2^{-n(s+d)}}{2^{s+d+1} - 2} \omega_{d-1} B\left(\frac{s + d}{2}, \delta + 1 \right),$$

and

$$\frac{\omega_{d-1}}{2^{s+d+1}} B\left(\frac{s + d}{2}, \delta + 1 \right) \leq \upsilon_{f_n,s} \leq \frac{\omega_{d-1}}{2^{s+d+1} - 2} B\left(\frac{s + d}{2}, \delta + 1 \right).$$

Proceeding along the same line, we obtain

$$\upsilon_{f_n,0} = \sum_{k=1}^{n} \upsilon_{\phi_{2^{-k}},0} = \frac{1 - 2^{-nd}}{2^{d+1} - 2} \omega_{d-1} B\left(\frac{d}{2}, \delta + 1 \right),$$

and

$$\frac{\omega_{d-1}}{2^{d+1}} B\left(\frac{d}{2}, \delta + 1 \right) \leq \upsilon_{f_n,0} \leq \frac{\omega_{d-1}}{2^{d+1} - 2} B\left(\frac{d}{2}, \delta + 1 \right).$$

Hence,

$$(2.10) \qquad \upsilon_{f_n,0} + \upsilon_{f_n,s} \leq \frac{\omega_{d-1}}{2} \left(\frac{B(d/2, \delta + 1)}{2^d - 1} + \frac{B((s + d)/2, \delta + 1)}{2^{s+d} - 1} \right).$$

By (2.3), a direct calculation gives

$$\| \phi_R \|_{L^1(\mathbb{R}^d)} = \frac{\Gamma(\delta + 1)}{\pi^\delta R^{\delta - d/2}} \int_{\mathbb{R}^d} \frac{|J_{\delta + d/2}(2\pi|x|R)|}{|x|^{\delta + d/2}} \mathrm{d}x = \frac{2^\delta \Gamma(\delta + 1)}{\pi^{\delta + d/2}} \int_{\mathbb{R}^d} \frac{|J_{\delta + d/2}(|x|)|}{|x|^{\delta + d/2}} \mathrm{d}x.$$

Invoking [Gra14, Appendix B.6, B.7], there exists C depending on ν such that

$$|J_\nu(x)| \le C \begin{cases} |x|^\nu & |x| \le 1, \\ |x|^{-1/2} & |x| > 1. \end{cases}$$

We get, there exists C depending only on d and δ such that

$$\| \phi_R \|_{L^1(\mathbb{R}^d)} = \frac{2^\delta \Gamma(\delta+1)}{\pi^{\delta+d/2}} \left(\int_{|x|\le 1} \frac{|J_{\delta+d/2}(|x|)|}{|x|^{\delta+d/2}} \mathrm{d}x + \int_{|x|>1} \frac{|J_{\delta+d/2}(|x|)|}{|x|^{\delta+d/2}} \mathrm{d}x \right)$$

$$\le C \left(\int_{|x|\le 1} \mathrm{d}x + \int_{|x|>1} |x|^{-1/2-\delta-d/2} \mathrm{d}x \right)$$

$$\le C \left(1 + \frac{1}{\delta - (d-1)/2} \right),$$

where we have used the fact $\delta > (d-1)/2$ in the last step. Therefore, $\| \phi_R \|_{L^1(\mathbb{R}^d)}$ is bounded by a constant that depends only on δ and d but is independent of R. Moreover,

$$\| f_n \|_{L^1(\mathbb{R}^d)} \le \sum_{k=1}^n \| \phi_{2^{-k}} \|_{L^1(\mathbb{R}^d)} \le n \| \phi_1 \|_{L^1(\mathbb{R}^d)},$$

and by the Hausdorff-Young inequality (2.1),

$$\| f_n \|_{L^1(\mathbb{R}^d)} \ge \| \widehat{f}_n \|_{L^\infty(\mathbb{R}^d)} = \widehat{f}_n(0) = n.$$

This means that $\| f_n \|_{L^1(\mathbb{R}^d)} = \mathcal{O}(n)$, which together with (2.10) immediately shows that the inequality (2.8) cannot be true for sufficiently large n. Hence, we conclude that $\widehat{\mathscr{B}}^s(\mathbb{R}^d)$ is not a Banach space. $\qquad \square$

2.2. **Embedding relations of the spectral Barron spaces.** In this part we discuss the embedding of the spectral Barron spaces.

Lemma 2.4. (1) *The moment inequality: For any $0 \le s_1 \le s \le s_2$ satisfying $s = \alpha s_1 + (1-\alpha)s_2$ with $0 \le \alpha \le 1$, and $f \in \mathscr{B}^{s_1}(\mathbb{R}^d)$, there holds*

$$(2.11) \qquad \qquad v_{f,s} \le v_{f,s_1}^\alpha v_{f,s_2}^{1-\alpha},$$

and

$$(2.12) \qquad \qquad \| f \|_{\mathscr{B}^s(\mathbb{R}^d)} \le \| f \|_{\mathscr{B}^{s_1}(\mathbb{R}^d)}^\alpha \| f \|_{\mathscr{B}^{s_2}(\mathbb{R}^d)}^{1-\alpha}.$$

(2) *Let $0 \le s_1 \le s_2$, there holds $\mathscr{B}^{s_2}(\mathbb{R}^d) \hookrightarrow \mathscr{B}^{s_1}(\mathbb{R}^d)$ with*

$$(2.13) \qquad \| f \|_{\mathscr{B}^{s_1}(\mathbb{R}^d)} \le \left(2 - \frac{s_1}{s_2} \right) \| f \|_{\mathscr{B}^{s_2}(\mathbb{R}^d)} \qquad \forall f \in \mathscr{B}^{s_2}(\mathbb{R}^d).$$

The embedding (2.13) has been stated in [Hör63, Theorem 2.2.2] without tracing the embedding constant.

Proof. We start with the moment inequality (2.11) for the spectral norm. For any $0 \le s_1 \le s \le s_2$ with $s = \alpha s_1 + (1 - \alpha)s_2$, using Hölder's inequality, we obtain

$$v_{f,s} = \int_{\mathbb{R}^d} \left(|\xi|^{s_1} |\widehat{f}(\xi)| \right)^{\alpha} \left(|\xi|^{s_2} |\widehat{f}(\xi)| \right)^{1-\alpha} \mathrm{d}\xi \le v_{f,s_1}^{\alpha} v_{f,s_2}^{1-\alpha}.$$

This gives (2.11).

Next, for $a, b, c > 0$, by Young's inequality, we have

$$\frac{a + b^{\alpha}c^{1-\alpha}}{(a+b)^{\alpha}(a+c)^{1-\alpha}} = \left(\frac{a}{a+b} \right)^{\alpha} \left(\frac{a}{a+c} \right)^{1-\alpha} + \left(\frac{b}{a+b} \right)^{\alpha} \left(\frac{c}{a+c} \right)^{1-\alpha}$$

$$\le \alpha \frac{a}{a+b} + (1-\alpha)\frac{a}{a+c} + \alpha\frac{b}{a+b} + (1-\alpha)\frac{c}{a+c}$$

$$= 1.$$

This yields

$$a + b^{\alpha}c^{1-\alpha} \le (a+b)^{\alpha}(a+c)^{1-\alpha}.$$

Let $a = v_{f,0}, b = v_{f,s_1}$ and $c = v_{f,s_2}$, we obtain

$$\| f \|_{\mathscr{B}^s(\mathbb{R}^d)} = v_{f,0} + v_{f,s} \le v_{f,0} + v_{f,s_1}^{\alpha} v_{f,s_2}^{1-\alpha} \le \| f \|_{\mathscr{B}^{s_1}(\mathbb{R}^d)}^{\alpha} \| f \|_{\mathscr{B}^{s_2}(\mathbb{R}^d)}^{1-\alpha}.$$

This implies (2.12).

Next, if we take $s_1 = 0$ in (2.11) and $s = (1 - \alpha)s_2$ with $\alpha = 1 - s/s_2$, then

$$\| f \|_{\mathscr{B}^s(\mathbb{R}^d)} \le v_{f,0} + v_{f,0}^{\alpha} v_{f,s_2}^{1-\alpha} \le (1+\alpha)v_{f,0} + (1-\alpha)v_{f,s_2} \le (1+\alpha)\| f \|_{\mathscr{B}^{s_2}(\mathbb{R}^d)}.$$

This leads to (2.13) and completes the proof. $\qquad\square$

The next lemma shows that $\mathscr{B}_p^s(\mathbb{R}^d)$ is a proper subspace of $\mathscr{B}^s(\mathbb{R}^d)$.

Lemma 2.5. *For $s \ge 0$ and $1 \le p \le 2$, there holds $\mathscr{B}_p^s(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$, and the inclusion is proper in the sense that for any $1 \le p < \infty$, there exists $f_p \in \mathscr{B}^s(\mathbb{R}^d)$ and $f_p \notin L^p(\mathbb{R}^d)$.*

Proof. It follows from the moment inequality (2.7) that $v_{f,0} \le C\| f \|_{\mathscr{B}_p^s(\mathbb{R}^d)}$. Hence

$$\| f \|_{\mathscr{B}^s(\mathbb{R}^d)} \le C\| f \|_{\mathscr{B}_p^s(\mathbb{R}^d)}.$$

This implies $\mathscr{B}_p^s(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$ for any $s \ge 0$ and $1 \le p \le 2$.

It remains to prove that the inclusion is proper. Let

$$f_p(x) := \left(|\xi|^{-d/p'} \chi_{[0,1)}(|\xi|) \right)^{\vee}(x),$$

where $\chi_{\Omega}(t)$ is the characteristic function on $\mathbb{R}$ that equals to one if $t \in \Omega$ and zero otherwise. For any $s \ge 0$, a straightforward calculation gives

$$v_{f_p,s} = \frac{\omega_{d-1}}{s + d/p}.$$

Hence, $f_p \in \mathscr{B}^s(\mathbb{R}^d)$.

What is left is to show that $f_p \notin L^p(\mathbb{R}^d)$, which is based on the following explicit formula for f_p shown in Appendix A.2:

$$(2.14) \qquad f_p(x) = {}_1F_2(d/(2p); 1 + d/(2p), d/2; -\pi^2|x|^2)p\nu_d,$$

where the generalized Hypergeometric function ${}_nF_m$ is defined as follows. For nonnegative integer n, m and none of the parameters $\{\beta_j\}_{j=1}^m$ is a negative integer or zero,

$$_nF_m(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_m; x) := \sum_{k=0}^{\infty} \frac{\prod_{j=1}^n (\alpha_j)_k}{\prod_{j=1}^m (\beta_j)_k} \frac{x^k}{k!}.$$

The generalized Hypergeometric function ${}_nF_m$ converges for all finite x if $n \leq m$. In particular ${}_nF_m(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_m; 0) = 1$. Hence $f_p(x)$ is finite for any x. Using [MS73, Appendix], we obtain

$$_1F_2(\alpha; \beta, \gamma; -x^2/4) \simeq \mathcal{O}(|x|^{\alpha-\beta-\gamma+1/2} + |x|^{-2\alpha}) \qquad \text{when} \quad |x| \to \infty.$$

Therefore,

$$f_p(x) \simeq \mathcal{O}(|x|^{-(d+1)/2} + |x|^{-d/p}) \qquad \text{when} \quad |x| \to \infty.$$

This immediately implies $f_p \notin L^p(\mathbb{R}^d)$. $\square$

Remark 2.6. The representation (2.14) is rather complicated, we give explicit formulas for certain special cases. For $d = p = 1$, using the relation [Luk69, § 6.2.1, Eq.(10)]

$$\sin x = {}_0F_1(; 3/2; -x^2/4)x,$$

we obtain

$$f_1(x) = 2{}_1F_2(1/2; 3/2, 1/2; -\pi^2x^2) = 2{}_0F_1(; 3/2; -\pi^2x^2) = \frac{\sin(2\pi x)}{\pi x}.$$

When $p = 2$, using the identity [Luk69, § 6.2.11, Eq. (41)]

$$C(\sqrt{2x/\pi}) = \sqrt{\frac{2x}{\pi}}{}_1F_2(1/4; 5/4, 1/2; -x^2/4) \qquad \text{when} \quad x > 0,$$

we obtain

$$f_2(x) = 4{}_1F_2(1/4; 5/4, 1/2; -\pi^2x^2) = \frac{2}{\sqrt{|x|}}C(2\sqrt{|x|}).$$

where C is the Fresnel Cosine integral given by

$$C(x) = \int_0^x \cos(\pi t^2/2)\mathrm{d}t \to \frac{1}{2} \qquad \text{when} \quad x \to \infty.$$

2.3. **Relations to some classical function spaces.** In this part, we establish the embedding between the spectral Barron space $\mathscr{B}^s(\mathbb{R}^d)$ and the Besov space, and hence we bridge $\mathscr{B}^s(\mathbb{R}^d)$ and the Sobolev space as in [MM22]. Let us begin with the definition of the Fourier-analytical based Besov space [Tri83].

Definition 2.7 (Besov space)**.** Let $\{\varphi_j\}_{j=0}^\infty \subset \mathscr{S}(\mathbb{R}^d)$ satisfies $0 \leq \varphi_j \leq 1$ and

$$
\begin{cases}
\text{supp}(\varphi_0) \subset \Gamma_0 := \left\{ x \in \mathbb{R}^d \mid |x| \leq 2 \right\}, \\
\text{supp}(\varphi_j) \subset \Gamma_j := \left\{ x \in \mathbb{R}^d \mid 2^{j-1} \leq |x| \leq 2^{j+1} \right\}, \qquad j = 1, 2, \ldots.
\end{cases}
$$

For every multi-index α, there exists a positive number c_α such that

$$
2^{j|\alpha|}|\nabla^\alpha \varphi_j(x)| \leq c_\alpha \quad \text{for all} \quad j = 0, \ldots, \quad \text{for all} \quad x \in \mathbb{R}^d,
$$

and

$$
\sum_{j=0}^\infty \varphi_j(x) = 1 \quad \text{for every} \quad x \in \mathbb{R}^d.
$$

Let $\alpha \in \mathbb{R}$ and $1 \leq p, q \leq \infty$. Define the *Besov space*

$$
B_{p,q}^\alpha(\mathbb{R}^d) := \left\{ f \in \mathscr{S}'(\mathbb{R}^d) \mid \| f \|_{B_{p,q}^\alpha(\mathbb{R}^d)} < \infty \right\}
$$

equipped with the norm

$$
\| f \|_{B_{p,q}^\alpha(\mathbb{R}^d)} := \left(\sum_{j=0}^\infty 2^{\alpha q j} \| (\varphi_j \widehat{f})^\vee \|_{L^p(\mathbb{R}^d)}^q \right)^{1/q} \qquad \text{when} \quad q < \infty,
$$

and

$$
\| f \|_{B_{p,\infty}^\alpha(\mathbb{R}^d)} := \sup_{j \geq 0} 2^{\alpha j} \| (\varphi_j \widehat{f})^\vee \|_{L^p(\mathbb{R}^d)}.
$$

Firstly we recall some well-known facts about the embedings of Besov spaces, which was firstly proved by Taibleson in the series of work [Tai64, Tai65, Tai66]. We retain the proof in Appendix A.3 for the readers' convenience.

Lemma 2.8. *There holds $B_{p_1,q_1}^{\alpha_1}(\mathbb{R}^d) \hookrightarrow B_{p_2,q_2}^{\alpha_2}(\mathbb{R}^d)$ if and only if $p_1 \leq p_2$ and one of the following conditions holds:*

(1) $\alpha_1 - d/p_1 > \alpha_2 - d/p_2$ *and* q_1, q_2 *are arbitrary;*
(2) $\alpha_1 - d/p_1 = \alpha_2 - d/p_2$ *and* $q_1 \leq q_2$.

The main result of the embedding is:

Theorem 2.9.   (1) *There holds*

$$
(2.15) \qquad\qquad B_{2,1}^{s+d/2}(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d) \hookrightarrow B_{\infty,1}^s(\mathbb{R}^d)
$$

with

$$
(2.16) \qquad 2^{-s}\| f \|_{B_{\infty,1}^s(\mathbb{R}^d)} \leq \| f \|_{\mathscr{B}^s(\mathbb{R}^d)} \leq 2^{s+1+d/2}\sqrt{\nu_d}\| f \|_{B_{2,1}^{s+d/2}(\mathbb{R}^d)}.
$$

(2) *The above embedding is optimal in the sense that* $B_{p,q}^\alpha(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$ *if and only if* $B_{p,q}^\alpha(\mathbb{R}^d) \hookrightarrow B_{2,1}^{s+d/2}(\mathbb{R}^d)$, *and* $\mathscr{B}^s(\mathbb{R}^d) \hookrightarrow B_{p,q}^\alpha(\mathbb{R}^d)$ *if and only if* $B_{\infty,1}^s(\mathbb{R}^d) \hookrightarrow B_{p,q}^\alpha(\mathbb{R}^d)$.

Remark 2.10. By tracking the embedding constant in (2.16), we observe that it is indenpendent of the dimension d because

$$2^{s+1+d/2}\sqrt{\nu_d} \sim \frac{2^{s+1}}{(d\pi)^{1/4}}\left(\frac{8\pi e}{d}\right)^{d/4} \to 0 \qquad \text{as} \quad d \to \infty.$$

This suggests that the embedding is effective in high dimension. It also indicates that the spectral Barron space is more suitable than the Besov space as a target function space for neural network approximation because

$$\| f \|_{\mathscr{B}(\mathbb{R}^d)} \le C d^{-(d+1)/4} \| f \|_{B_{2,1}^{s+d/2}(\mathbb{R}^d)} \qquad \text{for any} \quad f \in B_{2,1}^{s+d/2}(\mathbb{R}^d)$$

with a universal constant C.

Proof of Theorem 2.9. To prove (1), firstly, for any $f \in \mathscr{B}^s(\mathbb{R}^d)$,

$$\| f \|_{\mathscr{B}^s(\mathbb{R}^d)} = \sum_{j=0}^\infty \int_{\mathbb{R}^d} (1 + |\xi|^s)\varphi_j(\xi)|\widehat{f}(\xi)|\mathrm{d}\xi$$

$$\le \sum_{j=0}^\infty \left(\int_{\operatorname{supp}\varphi_j} (1 + |\xi|^s)^2 \mathrm{d}\xi\right)^{1/2} \| \varphi_j \widehat{f} \|_{L^2(\mathbb{R}^d)}.$$

A direct calculation gives: for $j = 0, 1, \ldots,$

$$\int_{\operatorname{supp}\varphi_j} (1 + |\xi|^s)^2 \mathrm{d}\xi \le \int_{0 \le |\xi| \le 2^{j+1}} (1 + |\xi|^s)^2 \mathrm{d}\xi$$

$$= \omega_{d-1}\int_0^{2^{j+1}} (1 + r^s)^2 r^{d-1} \mathrm{d}r$$

$$\le 2\omega_{d-1}\int_0^{2^{j+1}} (1 + r^{2s}) r^{d-1} \mathrm{d}r$$

$$\le 2\omega_{d-1}\left(\frac{2^{(j+1)d}}{d} + \frac{2^{(j+1)(2s+d)}}{2s+d}\right)$$

$$\le 4\nu_d 2^{(j+1)(2s+d)}.$$

Using the Plancherel's theorem, we get

$$\| f \|_{\mathscr{B}^s(\mathbb{R}^d)} \le 2^{s+1+d/2}\sqrt{\nu_d}\sum_{j=0}^\infty 2^{j(s+d/2)}\| \varphi_j\widehat{f} \|_{L^2(\mathbb{R}^d)}$$

$$= 2^{s+1+d/2}\sqrt{\nu_d}\sum_{j=0}^\infty 2^{j(s+d/2)}\| (\varphi_j\widehat{f})^\vee \|_{L^2(\mathbb{R}^d)}$$

$$= 2^{s+1+d/2}\sqrt{\nu_d}\| f \|_{B_{2,1}^{s+d/2}(\mathbb{R}^d)}.$$

Next, for any $f \in \mathscr{B}^s(\mathbb{R}^d)$, by Lemma 2.1, we have $\varphi_j \widehat{f} \in L^1(\mathbb{R}^d)$, using the Hausdorff-Young inequality (2.1), we obtain

$$\| f \|_{B^s_{\infty,1}(\mathbb{R}^d)} = \sum_{j=0}^{\infty} 2^{sj} \| (\varphi_j \widehat{f})^{\vee} \|_{L^{\infty}(\mathbb{R}^d)} \leq \sum_{j=0}^{\infty} 2^{sj} \| \varphi_j \widehat{f} \|_{L^1(\mathbb{R}^d)}$$

$$\leq \| \varphi_0 \widehat{f} \|_{L^1(\mathbb{R}^d)} + 2^s \sum_{j=1}^{\infty} \int_{\mathbb{R}^d} \varphi_j(\xi) |\xi|^s |\widehat{f}(\xi)| \mathrm{d}\xi$$

$$\leq \upsilon_{f,0} + 2^s \upsilon_{f,s}.$$

Therefore, $\| f \|_{B^s_{\infty,1}(\mathbb{R}^d)} \leq 2^s \| f \|_{\mathscr{B}^s(\mathbb{R}^d)}$. This proves (2.15) with

$$2^{-s} \| f \|_{B^s_{\infty,1}(\mathbb{R}^d)} \leq \| f \|_{\mathscr{B}^s(\mathbb{R}^d)} \leq 2^{s+1+d/2} \sqrt{\nu_d} \| f \|_{B^{s+d/2}_{2,1}(\mathbb{R}^d)}.$$

It remains to show that the embedding (2.15) is optimal. The "if"-part is apparent, it suffices to show the "only if"-part. On the one hand, suppose $B^\alpha_{p,q}(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$, one would have $B^\alpha_{p,q}(\mathbb{R}^d) \hookrightarrow B^s_{\infty,1}(\mathbb{R}^d)$. Using Lemma 2.8, we conclude that the triple (p,q,α) must be in the set

$$(2.17) \quad S_1 := \{ p, q \in [1, \infty], \alpha \in \mathbb{R} \mid \alpha > s + d/p, \text{ or } \alpha = s + d/p \text{ and } q = 1 \}.$$

We split this set into three subsets.

First, if $(p,q,\alpha) \in S_1$ with $1 \leq p \leq 2$, then by Lemma 2.8 we obtain $B^\alpha_{p,q}(\mathbb{R}^d) \hookrightarrow B^{s+d/2}_{2,1}(\mathbb{R}^d)$. This proves the first assertion.

Second, if $(p,q,\alpha) \in S_1$ with $2 < p < \infty$, then we shall exploit an example adopted from [LL01, Ch. 5, Ex. 9] to show that $B^\alpha_{p,q}(\mathbb{R}^d) \not\hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$. Let

$$\psi_n(x) = (1 + in)^{-d/2} e^{-\pi |x|^2/(1+in)}.$$

A direct calculation gives $\widehat{\psi}_n(\xi) = e^{-\pi(1+in)|\xi|^2}$. Hence $|\widehat{\psi}_n(\xi)| = e^{-\pi|\xi|^2} \in \mathscr{S}(\mathbb{R}^d)$ and

$$\| \psi_n \|_{\mathscr{B}^s(\mathbb{R}^d)} = 1 + \frac{\Gamma((s+d)/2)}{\Gamma(d/2)\pi^{s/2}},$$

which is independent of n. We shall show in Appendix A.4 that there exists C independent of n such that

$$(2.18) \qquad \| \psi_n \|_{B^\alpha_{p,q}(\mathbb{R}^d)} \leq C(1 + n^2)^{-d(p-2)/(4p)}$$

when $2 \leq p < \infty$ and $\alpha > 0$. Therefore, $\| \psi_n \|_{B^\alpha_{p,q}(\mathbb{R}^d)} \to 0$ when $p > 2$ and $n \to \infty$. This proves $B^\alpha_{p,q}(\mathbb{R}^d) \not\hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$.

Finally, if $(p,q,\alpha) \in S_1$ with $p = \infty$. Note that $\varphi_j(x) = \delta_{0j}$ when $|x| < 1$. A direct calculation gives

$$\| 1 \|_{B^\alpha_{\infty,q}(\mathbb{R}^d)} = \| (\varphi_0 \widehat{1})^{\vee} \|_{L^{\infty}(\mathbb{R}^d)} = \| 1 \|_{L^{\infty}(\mathbb{R}^d)} = 1,$$

while it is clear that the constant function $1 \notin \mathscr{B}^s(\mathbb{R}^d)$. This implies $B^\alpha_{\infty,q}(\mathbb{R}^d) \not\hookrightarrow \mathscr{B}^s(\mathbb{R}^d)$.

Summing up the above three cases, we conclude the first part of the assertion (2).

On the other hand, suppose $\mathscr{B}^s(\mathbb{R}^d) \hookrightarrow B_{p,q}^\alpha(\mathbb{R}^d)$, one would have $B_{2,1}^{s+d/2}(\mathbb{R}^d) \hookrightarrow B_{p,q}^\alpha(\mathbb{R}^d)$. Invoking Lemma 2.8 again, we conclude that the index (p, q, α) must be in the set

$$(2.19) \qquad S_2 := \{ p \in [2, \infty], q \in [1, \infty], \alpha \in \mathbb{R} \mid \alpha \leq s + d/p \}.$$

Again we split this set into two subsets.

First, if $(p, q, \alpha) \in S_2$ with $2 \leq p < \infty$, then we have proved that f_p constructed in Lemma 2.5 satisfying $f_p \in \mathscr{B}^s(\mathbb{R}^d)$ while $f_p \notin L^p(\mathbb{R}^d)$. A direct calculation gives

$$\| f_p \|_{B_{p,q}^\alpha(\mathbb{R}^d)} = \| (\varphi_0 \widehat{f_p})^\vee \|_{L^p(\mathbb{R}^d)} = \| f_p \|_{L^p(\mathbb{R}^d)}.$$

Hence $f_p \notin B_{p,q}^\alpha(\mathbb{R}^d)$ and it yields $\mathscr{B}^s(\mathbb{R}^d) \not\hookrightarrow B_{p,q}^\alpha(\mathbb{R}^d)$.

Second, if $(p, q, \alpha) \in S_2$ with $p = \infty$. As expected, Lemma 2.8 yields that $B_{\infty,1}^s(\mathbb{R}^d) \hookrightarrow B_{\infty,q}^\alpha(\mathbb{R}^d)$. Hence we prove the second part of the assertion (2). Therefore, we conclude that the embedding (2.15) is optimal. $\qquad\square$

As a consequence of Theorem 2.9 and Lemma 2.5, we establish the embedding between the spectral Barron space and the Sobolev spaces.

Definition 2.11 (Fractional Sobolev space)**.** Let $1 \leq p < \infty$ and non-integer $\alpha > 0$, then the fractional Sobolev space

$$W_p^\alpha(\mathbb{R}^d) := \left\{ f \in W_p^{\lfloor \alpha \rfloor}(\mathbb{R}^d) \mid \iint_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|\nabla^{\lfloor \alpha \rfloor} f(x) - \nabla^{\lfloor \alpha \rfloor} f(y)|^p}{|x - y|^{d + (\alpha - \lfloor \alpha \rfloor)p}} \mathrm{d}x \mathrm{d}y < \infty \right\}$$

equipped with the norm

$$\| f \|_{W_p^\alpha(\mathbb{R}^d)} := \| f \|_{W_p^{\lfloor \alpha \rfloor}(\mathbb{R}^d)} + \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|\nabla^{\lfloor \alpha \rfloor} f(x) - \nabla^{\lfloor \alpha \rfloor} f(y)|^p}{|x - y|^{d + (\alpha - \lfloor \alpha \rfloor)p}} \mathrm{d}x \mathrm{d}y \right)^{1/p}.$$

It is a straightforward corollary of the following embedding relation between the Sobolev space and $\mathscr{B}_p^s(\mathbb{R}^d)$.

Lemma 2.12 ([MM22, Theorem 4.3])**.** (1) *If* $1 \leq p \leq 2$ *and* $\alpha > s + d/p > 0$, *then*

$$W_p^\alpha(\mathbb{R}^d) \hookrightarrow \mathscr{B}_p^s(\mathbb{R}^d).$$

(2) *If* $s > -d$ *is not an integer or* $s > -d$ *is an integer and* $d \geq 2$, *then*

$$W_1^{s+d}(\mathbb{R}^d) \hookrightarrow \mathscr{B}_1^s(\mathbb{R}^d).$$

It follows from the above lemma and Lemma 2.5 that

Corollary 2.13. (1) *If* $1 \leq p \leq 2$ *and* $\alpha > s + d/p$, *there holds*

$$(2.20) \qquad W_p^\alpha(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d).$$

(2) *If* s *is not an integer or* s *is an integer and* $d \geq 2$, *then*

$$W_1^{s+d}(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d).$$

The first embedding with $p = 2$ and $s = 1$ was hidden in [Bar93, § II, Para. 7; § IX, 15].

Proof. By Lemma 2.12 and Lemma 2.5, when $\alpha > s + d/p$ and $1 \le p \le 2$, we have

$$W_p^\alpha(\mathbb{R}^d) \hookrightarrow \mathscr{B}_p^s(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d).$$

When s is not an integer or s is an integer and $d \ge 2$, there holds

$$W_1^{s+d}(\mathbb{R}^d) \hookrightarrow \mathscr{B}_1^s(\mathbb{R}^d) \hookrightarrow \mathscr{B}^s(\mathbb{R}^d).$$

It remains to prove the right-hand side of (2.20). Using Theorem 2.9,

$$\mathscr{B}^s(\mathbb{R}^d) \hookrightarrow B_{\infty,1}^s(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d)$$

due to Theorem 2.9, Lemma 2.8 and [Tri83, § 2.3.5, Eq. (1); § 2.5.7, Eq. (2), (9), (11)]. □

3. Application to deep neural network approximation

The embedding results proved in Theorem 2.9 and Corollary 2.13 indicate that s is a smoothness index. Consequently, we are interested in exploring the approximate rate when s is small with $\mathscr{B}^s$ as the target function space. To facilitate our analysis, we shall focus on the hypercube $\Omega := [0,1]^d$, and the spectral norm for function f with respect to Ω is

$$v_{f,s,\Omega} = \inf_{Ef|_\Omega = f} \int_{\mathbb{R}^d} |\xi|_1^s |\widehat{Ef}(\xi)| \mathrm{d}\xi.$$

Here we replace $|\xi|$ by $|\xi|_1$ in the definition of $v_{f,s,\Omega}$, the latter seems more natural for studying the approximation over the hypercube as suggested by [Bar93, § V].

Definition 3.1. A sigmoidal function is a bounded function $\sigma : \mathbb{R} \mapsto \mathbb{R}$ such that

$$\lim_{t \to -\infty} \sigma(t) = 0, \qquad \lim_{t \to +\infty} \sigma(t) = 1.$$

For example, the Heaviside function $\chi_{[0,\infty)}$ is a sigmoidal function.

A classical idea for the approximation error of neural networks with sigmoidal activation functions σ is to use the Heaviside function $\chi_{[0,\infty)}$ as a transition. CARAGEA ET. AL. [CPV23] pointed out that the gap between sigmoidal function σ and the Heaviside function $\chi_{[0,\infty)}$ cannot be dismissed in $L^\infty(\Omega)$. While this gap does not exist in $L^2(\Omega)$.

Lemma 3.2. *For fixed $\omega \in \mathbb{R}^d \backslash \{0\}$ and $b \in \mathbb{R}$,*

$$\lim_{\tau \to \infty} \| \sigma(\tau(\omega \cdot x + b)) - \chi_{[0,\infty)}(\omega \cdot x + b) \|_{L^2(\Omega)} = 0.$$

Proof. Note that

$$\lim_{t \to \pm\infty} |\sigma(t) - \chi_{[0,\infty)}(t)| = 0.$$

We divide the cube Ω into $\Omega_1 := \{ x \in \Omega \mid |\tau(\omega \cdot x + b)| < \delta \}$ and $\Omega_2 := \Omega \setminus \Omega_1$. With proper choice of $\delta > 0$ and $\tau > 0$ large enough, we can obtain that the L^2-distance between $\sigma(\tau(\omega \cdot x + b))$ and $\chi_{[0,\infty)}(\omega \cdot x + b)$ is arbitrarily small. $\square$

For a shallow neural network, the following lemma in [Bar93] is proved for the real-valued function, while it is straightforward to extend the proof to the complex-valued function.

Lemma 3.3 ([Bar93, Theorem 1]). *Let* $f \in \mathscr{B}^1(\mathbb{R}^d)$, *there exists*

(3.1) $$f_N(x) = \sum_{i=1}^{N} c_i \sigma(\omega_i \cdot x + b_i)$$

with $\omega_i \in \mathbb{R}^d, b_i \in \mathbb{R}$ *and* $c_i \in \mathbb{C}$ *such that*

$$\| f - f_N \|_{L^2(\Omega)} \leq \frac{2v_{f,1,\Omega}}{\sqrt{N}}.$$

In this part, we shall show the approximation error for the deep neural network. We use the (L, N)-network to describe a neural network with L hidden layers and at most N units per layer. Here L denotes the number of hidden layers, e.g., the shallow neural network, expressed as (3.1), is an $(1, N)$-network.

Definition 3.4 ((L, N)-network). An (L, N)-network represents a neural network with L hidden layers and at most N units per layer. The activation functions of the first $L-1$ layers are all ReLU and the activation function of the last layer is the sigmoidal function. The connection weights between the input layer and the hidden layer, and between the hidden layer and the hidden layer are all real numbers. The connection weights between the last hidden layer and the output layer are complex numbers.

There is relatively little work on the approximation rate of deep neural networks that utilize the spectral Barron space as the target space. For deep ReLU networks, [BN20b] has proven approximation results of $(sL/2)$-order. The main contribution of this section is to improve this result to an sL-order approximation, at the cost of introducing $v_{f,s,\Omega}$ in the estimate; cf. Theorem 3.5.

Theorem 3.5. *Let the positive integer* L *and* $f \in \mathscr{B}^s(\mathbb{R}^d)$ *with* $0 < sL \leq 1/2$. *For any positive integer* N *there exists an* $(L, N + 2)$-*network* f_N *such that*

(3.2) $$\| f - f_N \|_{L^2(\Omega)} \leq \frac{29v_{f,s,\Omega}}{N^{sL}}.$$

Moreover, if f *is a real-valued function, then the connection weights in* f_N *are all real.*

As far as we are aware, the above theorem represents the state-of-the-art in the literature to date. For shallow neural network $L = 1$, the authors in [MM22] established a 1/2-order convergence with target function space $\mathscr{B}_2^{1/2}(\mathbb{R}^d)$, which is a

subspace of $\mathscr{B}^{1/2}(\mathbb{R}^d)$. Their estimate depends on the dimension with a factor $d^{1/4}$. The upper bound in [SX22] depends on $\upsilon_{f,0} + \upsilon_{f,1/2}$, whereas (2.5) demonstrates that $\upsilon_{f,0}$ can be much larger than $\upsilon_{f,s}$ for certain functions in $\mathscr{B}^s(\mathbb{R}^d)$, and the estimate depends on the dimension exponentially. In contrast to these two results, the upper bound in Theorem 3.5 relies solely on $\upsilon_{f,s,\Omega}$, and is independent of the dimension.

For deep neural network, a similar result for the ReLU activation function has been presented in [BN20b] with a $(sL/2)$-order approximation. Compared to this result, our estimate exhibits a higher level of optimality. At first glance, our result might seem to contradict [BN20b, Theorem 2]. In reality, this is not the case because the upper bound in that reference is $\sqrt{\upsilon_{f,0}\upsilon_{f,s}} + \upsilon_{f,0}$, which requires $f \in \mathscr{B}^s(\mathbb{R}^d)$, but is typically smaller than $\| f \|_{\mathscr{B}^s(\mathbb{R}^d)}$ for oscillatory functions; cf. Lemma A.1.

In what follows, we make some preparations to prove Theorem 3.5. The analysis in this part owns the most to [BN20b] with certain improvements that will be detailed later on. For any function g defined on $[0,1]$ and it is symmetric about $x = 1/2$, We use the notation $g_{,n}$ to denote the function g in the $[0,1]$ interval of the period repeated n times, i.e.,

$$(3.3) \qquad g_{,n}(t) = g(nt - j), \quad j = 0, \ldots, n-1, \quad 0 \le nt - j \le 1.$$

Define

$$\beta(t) = \mathrm{ReLU}(2t) - 2\mathrm{ReLU}(2t-1) + \mathrm{ReLU}(2t-2) = \begin{cases} 2t, & 0 \le t \le 1/2, \\ 2 - 2t, & 1/2 \le t \le 1, \\ 0, & \text{otherwise}. \end{cases}$$

By definition (3.3), $\beta_{,n}$ represents a triangle function with n peaks and can be represented by $3n$ ReLUs:

$$\beta_{,n}(t) = \sum_{j=0}^{n-1} \beta(nt - j), \quad 0 \le t \le 1.$$

Lemma 3.6. *Let g be a function defined on $[0,1]$ and symmetric about $x = 1/2$, then $g_{,n_2} \circ \beta_{,n_1} = g_{,2n_1n_2}$ on $[0,1]$.*

The above lemma is a rigorous statement of [Tel16, Proposition 5.1]. A key example is $\cos(2\pi n_2 \beta_{,n_1}(t)) = \cos(4\pi n_1 n_2 t)$ when $t \in [0,1]$. A geometrical explanation may be founded in [EPGB21, Figure 3]. We postpone the rigorous proof in Appendix A.5.

For $r \in (0,1)$, we define

$$\alpha(t,r) = \chi_{[0,\infty)}(t - r/2) - \chi_{[0,\infty)}(t - (1-r)/2) = \begin{cases} \chi_{[r/2,(1-r)/2]}(t), & 0 < r \le 1/2, \\ -\chi_{[(1-r)/2,r/2]}(t), & 1/2 \le r < 1, \end{cases}$$

then $\mathrm{supp}(\alpha(\cdot, r)) \subset (0, 1/2)$ and $\alpha(t, r)$ is symmetric about $t = 1/4$. Define

$$\gamma(t, r) = \alpha(t + 1/4, r) - \alpha(t - 1/4, r) + \alpha(t - 3/4, r).$$

Then $\gamma(t, r)$ is symmetric about $t = 1/2$ because

$$\gamma(1 - t, r) = \alpha(5/4 - t, r) - \alpha(3/4 - t, r) + \alpha(1/4 - t, r)$$
$$= \alpha(t - 3/4, r) - \alpha(t - 1/4, r) + \alpha(t + 1/4, r) = \gamma(t, r).$$

By definition (3.3), $\gamma_{,n}(\cdot, r)$ is well defined on $[0, 1]$ and

$$\gamma_{,n}(t, r) = \begin{cases} \alpha(nt - j + 1/4, r), & 0 \leq nt - j \leq 1/4, \\ -\alpha(nt - j - 1/4, r), & 1/4 \leq nt - j \leq 3/4, \\ \alpha(nt - j - 3/4, r), & 3/4 \leq nt - j \leq 1, \end{cases} \quad j = 0, \ldots, n - 1$$

$$= \begin{cases} \alpha(nt - j + 1/4, r), & -1/4 \leq nt - j \leq 1/4, \\ -\alpha(nt - j - 1/4, r), & 1/4 \leq nt - j \leq 3/4, \end{cases} \quad j = 0, \ldots, n.$$

And $\gamma_{,n}(\cdot, r)$ on $[0, 1]$ can be represents by $4n$ Heaviside function $\chi_{[0,\infty)}$ due to $\alpha(nt + 1/4, r), \alpha(nt - n + 1/4, r)$ only need one Heaviside function each:

$$\gamma_{,n}(t, r) = \sum_{j=0}^{n} \alpha(nt - j + 1/4, r) - \sum_{j=0}^{n-1} \alpha(nt - j - 1/4, r).$$

A direct consequence of the above construction is

Lemma 3.7. *For $t \in [0, 1]$, there holds*

(3.4) $$\frac{\pi}{2} \int_0^1 \cos(\pi r) \gamma_{,n}(t, r) \mathrm{d}r = \cos(2\pi nt).$$

Proof. For any $t \in [0, 1/2]$, a direct calculation gives

$$\frac{\pi}{2} \int_0^1 \cos(\pi r) \alpha(t, r) \mathrm{d}r = \pi \int_0^{2t} \cos(\pi r) \mathrm{d}r = \sin(2\pi t).$$

Fix a $t \in [0, 1]$. If there exists an integer j satisfying $0 \leq j \leq n$ and $-1/4 \leq nt - j \leq 1/4$, then $\gamma_{,n}(t, r) = \alpha(nt - j + 1/4, r)$ and

$$\frac{\pi}{2} \int_0^1 \cos(\pi r) \alpha(nt - j + 1/4, r) \mathrm{d}r = \sin(2\pi(nt - j + 1/4)) = \cos(2\pi nt).$$

Otherwise there exists an integer j satisfying $0 \leq j \leq n - 1$ and $1/4 \leq nt - j \leq 3/4$. Then $\gamma_{,n}(t, r) = -\alpha(nt - j - 1/4, r)$ and

$$-\frac{\pi}{2} \int_0^1 \cos(\pi r) \alpha(nt - j - 1/4, r) \mathrm{d}r = -\sin(2\pi(nt - j - 1/4)) = \cos(2\pi nt).$$

This completes the proof of (3.4). □

The following Lemma is the key point to prove Theorem 3.5, which follows the framework of [BN20b], while we achieve a higher order convergence rate and the constant is dimension-free for high-frequency function.

Lemma 3.8. *Let the positive integer L and $f \in \mathscr{B}^s(\mathbb{R}^d)$ with $0 < sL \leq 1/2$ and* supp $\widehat{f} \subset \left\{ \xi \in \mathbb{R}^d \mid |\xi|_1 \geq 1 \right\}$. *For any positive integer N there exists an (L, N)-network f_N such that*

$$(3.5) \qquad \| f - f_N \|_{L^2(\Omega)} \leq \frac{22 v_{f,s,\Omega}}{N^{sL}}.$$

Proof. By Lemma 2.1, for $f \in \mathscr{B}^s(\mathbb{R}^d)$, assume f is real-valued, then

$$f(x) = \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \mathrm{d}\xi = \int_{\mathbb{R}^d} |\widehat{f}(\xi)| \cos(2\pi(\xi \cdot x + \theta(\xi))) \mathrm{d}\xi,$$

with proper choice $\theta(\xi)$ such that $0 \leq \xi \cdot x + \theta(\xi) \leq |\xi|_1 + 1$. For fixed ξ, choose $n_\xi = 2^{L-1} \lceil (|\xi|_1 + 1)^{1/L} \rceil^L$ and $t_\xi(x) = (\xi \cdot x + \theta(\xi))/n_\xi$, then $0 \leq t_\xi(x) \leq 1$ and by Lemma 3.7, we reshape $f(x)$ as

$$f(x) = \int_{\mathbb{R}^d} |\widehat{f}(\xi)| \cos(2\pi n_\xi t_\xi(x)) \mathrm{d}\xi = \frac{\pi}{2} \int_{\mathbb{R}^d} |\widehat{f}(\xi)| \mathrm{d}\xi \int_0^1 \cos(\pi r) \gamma_{,n_\xi}(t_\xi(x), r) \mathrm{d}r.$$

Define the probability measure

$$(3.6) \qquad \mu(\mathrm{d}\xi, \mathrm{d}r) = \frac{1}{Q} |\xi|_1^{-s} |\widehat{f}(\xi)| \chi_{(0,1)}(r) \mathrm{d}\xi \mathrm{d}r,$$

where Q is the normalized factor that

$$Q = \int_{\mathbb{R}^d} |\xi|_1^{-s} |\widehat{f}(\xi)| \mathrm{d}\xi \int_0^1 \mathrm{d}r \leq v_{f,s,\Omega}.$$

Therefore $f(x) = \mathbb{E}_{(\xi,r) \sim \mu} F(x, \xi, r)$ with

$$F(x, \xi, r) = \frac{\pi Q}{2} |\xi|_1^s \cos(\pi r) \gamma_{,n_\xi}(t_\xi(x), r).$$

If $\{\xi_i, r_i\}_{i=1}^m$ is an i.i.d. sequence of random samples from μ, and

$$\tilde{f} = \frac{1}{m} \sum_{i=1}^m F(x, \xi_i, r_i),$$

then using Fubini's theorem, we obtain

$$\mathbb{E}_{(\xi_i, r_i) \sim \mu} \| f - \tilde{f} \|_{L^2(\Omega)}^2 = \int_\Omega \mathbb{E}_{(\xi_i, r_i) \sim \mu} |\mathbb{E}_{(\xi,r) \sim \mu} F(x, \xi, r) - \tilde{f}(x)|^2 \mathrm{d}x$$

$$= \frac{1}{m} \int_\Omega \mathrm{Var}_{(\xi,r) \sim \mu} F(x, \xi, r) \mathrm{d}x$$

$$\leq \frac{1}{m} \mathbb{E}_{(\xi,r) \sim \mu} \| F(\cdot, \xi, r) \|_{L^\infty(\Omega)}^2.$$

Note that

$$\| F(\cdot, \xi, r) \|_{L^\infty(\Omega)} \leq \frac{\pi Q}{2} |\xi|_1^s,$$

we obtain

$$\mathbb{E}_{(\xi_i, r_i) \sim \mu} \| f - \tilde{f} \|_{L^2(\Omega)}^2 \leq \frac{1}{m} \mathbb{E}_{(\xi,r) \sim \mu} \| F(\cdot, \xi, r) \|_{L^\infty(\Omega)}^2 \leq \frac{\pi^2 Q v_{f,s,\Omega}}{4m}.$$

By Markov's inequality, with probability at least $(1+\varepsilon)/(2+\varepsilon)$, for some $\varepsilon > 0$ to be chosen later on, we obtain

$$(3.7) \qquad \| f - \tilde{f} \|_{L^2(\Omega)}^2 \leq \frac{(2+\varepsilon)\pi^2 Q v_{f,s,\Omega}}{4m}$$

It remains to calculate the number of units in each layer. For each $\gamma_{,n_\xi}(t_\xi(x), r)$, choose $n_1 = \cdots = n_L = \lceil (|\xi|_1+1)^{1/L} \rceil$, then $n_\xi = 2^{L-1} n_1 \ldots n_L$, and by Lemma 3.6, $\gamma_{,n_\xi}(\cdot, r) = \gamma_{,n_L}(\cdot, r) \circ \beta_{,n_{L-1}} \circ \cdots \circ \beta_{,n_1}$ on $[0,1]$. Lemma 3.2 shows the Heaviside function $\chi_{[0,\infty)}$ can be approximated by σ with at most

$$\max\{3n_1, \ldots, 3n_{L-1}, 4n_L\} \leq 4\lceil (|\xi|_1 + 1)^{1/L} \rceil \leq 12 |\xi|_1^{1/L}$$

units in each layer to represent $\gamma_{,n_\xi}(t_\xi(x), r)$. Denote N the total number of units in each layer, then $N \leq 12 \sum_{i=1}^m |\xi_i|_1^{1/L}$ and

$$\mathbb{E}_{(\xi_i, r_i) \sim \mu} N^{2sL} \leq 12 \sum_{i=1}^m \mathbb{E}_{(\xi_i, r_i) \sim \mu} |\xi_i|_1^{2s} \leq \frac{12 m v_{f,s,\Omega}}{Q}.$$

Invoking Markov inequality again, with probability at least $(1+\varepsilon)/(2+\varepsilon)$, we obtain

$$(3.8) \qquad \frac{Q}{m} \leq \frac{12(2+\varepsilon) v_{f,s,\Omega}}{N^{2sL}}.$$

Combining (3.7) and (3.8), with probability at least $\varepsilon/(2+\varepsilon)$, there exists an (L, N)-network f_N such that

$$\| f - f_N \|_{L^2(\Omega)} \leq \frac{\sqrt{3}(2+\varepsilon)\pi v_{f,s,\Omega}}{N^{sL}} \leq \frac{11 v_{f,s,\Omega}}{N^{sL}},$$

with proper choice of ε in the last step. Finally, if f is complex-valued, we approximate the real and imaginary parts of the function separately to obtain (3.5). $\square$

Remark 3.9. We assume $\mathrm{supp}\ \widehat{f} \subset \{ \xi \in \mathbb{R}^d \mid |\xi|_1 \geq 1 \}$ in Lemma 3.8 because we want to obtain an upper bound depending only on $v_{f,s,\Omega}$. If we give up this condition, then the upper bound in Theorem 3.5 changes to $C\| f \|_{\mathscr{B}^s(\mathbb{R}^d)}/N^{sL}$ for some dimension-free constant C. The proof is essentially the same provided that the probability measure (3.6) is replaced by

$$\mu(\mathrm{d}\xi, \mathrm{d}r) = \frac{1}{Q}(1 + |\xi|_1)^{-s} |\widehat{f}(\xi)| \chi_{(0,1)}(r) \mathrm{d}\xi \mathrm{d}r,$$

We leave it to the interested reader.

Proof of Theorem 3.5. We write $f = f_1 + f_2$ with

$$f_1(x) = \int_{|\xi|_1 < 1} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \mathrm{d}\xi, \qquad f_2(x) = \int_{|\xi|_1 \geq 1} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \mathrm{d}\xi.$$

Then $v_{f_1,1,\Omega} \leq v_{f,s,\Omega}$ and $v_{f_2,s,\Omega} \leq v_{f,s,\Omega}$ because

$$\widehat{f_1}(\xi) = \widehat{f}(\xi) \chi_{[0,1)}(|\xi|_1) \qquad \text{and} \qquad \widehat{f_2}(\xi) = \widehat{f}(\xi) \chi_{[1,\infty)}(|\xi|_1).$$

We approximate f_1 with an (L, n_1)-network with $n_1 = \lceil N/6 \rceil$. By applying Lemma 3.3 to f_1, we obtain, there exists an $(1, n_1)$-network f_{1,n_1} such that

$$\| f_1 - f_{1,n_1} \|_{L^2(\Omega)} \leq \frac{2\upsilon_{f_1,1,\Omega}}{n_1^{1/2}} \leq \frac{2\sqrt{6}\upsilon_{f,s,\Omega}}{N^{sL}}.$$

Noting that an $(1, n_1)$-network can be represented by an (L, n_1)-network. We just need to fill the rest of the hidden layers with

$$t = \begin{cases} \mathrm{ReLU}(t), & t \geq 0, \\ -\mathrm{ReLU}(-t), & t < 0. \end{cases}$$

Secondly, we approximate f_2 with an (L, n_2)-network with $n_2 = \lceil 5N/6 \rceil$ and obtain the error estimate. Applying Lemma 3.8 we obtain, there exists an (L, n_2) network f_{2,n_2} such that

$$\| f_2 - f_{2,n_2} \|_{L^2(\Omega)} \leq \frac{22\pi\upsilon_{f_2,s,\Omega}}{n_2^s} \leq \frac{22\sqrt{6/5}\pi\upsilon_{f,s,\Omega}}{N^{sL}}.$$

These together with the triangle inequality give the estimate (3.2) and the total number of units in each layer is

$$n_1 + 2n_2 = \lceil N/6 \rceil + \lceil 5N/6 \rceil \leq N + 2.$$

If f is a real-valued function, then we let $f_N = \mathrm{Re}(f_{1,n_1} + f_{2,n_2})$, and the upper bound (3.2) still holds true. $\qquad\square$

Remark 3.10. The activation function of the last hidden layer of the (L, N)-network in Theorem 3.5 may be replaced by many other familiar activation functions such as Hyperbolic tangent, SoftPlus, ELU, Leaky ReLU, ReLU^k and so on. Because all these activation functions can be reduced to sigmoidal functions by certain shifting and scaling argument; e.g., for SoftPlus, we observe that $\mathrm{SoftPlus}(t) - \mathrm{SoftPlus}(t-1)$ is a sigmoidal function. Unfortunately, it is not easy to change ReLU of the first $L - 1$ hidden layers by other activation functions.

In what follows, we shall show that Theorem 3.5 is sharp if the activation function of the last hidden layer is Heaviside function. This example is adopted from [BN20b]. For readers' convenience, We reserve the proof in Appendix A.6.

Theorem 3.11. *For any fixed positive integers L, N and real numbers ε, s with $0 < \varepsilon, sL \leq 1/2$, there exists $f \in \mathscr{B}^s(\mathbb{R}^d)$ satisfying $\upsilon_{f,s,\Omega} \leq 1 + \varepsilon$ such that for any (L, N)-network f_N whose activation function σ in the last layer is the Heaviside function $\chi_{[0,\infty)}$, there holds*

$$(3.9) \qquad \| f - f_N \|_{L^2(\Omega)} \geq \frac{1 - \varepsilon}{8N^{sL}}.$$

4. Conclusion

We discuss the analytical functional properties of the spectral Barron space. The sharp embedding between the spectral Barron spaces and various classical function spaces have been established. The approximation rate has been proved for the deep ReLU neural networks when the spectral Barron space with a small smoothness index is employed as the target function space. There are still some unsolved problems, such as the sup-norm error and the higher-order convergence results for larger s, the relations among Barron type spaces, variational space and the Radon bounded variation space as well as understanding how these spaces are related to the classical function spaces, which will be pursued in the subsequent works.

References

[Bar92] A. R. Barron, *Neural net approximation*, Proc. 7th Yale Workshop on Adaptive and Learning Systems **1** (1992), 69–72.

[Bar93] ———, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory **39** (1993), no. 3, 930–945.

[Bar94] ———, *Approximation and estimation bounds for artificial neural networks*, Mach. Learn. **14** (1994), 115–133.

[BCDD08] A. R. Barron, A. Cohen, W. Dahmen, and R. A. DeVore, *Approximation and learning by greedy algorithms*, Ann. Statist. **36** (2008), no. 1, 64–94.

[BN20a] G. Bresler and D. Nagaraj, *A corrective view of neural networks: Representation, memorization and learning*, Proceedings of Machine Learning Research **125** (2020), 848–901.

[BN20b] ———, *Sharp representation theorems for ReLU networks with precise dependence on depth*, Advances in Neural Information Processing Systems **33** (2020), 10697–10706.

[BO13] Á. Bényi and T. Oh, *The Sobolev inequality on the torus revisited*, Publ. Math. Debrecen **83** (2013), no. 3, 359–374.

[Bre11] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext, Springer New York, NY, 2011.

[CLL21] Z. A. Chen, J. F. Lu, and Y. L. Lu, *On the representation of solutions to elliptic PDEs in Barron spaces*, Advances in Neural Information Processing Systems **34** (2021), 6454–6465.

[CLLZ23] Z. A. Chen, J. F. Lu, Y. L. Lu, and S. X. Zhou, *A regularity theory for static Schrödinger equations on $\mathbb{R}^d$ in spectral Barron spaces*, SIAM J. Math. Anal. **55** (2023), no. 1, 557–570.

[CPV23] A. Caragea, P. Petersen, and F. Voigtlaender, *Neural network approximation and estimation of classifiers with classification boundary in a Barron class*, Ann. Appl. Probab. **33** (2023), no. 4, 3039–3079.

[EMW19] W. E, C. Ma, and L. Wu, *A priori estimates of the population risk for two-layer neural networks*, Commun. Math. Sci. **17** (2019), no. 5, 1407–1425.

[EMW22] ———, *The Barron space and the flow-induced function spaces for neural network models*, Constr. Approx. **55** (2022), 369–406.

[EPGB21] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei, *Deep neural network approximation theory*, IEEE Trans. Inform. Theory **67** (2021), no. 5, 2581–2623.

[ES16] R. Eldan and O. Shamir, *The power of depth for feedforward neural networks*, Proceedings of Machine Learning Research **49** (2016), 907–940.

[EW22a] W. E and S. Wojtowytsch, *Representation formulas and pointwise properties for Barron functions*, Calc. Var. **61** (2022), 46.

[EW22b] ———, *Some observations on high-dimensional partial differential equations with Barron data*, Proceedings of Machine Learning Research **145** (2022), 253–269.

[GG02] Y. V. Galperin and K. Gröchenig, *Uncertainty principles as embeddings of modulation spaces*, J. Math. Anal. Appl. **274** (2002), no. 1, 181–202.

[GIJK23] P. Grohs, S. Ibragimov, A. Jentzen, and S. Koppensteiner, *Lower bounds for artificial neural network approximations: A proof that shallow neural networks fail to overcome the curse of dimensionality*, J. Complexity **77** (2023), Paper No. 101746, 47 pp.

[GKNV22] R. Gribonval, G. Kutyniok, M. Nielsen, and F. Voigtlaender, *Approximation spaces of deep neural networks*, Constr. Approx. **55** (2022), no. 1, 259–367.

[Gra14] L. Grafakos, *Classical Fourier Analysis*, 3rd ed., Graduate Texts in Mathematics, vol. 249, Springer New York, NY, 2014.

[GV23] P. Grohs and F. Voigtlaender, *Sobolev-type embeddings for neural network approximation spaces*, Constr. Approx. **57** (2023), no. 2, 579–599.

[Hör63] L. Hörmander, *Linear Partial Differential Operators*, Grundlehren der mathematischen Wissenschaften, vol. 116, Springer Berlin, Heidelberg, 1963.

[HKKR59] H. Helson, J.-P. Kahane, Y. Katznelson, and W. Rudin, *The function which operate on Fourier transforms*, Acta Math. **102** (1959), 135–157.

[IV21] I. Ivec and I. Vojnović, *H-distributions on Hörmander spaces*, J. Math. Anal. Appl. **495** (2021), no. 1, 124717.

[KB18] J. M. Klusowski and A. R. Barron, *Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ^1 and ℓ^0 controls*, IEEE Trans. Inform. Theory **64** (2018), no. 12, 7649–7656.

[KST20] T. Kato, M. Sugimoto, and N. Tomita, *Nonlinear operations on a class of modulation spaces*, J. Funct. Anal. **278** (2020), no. 9, 108447.

[Lév35] P. Lévy, *Sur la convergence absolue des séries de Fourier*, Compositio Math. **1** (1935), 1–14.

[LGM+17] H. Lee, R. Ge, T. Y. Ma, A. Risteski, and S. Arora, *On the ability of neural nets to express distributions*, Proceedings of Machine Learning Research **65** (2017), 1271–1296.

[LL01] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics, vol. 14, American Mathematical Society, 2001.

[LLW21] J. F. Lu, Y. L. Lu, and M. Wang, *A priori generalization analysis of the Deep Ritz Method for solving high dimensional elliptic partial differential equations*, Proceedings of Machine Learning Research **134** (2021), 3196–3241.

[LST12] E. Liflyand, S. Samko, and R. Trigub, *The Wiener algebra of absolutely convergent Fourier integrals: An overview*, Anal. Math. Phys. **2** (2012), no. 1, 1–68.

[LSYZ21] J. F. Lu, Z. W. Shen, H. Z. Yang, and S. J. Zhang, *Deep network approximation for smooth functions*, SIAM J. Math. Anal. **53** (2021), no. 5, 5465–5506.

[Luk62] Y. L. Luke, *Integrals of Bessel Functions*, McGraw-Hill Book Company, 1962.

[Luk69] ———, *The Special Functions and Their Approximations, Volume I*, Mathematics in Science and Engineering, vol. 53, Academic Press, 1969.

[MD19] H. Montanelli and Q. Du, *New error bounds for deep ReLU networks using sparse grids*, SIAM J. Math. Data Sci. **1** (2019), no. 1, 78–92.

[MM14] V. A. Mikhailets and A. A. Murach, *Hörmander Spaces, Interpolation, and Elliptic Problems*, De Gruyter Studies in Mathematics, vol. 60, Walter de Gruyter GmbH, Berlin/Boston, 2014.

[MM22] Y. Meng and P. B. Ming, *A new function space from Barron class and application to neural network approximation*, Commun. Comput. Phys. **32** (2022), no. 5, 1361–1400.

[MR01] F. Messina and L. Rodino, *Local solvability for nonlinear partial differential equations*, Nonlinear Anal. **47** (2001), no. 5, 2917–2927.

[MS73] A. M. Mathai and R. K. Saxena, *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Lecture Notes in Mathematics, vol. 348, Springer Berlin, Heidelberg, 1973.

[MSX22] L. M. Ma, J. W. Siegel, and J. C. Xu, *Uniform approximation rates and metric entropy of shallow neural networks*, Res. Math. Sci. **9** (2022), no. 3, 46.

[PN22] R. Parhi and R. D. Nowak, *What kinds of functions do deep neural networks learn? Insights from variational spline theory*, SIAM J. Math. Data Sci. **4** (2022), no. 2, 464–489.

[PN23] ———, *Near-minimax optimal estimation with shallow ReLU neural networks*, IEEE Trans. Inform. Theory **69** (2023), no. 2, 1125–1140.

[PT97] M. A. Pinsky and M. E. Taylor, *Pointwise Fourier inversion: a wave equation approach*, J. Fourier Anal. Appl. **3** (1997), no. 6, 647–703.

[PTT10] S. Pilipović, N. Teofanov, and J. Toft, *Micro-local analysis in Fourier Lebesgue and modulation spaces: part II*, J. Pseudo-Differ. Oper. Appl. **1** (2010), no. 3, 341–376.

[RS00] H. Reiter and J. D. Stegeman, *Classical Harmonic Analysis and Locally Compact Groups*, Clarendon Press, Oxford, 2000.

[SES22] I. Safran, R. Eldan, and O. Shamir, *Depth separations in neural networks: What is actually being separated?*, Constr. Approx. **55** (2022), no. 1, 225–257.

[SH20] J. Schmidt-Hieber, *Nonparametric regression using deep neural networks with ReLU activation function*, Ann. Statist. **48** (2020), no. 4, 1875–1897.

[ST95] W. Sickel and H. Triebel, *Hölder inequalities and sharp embeddings in function spaces of $B_{p,q}^s$ and $F_{p,q}^s$ type*, Z. Anal. Anwend. **14** (1995), no. 1, 105–140.

[SW71] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Mathematical Series, Princeton University Press, 1971.

[SX20] J. W. Siegel and J. C. Xu, *Approximation rates for neural networks with general activation functions*, Neural Networks **128** (2020), 313–321.

[SX22] ———, *High-order approximation rates for shallow neural networks with cosine and $ReLU^k$ activation functions*, Appl. Comput. Harmon. Anal. **58** (2022), 1–26.

[SX23] ———, *Characterization of the variation spaces corresponding to shallow neural networks*, Constr. Approx. **57** (2023), no. 3, 1109–1132.

[SX24] ———, *Sharp bounds on the approximation rates, metric entropy, and n-widths of shallow neural networks*, Found. Comput. Math. **24** (2024), no. 2, 481–537.

[Tai64] M. H. Taibleson, *On the theory of Lipschits spaces of distributions on Euclidean n-space. I. Principal properties*, J. Math. Mech. **13** (1964), no. 3, 407–479.

[Tai65] ———, *On the theory of Lipschits spaces of distributions on Euclidean n-space. II. Translation invariant operators, duality, and interpolation*, J. Math. Mech. **14** (1965), no. 5, 821–839.

[Tai66] ———, *On the theory of Lipschits spaces of distributions on Euclidean n-space II. Smoothness and integrability of Fourier transforms, smoothness of convolution kernels*, J. Math. Mech. **15** (1966), no. 6, 973–981.

[Tel16] M. Telgarsky, *Benefits of depth in neural networks*, Proceedings of Machine Learning Research **49** (2016), 1517–1539.

[Tri83] H. Triebel, *Theory of Function Spaces*, Modern Birkhäuser Classics, Birkhäuser, 1983.

[TS21] K. Tsuji and T. Suzuki, *Estimation error analysis of deep learning on the regression problem on the variable exponent Besov space*, Electron. J. Stat. **15** (2021), no. 1, 1869–1908.

[Wie32] N. Wiener, *Tauberian theorems*, Ann. of Math. (2) **33** (1932), no. 1, 1–100.

[Xu20] J. C. Xu, *Finite neuron method and convergence analysis*, Commun. Comput. Phys. **28** (2020), no. 5, 1707–1745.

APPENDIX A. SOME PROOF DETAILS

A.1. **Proof for** (2.3) **and** (2.4).

Proof. Note that $\widehat{\phi}_R$ is a radial function. By Lemma 2.1 and $\widehat{\phi}_R \in L^1(\mathbb{R}^d)$,

$$
\phi_R(x) = \int_{B_R} \left(1 - \frac{|\xi|^2}{R^2}\right)^\delta e^{2\pi i x \cdot \xi} \mathrm{d}\xi
$$

$$
= \int_{-R}^R e^{2\pi i |x|\xi_1} \mathrm{d}\xi_1 \int_{\xi_2^2 + \ldots \xi_d^2 < R^2 - \xi_1^2} \left(1 - \frac{|\xi|^2}{R^2}\right)^\delta \mathrm{d}\xi_2 \ldots \mathrm{d}\xi_d.
$$

Performing the polar transformation and changing the variable $t = r^2/(R^2 - \xi_1^2)$, we obtain

$$
\int_{\xi_2^2 + \ldots \xi_d^2 < R^2 - \xi_1^2} \left(1 - \frac{|\xi|^2}{R^2}\right)^\delta \mathrm{d}\xi_2 \ldots \mathrm{d}\xi_d
$$

$$
= \omega_{d-2} \int_0^{\sqrt{R^2 - \xi_1^2}} r^{d-2} \left(1 - \frac{\xi_1^2 + r^2}{R^2}\right)^\delta \mathrm{d}r
$$

$$
= \frac{\omega_{d-2}}{2} R^{d-1} \left(1 - \frac{\xi_1^2}{R^2}\right)^{\delta + (d-1)/2} \int_0^1 t^{(d-3)/2}(1-t)^\delta \mathrm{d}t
$$

$$
= \frac{\omega_{d-2}}{2} R^{d-1} \left(1 - \frac{\xi_1^2}{R^2}\right)^{\delta + (d-1)/2} B\left(\frac{d-1}{2}, \delta + 1\right).
$$

Substituting this equation into the previous one and changing the variable $\xi_1 = R\cos\theta$, we get

$$
\phi_R(x) = \frac{\omega_{d-2}}{2} R^{d-1} B\left(\frac{d-1}{2}, \delta + 1\right) \int_{-R}^R \left(1 - \frac{\xi_1^2}{R^2}\right)^{\delta + (d-1)/2} e^{2\pi i |x|\xi_1} \mathrm{d}\xi_1
$$

$$
= \frac{\pi^{(d-1)/2}\Gamma(\delta + 1)R^d}{\Gamma(\delta + (d+1)/2)} \int_0^\pi \cos(2\pi|x|R\cos\theta) \sin^{2\delta + d}\theta \mathrm{d}\theta
$$

$$
= \frac{\Gamma(\delta + 1)}{\pi^\delta |x|^{\delta + d/2}} R^{-\delta + d/2} J_{\delta + d/2}(2\pi|x|R),
$$

where we have used

$$
J_\nu(x) = \frac{(x/2)^\nu}{\pi^{1/2}\Gamma((d+1)/2)} \int_0^\pi \cos(x\cos\theta)\sin^{2\nu}\theta \mathrm{d}\theta, \qquad \nu > -\frac{1}{2}
$$

in the last step. The above integral representation of the first kind of Bessel function may be found in [Luk62, § 1.4.5, Eq. (4)].

For $s \geq 0$, a direct calculation gives

$$
\begin{aligned}
\upsilon_{\phi_R,s} &= \int_{B_R} |\xi|^s \left(1 - \frac{|\xi|^2}{R^2} \right)^\delta \mathrm{d}\xi = \omega_{d-1} \int_0^R r^{s+d-1} \left(1 - \frac{r^2}{R^2} \right)^\delta \mathrm{d}r \\
&= \frac{\omega_{d-1}}{2} R^{s+d} \int_0^1 t^{(s+d)/2-1} (1-t)^\delta \mathrm{d}t \\
&= \frac{\omega_{d-1}}{2} B \left(\frac{s+d}{2}, \delta + 1 \right) R^{s+d}.
\end{aligned}
$$

Therefore $\phi_R \in \mathscr{B}^s(\mathbb{R}^d)$. $\qquad\qquad\square$

A.2. **Proof for** (2.14).

Proof. To prove (2.14), we start with the following representation formula. If $\widehat{f} \in L^1(\mathbb{R}^d)$ is a radial function with $\widehat{f}(\xi) = g_0(|\xi|)$, then

(A.1)
$$
f(x) = \frac{2\pi}{|x|^{d/2-1}} \int_0^\infty g_0(r) r^{d/2} J_{d/2-1}(2\pi|x|r) \mathrm{d}r.
$$

If $d = 1$, then using Lemma 2.1, we obtain

$$
f(x) = \int_{\mathbb{R}} g_0(|\xi|) e^{2\pi i x \xi} \mathrm{d}\xi = 2 \int_0^\infty g_0(r) \cos(2\pi|x|r) \mathrm{d}r,
$$

which gives (A.1), where we have used the relation [Luk62, § 1.4.6, Eq. (7)]

$$
J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cos(x)
$$

in the last step.

For $d \geq 2$, combining Lemma 2.1 and [SW71, Ch. IV, Theorem 3.3], we obtain (A.1), which immediately implies

$$
\begin{aligned}
f_p(x) &= \frac{2\pi}{|x|^{d/2-1}} \int_0^1 r^{d(1/2-1/p')} J_{d/2-1}(2\pi|x|r) \mathrm{d}r \\
&= \omega_{d-1} \int_0^1 r^{d/p-1} {}_0F_1(; d/2; -\pi^2|x|^2 r^2) \mathrm{d}r,
\end{aligned}
$$

where we have used the relation [Luk62, § 1.4.1, Eq. (1)]

$$
J_\nu(x) = \frac{(x/2)^\nu}{\Gamma(\nu+1)} {}_0F_1(; \nu+1; -x^2/4)
$$

in the last step. Changing the variable $t = r^2$ and using the identity [Luk62, § 1.3.2, Eq. (2)]

$$
{}_1F_2(\rho; \rho + \sigma, \beta; x) = \frac{1}{B(\rho, \sigma)} \int_0^1 t^{\rho-1} (1-t)^{\sigma-1} {}_0F_1(; \beta; xt) \mathrm{d}t,
$$

we get (2.14). $\qquad\qquad\square$

A.3. **Proof for Lemma 2.8.**

Proof. The "if"-part is standard by [Tri83, § 2.3.2, Proposition 2; § 2.7.1, Theorem]. We illustrate the "only if"-part with an example, which is taken from [Tri83, § 2.3.9, Proof of Theorem].

Let $f_0 \in \mathscr{S}(\mathbb{R}^d)$ with $\mathrm{supp}(\widehat{f_0}) \subset \{ x \in \mathbb{R}^d \mid 1 \leq |x| \leq 3/2 \}$. Let $f_n(x) = f_0(2^{-n}x)$ for an integer n, then

$$\widehat{f_n}(\xi) = 2^{-dn}\widehat{f_0}(2^{-n}\xi) \qquad \text{and} \qquad \mathrm{supp}(\widehat{f_n}) \subset \{ x \in \mathbb{R}^d \mid 2^n \leq |x| \leq 3 \times 2^{n-1} \}.$$

Choose proper $\{\varphi_j\}_{j=0}^\infty \subset \mathscr{S}(\mathbb{R}^d)$ in the definition of Besov space such that $\varphi_0(x) = 1$ when $|x| \leq 3/2$ and $\varphi_j = 1$ on $\mathrm{supp}(\widehat{f_j})$ for $j \geq 1$, then

$$\mathrm{supp}(\widehat{f_n}) \cap \mathrm{supp}(\varphi_j) = \emptyset \quad \text{if} \quad n \geq 0 \quad \text{and} \quad n \neq j.$$

A direct calculation gives that

$$(\varphi_j \widehat{f_n})^\vee = \delta_{0n} f_n, \quad \text{if} \quad n \leq 0 \qquad \text{and} \qquad (\varphi_j \widehat{f_n})^\vee = \delta_{jn}\widehat{f_n}, \quad \text{if} \quad n > 0.$$

By definition, when $n < 0$,

$$\| f_n \|_{B_{p,q}^\alpha(\mathbb{R}^d)} = \| f_n \|_{L^p(\mathbb{R}^d)} = 2^{-dn/p}\| f_0 \|_{L^p(\mathbb{R}^d)}.$$

Let $n \to -\infty$ with the embedding relation $B_{p_1,q_1}^{\alpha_1}(\mathbb{R}^d) \hookrightarrow B_{p_2,q_2}^{\alpha_2}(\mathbb{R}^d)$ yields $p_1 \leq p_2$. Similarly, when $n > 0$,

$$\| f_n \|_{B_{p,q}^\alpha(\mathbb{R}^d)} = 2^{\alpha n}\| f_n \|_{L^p(\mathbb{R}^d)} = 2^{(\alpha-d/p)n}\| f_0 \|_{L^p(\mathbb{R}^d)}.$$

Let $n \to +\infty$ with the embedding relation implies $\alpha_1 - d/p_1 \geq \alpha_2 - d/p_2$. Finally if $\alpha_1 - d/p_1 = \alpha_2 - d/p_2$, then $q_1 \leq q_2$ proved in [ST95, Theorem 3.2.1]. $\qquad\square$

A.4. **Proof for** (2.18).

Proof. If $1 \leq p < \infty$ and $\alpha > 0$, then $B_{p,1}^\alpha(\mathbb{R}^d)$ is equivalent to the space defined in [Tri83, § 2.5.7, Theorem]

$$\Lambda_{p,q}^\alpha(\mathbb{R}^d) := \left\{ f \in W_p^{[\alpha]}(\mathbb{R}^d) \mid \int_{\mathbb{R}^d} \frac{\| \Delta_h^2(\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)}^q}{|h|^{d+\{\alpha\}q}} \mathrm{d}h < \infty \right\}$$

equipped with the norm

$$\| f \|_{\Lambda_{p,q}^\alpha(\mathbb{R}^d)} := \| f \|_{W_p^{[\alpha]}(\mathbb{R}^d)} + \left(\int_{\mathbb{R}^d} \frac{\| \Delta_h^2(\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)}^q}{|h|^{d+\{\alpha\}q}} \mathrm{d}h \right)^{1/q}$$

for $1 \leq q < \infty$, and

$$\Lambda_{p,\infty}^\alpha(\mathbb{R}^d) := \left\{ f \in W_p^{[\alpha]}(\mathbb{R}^d) \mid \sup_{h \in \mathbb{R}^d \setminus \{0\}} |h|^{-\{\alpha\}}\| \Delta_h^2(\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)} < \infty \right\}$$

equipped with the norm

$$\| f \|_{\Lambda_{p,\infty}^\alpha(\mathbb{R}^d)} := \| f \|_{W_p^{[\alpha]}(\mathbb{R}^d)} + \sup_{h \in \mathbb{R}^d \setminus \{0\}} |h|^{-\{\alpha\}}\| \Delta_h^2(\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)}.$$

Here $\alpha = [\alpha] + \{\alpha\}$ with integer $[\alpha]$ and $0 < \{\alpha\} \le 1$, and $\Delta_h^2 f(x) = f(x + 2h) - 2f(x + h) + f(x)$; see [Tri83, § 2.2.2, Eq. (9)].

For any nonnegative integer k, a direct calculation gives

$$\nabla^k \psi_n(x) = \sum_{j=0}^k \frac{c_j x^{\beta_j}}{(1 + in)^{d/2 + j}} e^{-\pi |x|^2 / (1 + in)}$$

for some constants $\{c_j\}_{j=0}^k$, and the multi-index $\beta_j = (\beta_{j1}, \ldots, \beta_{jd})$ satisfies $|\beta_j| \le j$ with $x^{\beta_j} = x_1^{\beta_{j1}} \ldots x_d^{\beta_{jd}}$. Then

$$\| \nabla^k \psi_n \|_{L^p(\mathbb{R}^d)} \le C \sum_{j=0}^k \frac{1}{(1 + n^2)^{d/4 + j/2}} \| |x|^{|\beta_j|} e^{-\pi |x|^2 / (1 + n^2)} \|_{L^p(\mathbb{R}^d)}.$$

A direct calculation gives

$$\int_{\mathbb{R}^d} |x|^{|\beta_j| p} e^{-\pi p |x|^2 / (1 + n^2)} \mathrm{d}x = (1 + n^2)^{(d + |\beta_j| p)/2} \int_{\mathbb{R}^d} |y|^{|\beta_j| p} e^{-\pi p |y|^2} \mathrm{d}y$$

$$= \frac{\omega_{d-1} \Gamma\left((|\beta_j| p + d)/2\right)}{2(p\pi)^{(|\beta_j| p + d)/2}} (1 + n^2)^{(d + |\beta_j| p)/2}.$$

Therefore, there exists C depending only on d, p, k such that

$$(A.2) \qquad \| \nabla^k \psi_n \|_{L^p(\mathbb{R}^d)} \le C \sum_{j=0}^k \frac{(1 + n^2)^{d/(2p) + |\beta_j|/2}}{(1 + n^2)^{d/4 + j/2}} \le C(1 + n^2)^{-d(p-2)/(4p)}.$$

If $f \in W_p^2(\mathbb{R}^d)$, then

$$\Delta_h^2 f(x) = \int_0^1 \mathrm{d}t \int_t^{1+t} \nabla^2 f(x + sh) \mathrm{d}s h \cdot h = \int_0^2 \nabla^2 f(x + sh) \mathrm{d}s \int_{\max(s-1,0)}^{\min(s,1)} \mathrm{d}t h \cdot h.$$

Therefore,

$$|\Delta_h^2 f(x)| \le |h|^2 \int_0^2 |\nabla^2 f(x + sh)| \mathrm{d}s.$$

By the Minkowski's inequality, we obtain

$$\| \Delta_h^2 f(x) \|_{L^p(\mathbb{R}^d)} \le |h|^2 \int_0^2 \| \nabla^2 f(\cdot + sh) \|_{L^p(\mathbb{R}^d)} \mathrm{d}s = 2|h|^2 \| \nabla^2 f \|_{L^p(\mathbb{R}^d)}.$$

Splitting the integral part of the $\Lambda_{p,q}^\alpha$-norm into two parts, we get

$$\int_{|h| < 1} \frac{\| \Delta_h^2 f \|_{L^p(\mathbb{R}^d)}^q}{|h|^{d + \{\alpha\} q}} \mathrm{d}h + \int_{|h| > 1} \frac{\| \Delta_h^2 f \|_{L^p(\mathbb{R}^d)}^q}{|h|^{d + \{\alpha\} q}} \mathrm{d}h$$

$$\le 2^q \| \nabla^2 f \|_{L^p(\mathbb{R}^d)}^q \int_{|h| < 1} h^{(2 - \{\alpha\}) q - d} \mathrm{d}h + 4^q \| f \|_{L^p(\mathbb{R}^d)}^q \int_{|h| > 1} h^{-d - \{\alpha\} q} \mathrm{d}h$$

$$= 2^q \omega_{d-1} \left(\frac{\| \nabla^2 f \|_{L^p(\mathbb{R}^d)}^q}{(2 - \{\alpha\}) q} + \frac{2^q \| f \|_{L^p(\mathbb{R}^d)}^q}{\{\alpha\} q} \right)$$

$$\le \frac{2^{2q} \omega_{d-1}}{\{\alpha\} q} \| f \|_{W_p^2(\mathbb{R}^d)}^q.$$

In the same manner we can see that when $q = \infty$,

$$\sup_{h \in \mathbb{R}^d \backslash \{0\}} |h|^{-\{\alpha\}} \| \Delta_h^2 (\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)}$$

$$\leq \sup_{0 < |h| < 1} |h|^{-\{\alpha\}} \| \Delta_h^2 (\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)} + \sup_{|h| \geq 1} |h|^{-\{\alpha\}} \| \Delta_h^2 (\nabla^{[\alpha]} f) \|_{L^p(\mathbb{R}^d)}$$

$$\leq 2 \| \nabla^2 f \|_{L^p(\mathbb{R}^d)} \sup_{0 < |h| < 1} |h|^{2 - \{\alpha\}} + 4 \| f \|_{L^p(\mathbb{R}^d)} \sup_{|h| \geq 1} |h|^{-\alpha}$$

$$\leq 4 \| f \|_{W_p^2(\mathbb{R}^d)}.$$

Note that $\nabla^{[\alpha]} \psi_n \in W_p^2(\mathbb{R}^d)$, a combination of the above inequality and (A.2) yields

$$\| \psi_n \|_{\Lambda_{p,q}^\alpha(\mathbb{R}^d)} \leq C \| \psi_n \|_{W_p^{[\alpha]+2}(\mathbb{R}^d)} \leq C(1 + n^2)^{-d(p-2)/(4p)},$$

where C is a constant depending on p, α and d but independent of n. So does $\| \psi_n \|_{B_{p,q}^\alpha(\mathbb{R}^d)}$. $\square$

A.5. Proof for Lemma 3.6.

Proof. Firstly we show that $g_{,n}$ is symmetric about $x = 1/2$. By definition,

$$g_{,n}(1 - t) = g(n(1 - t) - j) = g(nt - n + j + 1) = g(nt - k) = g_{,n}(t)$$

for some integers j, k satisfying $0 \leq j, k \leq n - 1$ and $k = n - j - 1$.

For a fixed $t \in [0, 1]$, there exist integers j, k satisfying $0 \leq j \leq 2n_1 - 1, 0 \leq k \leq n_2 - 1$ such that $0 \leq 2n_1 n_2 t - n_2 j - k \leq 1$, then $0 \leq n_2(2n_1 t - j) - k \leq 1$ and

$$g_{,2n_1 n_2}(t) = g(2n_1 n_2 t - n_2 j - k) = g(n_2(2n_1 t - j) - k) = g_{,n_2}(2n_1 t - j).$$

By definition,

$$\beta_{,n}(t) = \begin{cases} 2nt - 2j, & 0 \leq nt - j \leq 1/2, \\ 2 + 2j - 2nt, & 1/2 \leq nt - j \leq 1, \end{cases} \quad j = 0, \ldots, n - 1.$$

If j is even, then $j = 2l$ for some integer l satisfying $0 \leq l \leq n_1 - 1$ and $0 \leq n_1 t - l \leq 1/2$. Therefore

$$g_{,n_2}(\beta_{,n_1}(t)) = g_{,n_2}(2n_1 t - 2l) = g_{,n_2}(2n_1 t - j).$$

Otherwise j is odd, then $j = 2l + 1$ for some integer l satisfying $0 \leq l \leq n_1 - 1$ and $1/2 \leq n_1 t - l \leq 1$. Therefore

$$g_{,n_2}(\beta_{,n_1}(t)) = g_{,n_2}(2 + 2l - 2n_1 t) = g_{,n_2}(1 + j - 2n_1 t) = g_{,n_2}(2n_1 t - j).$$

This gives $g_{,n_2} \circ \beta_{,n_1} = g_{,2n_1 n_2}$ on $[0, 1]$. $\square$

A.6. **Proof for Theorem 3.11.**

Lemma A.1. *Given* $n, R > 0$ *and let* $f(x) = \cos(2\pi n x_1)e^{-\pi|x|^2/R}$, *then* $f \in \mathscr{B}^s(\mathbb{R}^d)$ *with*

$$(A.3) \qquad \upsilon_{f,s,\Omega} \leq \left(n + \frac{d}{\pi\sqrt{R}}\right)^s \qquad for \quad 0 \leq s \leq 1.$$

Proof. For any $R > 0$, the Fourier transform of the dilated Guass function $e^{-\pi x_j^2/R}$ reads as

$$\widehat{e^{-\pi x_j^2/R}} = \sqrt{R}e^{-\pi R \xi_j^2}.$$

A direct calculation gives

$$\begin{aligned}\widehat{f}(\xi) =& \frac{1}{2}\int_{\mathbb{R}} e^{-\pi x_1^2/R}\left(e^{-2\pi i x_1(\xi_1-n)} + e^{-2\pi i x_1(\xi_1+n)}\right)\mathrm{d}x_1 \prod_{j=2}^d \int_{\mathbb{R}} e^{-\pi x_j^2/R - 2\pi i x_j \xi_j}\,\mathrm{d}x_j \\ =& \frac{R^{d/2}}{2}\left(e^{-\pi R(\xi_1-n)^2} + e^{-\pi R(\xi_1+n)^2}\right)\prod_{j=2}^d e^{-\pi R \xi_j^2} \\ =& R^{d/2}e^{-\pi R(|\xi|^2+n^2)}\cosh(2\pi n R\xi_1).\end{aligned}$$

It is clear that $f, \widehat{f} \in L^1(\mathbb{R}^d)$ and the pointwise Fourier inversion theorem holds true, and

$$\upsilon_{f,0,\Omega} = \int_{\mathbb{R}^d}|\widehat{f}(\xi)|\mathrm{d}\xi = \int_{\mathbb{R}^d}\widehat{f}(\xi)\mathrm{d}\xi = f(0) = 1,$$

where we have used the positiveness of $\widehat{f}$.

Next, using the elementary identities

$$\sqrt{R}\int_{\mathbb{R}}e^{-\pi R\xi_j^2}\mathrm{d}\xi_j = 1 \quad \text{and} \quad \sqrt{R}\int_{\mathbb{R}}|\xi_j|e^{-\pi R\xi_j^2}\mathrm{d}\xi_j = \frac{1}{\pi\sqrt{R}},$$

we obtain

$$\begin{aligned}\upsilon_{f,1,\Omega} =& R^{d/2}\int_{\mathbb{R}}|\xi_1|e^{-\pi R(\xi_1-n)^2}\mathrm{d}\xi_1\prod_{j=2}^d\int_{\mathbb{R}}e^{-\pi R\xi_j^2}\mathrm{d}\xi_j \\ &+ R^{d/2}\int_{\mathbb{R}}e^{-\pi R(\xi_1-n)^2}\mathrm{d}\xi_1\sum_{j=2}^d\int_{\mathbb{R}}|\xi_j|e^{-\pi R\xi_j^2}\mathrm{d}\xi_j\prod_{k\neq 1,j}\int_{\mathbb{R}}e^{-\pi R\xi_k^2}\mathrm{d}\xi_k \\ \leq& \sqrt{R}\int_{\mathbb{R}}(|\xi_1 - n| + n)e^{-\pi R(\xi_1-n)^2}\mathrm{d}\xi_1 + \frac{d-1}{\pi\sqrt{R}} \\ =& n + \frac{d}{\pi\sqrt{R}}.\end{aligned}$$

Using the interpolation inequality (2.11), we obtain (A.3). $\qquad\square$

Proof for Theorem 3.11. Define $n = 2^{L+2}N^L$ and $f(x) = n^{-s}\cos(2\pi n x_1)e^{-\pi|x|^2/R}$ with large enough R such that $\upsilon_{f,s,\Omega} \leq 1 + \varepsilon$ by Lemma A.1 and $e^{-\pi|x|^2/R} \geq 1 - \varepsilon$

when $x \in \Omega$. Fix $x_2, \ldots, x_d$, then any (L, N)-network f_N can be viewed as an one-dimensional (L, N)-network, i.e. $f_N(\cdot, x_2, \ldots, x_d) : [0, 1] \to \mathbb{C}$. Divide $[0, 1]$ into n-internals of $[j/n, (j+1)/n]$ with $j = 0, \ldots, n-1$. There exists at least $n - 2^{L+1}N^L = 2^{L+1}N^L$ intervals such that f_N does not change sign on those intervals [Tel16, Lemma 3.2]. Without loss of generality, we assume $f_N(\cdot, x_2, \ldots, x_d) \geq 0$ on some interval $[j/n, (j+1)/n]$, then

$$\int_{j/n}^{(j+1)/n} (f(x) - f_N(x))^2 \mathrm{d}x_1 \geq \frac{(1-\varepsilon)^2}{n^{2s}} \int_{(4j+1)/(4n)}^{(4j+3)/(4n)} \cos^2(2\pi n x_1) \mathrm{d}x_1 \geq \frac{(1-\varepsilon)^2}{4n^{2s+1}},$$

because $\cos(2\pi n x_1) \leq 0$ when $2\pi j + \pi/2 \leq 2\pi n x_1 \leq 2\pi j + 3\pi/2$.

Summing up these $n - 2^{L+1}N^L$ intervals gives

$$\| f - f_N \|_{L^2(\Omega)}^2 \geq \int_{[0,1]^{d-1}} \mathrm{d}x_2 \ldots \mathrm{d}x_d \int_0^1 (f(x) - f_N(x))^2 \mathrm{d}x_1$$

$$\geq \frac{2^{L+1}N^L(1-\varepsilon)^2}{4n^{2s+1}} \geq \frac{(1-\varepsilon)^2}{2^{2sL+4s+3}N^{2sL}} \geq \frac{(1-\varepsilon)^2}{64N^{2sL}}.$$

Simultaneously squaring off both sides of the inequality, we obtain

$$\| f - f_N \|_{L^2(\Omega)} \geq \frac{1-\varepsilon}{8N^{sL}}.$$

$\square$