Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2022-0151

Vol. **x**, No. x, pp. 1-40
xxx 202x

A New Function Space from Barron Class and Application to Neural Network Approximation

Yan Meng[1] and Pingbing Ming[2,3,*]

[1] *School of Mathematics, Renmin University, Beijing 100872, China.*
[2] *LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, No. 55, East Road Zhong-Guan-Cun, Beijing 100190, China.*
[3] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.*

Received 27 May 2022; Accepted (in revised version) 12 October 2022

Abstract. We introduce a new function space, dubbed as the Barron spectrum space, which arises from the target function space for the neural network approximation. We give a Bernstein type sufficient condition for functions in this space, and clarify the embedding among the Barron spectrum space, the Bessel potential space, the Besov space and the Sobolev space. Moreover, the unexpected smoothness and the decaying behavior of the radial functions in the Barron spectrum space have been investigated. As an application, we prove a dimension explicit L^q error bound for the two-layer neural network with the Barron spectrum space as the target function space, the rate is dimension independent.

AMS subject classifications: 65N30, 65M12, 41A46, 35J25

Key words: Fourier transform, Besov space, Sobolev space, radial function, neural network.

1 Introduction

Several target function spaces have been scattered in the literature for the neural network approximation, such as the potential space, the reproducing kernel space [23, 29, 42], the Sobolev space [44], the Besov space [17] and its variant of the dominating mixed smoothness [54], among many others. In Barron's seminal work [2], he proved that for a function that has a finite first moment of the magnitude of the Fourier transform, the convergence rate for a feedforward artificial neural network with sigmoidal nonlinearity is

*Corresponding author. *Email addresses:* `mengyan@ruc.edu.cn` (Y. Meng), `mpb@lsec.cc.ac.cn` (P. B. Ming)

$\mathcal{O}(n^{-1/2})$ with n the number of the neurons. Roughly speaking, such function class may be rephrased as

$$f \quad \text{has Fourier transform} \quad \widehat{f} \quad \text{with} \quad \int_{\mathbb{R}^d} |x||\widehat{f}(x)|\mathrm{d}x < \infty. \tag{1.1}$$

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

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

The novelty of this result lies in the fact that the convergence rate is independent of the dimension of the ambient space, and cracks the curse of the dimensionality [4, 16]. The spectrum norm was extended to second order to study the approximation rate for hinging functions by Breiman in [9]. Hornik et al [26] introduced the spectrum norm of arbitrary positive integer order m:

$$\int_{\mathbb{R}^d} \max\{1, |x|^m\} |\widehat{f}(x)| \mathrm{d}x.$$

Several works have been devoted to study the spectrum norm in the literature; see [10, §7.2] and [16, 45]. In a series of work [19, 20], E et al have defined a function class, which is dubbed as Barron space. Barron space contains *infinitely wide* neural networks with certain controls over the parameters, which depends on the activation function used in the neural network. The relation among Barron spaces and the so-called *Fourier-analytic Barron space* have been established in [11, §7]. They have studied the pointwise behavior of the functions in Barron space, and have proved the direct and inverse approximation theorems for the two-layer neural network approximation. It is worth mentioning that the approximation class for the deep neural network have recently been investigated in [19, 24].

Motivated by (1.1), for $s \in \mathbb{R}$ and $1 \le p \le 2$, we introduce a new function space $\mathcal{B}_{s,p}(\mathbb{R}^d)$, called Barron spectrum space, which consists of $f \in L^p(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} |x|^s |\widehat{f}(x)| \mathrm{d}x < \infty$ (see Definition 2.1). Compared with the original Barron class or the Fourier-analytic Barron space, we include the L^p-norm in addition to the spectrum norm and our definition is independent of the choice of the activation functions. It is a Banach space as shown in the next part. We shall study the properties of $\mathcal{B}_{s,p}(\mathbb{R}^d)$ and clarify the relationship among $\mathcal{B}_{s,p}(\mathbb{R}^d)$ and certain commonly used target function spaces for the neural network approximation. For any $s > -d/p$ and $1 \le p \le 2$, we prove the embedding

$$B^{s+d/p}_{p,1}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,p}(\mathbb{R}^d) \hookrightarrow B^s_{\infty,1}(\mathbb{R}^d)$$

holds, where $B^\alpha_{p,q}(\mathbb{R}^d)$ is the Besov space with $\alpha \in \mathbb{R}$ and $1 \le p, q \le \infty$. With the aid of this inclusion, we establish the embedding between $\mathcal{B}_{s,p}(\mathbb{R}^d)$ and the Sobolev space. Moreover, we prove the relation between $\mathcal{B}_{s,p}(\mathbb{R}^d)$ and the Bessel potential space. Another consequence of the above inclusion is that the function in $\mathcal{B}_{s,p}$ with a nonnegative index s is smooth. We give a precise estimate for this statement; cf. Theorem 4.4. Moreover, the smoothness of the functions in $\mathcal{B}_{s,p}$ with a negative index s may be characterized by L^p-modulus of continuity as addressed in Section 3.

To understand the radial basis function neural networks [41], we study the radial function in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ with a nonnegative index s. Roughly speaking, the radial functions in $\mathcal{B}_{s,p}$ with certain p become smoother as the increasing of the dimension, though at the expense of an exponentially large embedding constant with respect to the dimension. Moreover, we prove a decaying result in the spirit of Strauss' inequality [53], which shows that the decaying rate of the radial function in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ at infinity is $(1-d)/2$.

As an application, we derive the approximation rate of the two-layer neural network with $\mathcal{B}_{s,p}(\mathbb{R}^d)$ as the target function space. For the two-layer neural networks with commonly used activation functions, we prove L^q approximation rate with an explicit dimension dependence. The rate is $\mathcal{O}(n^{-\gamma(q)})$ with $\gamma(q)=\min(1-1/q,1/2)$ and n the number of neurons, which is dimension independent. This recovers the classical approximation results [2, 9, 30, 46, 48] when $q=2$, while we only need a target function with a smaller smoothness index s. In case of $1\le q<2$, the rate $1-1/q$ may be improved to $1/2$ when $s\ge 1/2$. We conclude that $\mathcal{B}_{s,p}(\mathbb{R}^d)$ may also be used as a target function space for the neural network approximation in high dimension.

Barron spectrum space $\mathcal{B}_{s,p}(\mathbb{R}^d)$ may be regarded as one realization of Barron's definition (1.1). It definitely admits other realizations such as f is a tempered distribution, and endowed with merely the spectrum norm. This realization is a special case of the so-called Fourier Besov space [27] or Fourier-Herz space [12, 33], which has been exploited to prove the global existence of the solution for Navier-Stokes equation [34] and Keller-Segel equations [27]. We shall study the relationship among $\mathcal{B}_{s,p}(\mathbb{R}^d)$ and this realization as well as some other target function spaces in the neural network approximation [3, 10, 37, 48], just name a few of them, in the future work.

The paper is organized as follows. In Section 2, we introduce the Barron spectrum space and prove some basic properties. In Section 3, we give some sufficient conditions for functions in the Barron spectrum spaces. In Section 4, we clarify its relation to the Sobolev space, the Besov space and the Bessel potential space with suitable indices. In Section 5, we discuss the decay behavior of the radial functions in the Barron spectrum spaces. And in the last section, we show an application of the Barron spectrum space for the approximation of two-layer neural networks.

Finally, we make some conventions on notation. Throughout this paper, we denote by C a positive constant which is independent of the main parameters, but may vary from line to line. The symbol $f\lesssim g$ means that there exists a positive constant C such that $f\le Cg$, and $f\gtrsim g$ means that there exists a positive constant C such that $f\ge Cg$. Moreover, $f\sim g$ abbreviates $f\lesssim g\lesssim f$. Given any $p\in[1,\infty]$, let $p':=p/(p-1)$ be its conjugate index. Also, for any subset $E\subset\mathbb{R}^d$, χ_E denotes its characteristic function. For any $x\in\mathbb{R}^d$, let $\|x\|_p:=(\sum_{i=1}^d|x_i|^p)^{1/p}$ be the ℓ_p norm of x with $1\le p<\infty$. We may use $|x|$ in lieu of $\|x\|_2$ without abuse of the notations. We denote the unit ball of the ℓ_1 norm on $\mathbb{R}^d$ by B_1^d, whose volume is $2^d/d!$. The ball of the ℓ_1 norm on $\mathbb{R}^d$ centered at the origin with radius $r>0$ is denoted by $B_1^d(r)$, whose volume is $2^d r^d/d!$. We denote the surface area of the unit sphere by ω_{d-1}, i.e., $\omega_{d-1}=2\pi^{d/2}/\Gamma(d/2)$.

2 Barron spectrum space and its basic properties

Definition 2.1. Let $1\le p\le 2$ and $s\in\mathbb{R}$. The Barron spectrum space consists of functions $f\in L^p(\mathbb{R}^d)$ satisfying

$$\|f\|_{v_s}:=\int_{\mathbb{R}^d}\|x\|_1^s|\widehat{f}(x)|\mathrm{d}x<\infty.$$

Moreover, define

$$\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}:=\|f\|_{L^p(\mathbb{R}^d)}+\|f\|_{v_s}.$$

We firstly show that $\mathcal{B}_{s,p}(\mathbb{R}^d)$ is a Banach space.

Theorem 2.1. (i) *For $1\le p\le 2$ and $-\infty<s<\infty$, $\mathcal{B}_{s,p}(\mathbb{R}^d)$ is a Banach space;*

(ii) *$\mathcal{B}_{s,p}(\mathbb{R}^d)$ is not a Banach space if the norm $\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}$ is replaced by $\|f\|_{v_s}$ whenever $s+d/p>0$.*

We shall frequently use the Hausdorff-Young inequality [50]. For any $1\le p\le 2$ and $f\in L^p(\mathbb{R}^d)$, there holds

$$\left\|\widehat{f}\right\|_{L^{p'}(\mathbb{R}^d)}\le\|f\|_{L^p(\mathbb{R}^d)}. \tag{2.1}$$

Proof of Theorem 2.1. (i) It is easy to verify that $\|\cdot\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}$ is a norm and $\mathcal{B}_{s,p}(\mathbb{R}^d)$ is a linear metric space. Thus, it remains to check that $\mathcal{B}_{s,p}(\mathbb{R}^d)$ is complete. For any Cauchy series $\{f_j\}_{j=1}^\infty$ in $\mathcal{B}_{s,p}(\mathbb{R}^d)$, there exists $f\in L^p(\mathbb{R}^d)$ such that

$$\lim_{j\to\infty}\left\|f-f_j\right\|_{L^p(\mathbb{R}^d)}=0.$$

From the Hausdorff-Young inequality (2.1), it follows that

$$\lim_{j\to\infty}\left\|\widehat{f}-\widehat{f_j}\right\|_{L^{p'}(\mathbb{R}^d)}=0.$$

Therefore, there exists a subsequence $\{f_{j_k}\}$ such that

$$\lim_{k\to\infty}\widehat{f}_{j_k}(x)=\widehat{f}(x)\qquad a.e.\quad x\in\mathbb{R}^d.$$

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

$$\mu(E):=\int_E\|x\|_1^s\,\mathrm{d}x.$$

From the fact that $\{f_j\}$ is the Cauchy sequence in $\mathcal{B}_{s,p}(\mathbb{R}^d)$, it follows that

$$\lim_{k,i\to\infty}\int_{\mathbb{R}^d}|\widehat{f}_{j_k}(x)-\widehat{f}_{j_i}(x)|d\mu(x)=0.$$

This implies that there exists a subsequence $\{f_{j_{k_m}}\}$ and $g\in L^1(\mu)$ such that

$$\lim_{m\to\infty}\widehat{f}_{j_{k_m}}(x)=g(x)\qquad \mu-a.e.\quad x\in\mathbb{R}^d.$$

It is easy to verify that for any measurable set $E\subset\mathbb{R}^d$, $\mu(E)=0$ is equivalent to $|E|=0$. This leads to

$$\lim_{m\to\infty}\widehat{f}_{j_{k_m}}(x)=g(x)\qquad a.e.\quad x\in\mathbb{R}^d.$$

Therefore,

$$g(x)=\widehat{f}(x)\qquad a.e.\quad x\in\mathbb{R}^d,$$

and then $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$. This proves that $\mathcal{B}_{s,p}(\mathbb{R}^d)$ is a Banach space.

(ii) If $\mathcal{B}_{s,p}(\mathbb{R}^d)$ is also complete equipped with the norm $\|\cdot\|_{v_s}$, then by Banach's theorem, we would have that for any $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$,

$$\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}\sim\|f\|_{v_s}.$$

Therefore,

$$\|f\|_{L^p(\mathbb{R}^d)}\lesssim\|f\|_{v_s}. \tag{2.2}$$

We shall show this is not the case by the following example.

Define

$$f_n(x):=\left(\sum_{k=1}^n 2^{k(s+2d)}\phi_k\right)^{\vee}(x),$$

where $\phi_k(x)=\prod_{i=1}^d\psi_k(x_i)$ with

$$\psi_k(t)=\begin{cases}2^{-k}-|t|, & \frac{3}{4}2^{-k}<|t|\le 2^{-k},\\ |t|-2^{-k-1}, & 2^{-k-1}\le|t|<\frac{3}{4}2^{-k},\\ 0, & \text{otherwise.}\end{cases}$$

We only give the details for $d=1$ because the extension to $d>1$ is straightforward.

A direct calculation gives

$$f_n(x)=\frac{1}{2\pi^2x^2}\sum_{k=1}^n 2^{k(s+2)}\left(2\cos\frac{3\pi x}{2^k}-\cos\frac{2\pi x}{2^k}-\cos\frac{\pi x}{2^k}\right).$$

By trigonometric identity, we have

$$2\cos\frac{3\pi x}{2^k}-\cos\frac{2\pi x}{2^k}-\cos\frac{\pi x}{2^k}=-2\sin\frac{\pi x}{2^{k+1}}\left(2\sin\frac{5\pi x}{2^{k+1}}+\sin\frac{3\pi x}{2^{k+1}}\right).$$

Using the elementary inequality $|\sin x|\leq|x|$, we obtain

$$\left|2\cos\frac{3\pi x}{2^k}-\cos\frac{2\pi x}{2^k}-\cos\frac{\pi x}{2^k}\right|\leq\frac{13}{2}\frac{\pi^2x^2}{4^k}.$$

This gives

$$|f_n(x)|\leq 4\sum_{k=1}^n 2^{ks},\qquad |x|\leq 1.$$

Moreover, for $|x|>1$, we have

$$|f_n(x)|\leq\frac{2}{\pi^2x^2}\sum_{k=1}^n 2^{k(s+2)}.$$

Using the above two estimates, we get

$$\begin{aligned}\|f_n\|_{L^p(\mathbb{R})}^p&=2\int_0^\infty|f_n(x)|^p\,\mathrm{d}x\\&\leq 2\left(\sum_{k=1}^n 2^{k(s+2)}\right)^p\left(4^p+\left(\frac{2}{\pi^2}\right)^p\int_1^\infty x^{-2p}\,\mathrm{d}x\right)\\&=2\left(\sum_{k=1}^n 2^{k(s+2)}\right)^p\left(4^p+\left(\frac{2}{\pi^2}\right)^p\frac{1}{2p-1}\right).\end{aligned}$$

Hence,

$$\|f_n\|_{L^p(\mathbb{R})}\leq 2^{1/p}\left(4+\frac{2}{\pi^2}\left(\frac{1}{2p-1}\right)^{1/p}\right)\sum_{k=1}^n 2^{k(s+2)}\lesssim 2^{n(s+2)}.$$

This means that for any fixed n,

$$\|f_n\|_{L^p(\mathbb{R})}<\infty.$$

By the Hausdorff-Young inequality,

$$\|f_n\|_{L^p(\mathbb{R})}\geq\left\|\widehat{f}_n\right\|_{L^{p'}(\mathbb{R})}=\frac{1}{4}(p'+1)^{1/p'}\left(\sum_{k=1}^n 2^{k((s+1)p'-1)}\right)^{1/p'}\sim 2^{n(s+1/p)}.$$

On the other hand, a direct computation gives

$$\|f_n\|_{v_s}=\sum_{k=1}^n 2^{k(s+2)}\int_{\mathbb{R}}|x|^s\psi_k(x)\,\mathrm{d}x=\frac{1}{2(s+2)}\sum_{k=1}^n\frac{1}{2^k}<\frac{1}{2(s+2)}.$$

For such f_n with sufficiently large n, it is impossible for the inequality (2.2) to be true because the right-hand side of (2.2) is $\mathcal{O}(1)$, while the left-hand side is $\mathcal{O}(2^{ns})$ when $s>-1/p$. Therefore, we conclude that $\mathcal{B}_{s,p}(\mathbb{R})$ is not complete endowed with merely the spectral norm $\|\cdot\|_{v_s}$. This proves the second statement and completes the proof of Theorem 2.1. □

A direct consequence of the above theorem is

Corollary 2.1. *For $s\in\mathbb{R}$, $\mathcal{B}_{s,1}(\mathbb{R}^d)$ is a commutative Banach algebra under convolution.*

Proof. By Young's inequality, we obtain that for any $f,g\in\mathcal{B}_{s,1}(\mathbb{R}^d)$,

$$\|f*g\|_{L^1(\mathbb{R}^d)}\le\|f\|_{L^1(\mathbb{R}^d)}\|g\|_{L^1(\mathbb{R}^d)}.$$

Moreover,

$$\|f*g\|_{v_s}=\int_{\mathbb{R}^d}\|x\|_1^s|\widehat{f*g}(x)|\mathrm{d}x=\int_{\mathbb{R}^d}\|x\|_1^s|\widehat{f(x)}||\widehat{g(x)}|\mathrm{d}x\le\|f\|_{v_s}\|g\|_{L^1(\mathbb{R}^d)}.$$

A combination of the above two inequalities leads to

$$\|f*g\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)}\le\|f\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)}\|g\|_{L^1(\mathbb{R}^d)}\le\|f\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)}\|g\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)}.$$

This proves the assertion. □

It is natural to discuss the chain of embedding for $\mathcal{B}_{s,p}(\mathbb{R}^d)$ with varying s and p.

Theorem 2.2 (Monotonicity of $\mathcal{B}_{s,p}(\mathbb{R}^d)$)**.** (i) *Let $1\le p\le 2$ and $-d/p<s_1<s_2$. There holds*

$$\mathcal{B}_{s_2,p}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s_1,p}(\mathbb{R}^d). \tag{2.3}$$

(ii) *Let $1\le p\le 2$ and $s\ge 0$. There holds*

$$\mathcal{B}_{s,1}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,2}(\mathbb{R}^d).$$

Remark 2.1. We point that the inclusion (2.3) is proper. Indeed, consider the distance function on $\mathbb{R}$:

$$f(x):=\max(1-|x|,0).$$

Thus, for any $x\in\mathbb{R}$,

$$\widehat{f}(x)=\left(\frac{\sin\pi x}{\pi x}\right)^2.$$

It is clear that

$$\|f\|_{L^p(\mathbb{R})}=(2/(1+p))^{1/p}$$

with $1\le p\le 2$. And for any $-1<s<1$,

$$\begin{aligned}\|f\|_{v_s}&=2\int_0^\infty x^s\frac{\sin^2(\pi x)}{\pi^2x^2}\mathrm{d}x\le 2\int_0^1 x^s\mathrm{d}x+\frac{2}{\pi^2}\int_1^\infty x^{s-2}\mathrm{d}x\\&=\frac{2}{1+s}+\frac{2}{\pi^2(1-s)}<\infty.\end{aligned}$$

This implies that $f\in\mathcal{B}_{s,p}(\mathbb{R})$ with $-1<s<1$, while

$$\begin{aligned}\|f\|_{v_1}&=\frac{2}{\pi^2}\int_0^\infty\frac{\sin^2(\pi x)}{x}\mathrm{d}x\ge\frac{2}{\pi^2}\int_1^\infty\frac{\sin^2(\pi x)}{x}\mathrm{d}x\\&=\frac{1}{\pi^2}\left(\int_1^\infty\frac{1}{x}\mathrm{d}x-\int_1^\infty\frac{\cos(2\pi x)}{x}\mathrm{d}x\right)=\infty.\end{aligned}$$

Hence $f\notin\mathcal{B}_{1,p}(\mathbb{R})$. This means that the inclusion (2.3) is proper.

To prove Theorem 2.2, we need an interpolation inequality that connects the spectrum norms of different orders, which is key to the properties of $\mathcal{B}_{s,p}(\mathbb{R}^d)$. This inequality also motivates our definition for $\mathcal{B}_{s,p}(\mathbb{R}^d)$.

Lemma 2.1. *Let $1\le p\le 2$ and $-d/p<s_1<s_2$. There exists $C(p;s_1,s_2)$ such that for any $f\in\mathcal{B}_{s_2,p}(\mathbb{R}^d)$,*

$$\|f\|_{v_{s_1}}\le C(p;s_1,s_2)\|f\|^{\gamma}_{L^p(\mathbb{R}^d)}\|f\|^{1-\gamma}_{v_{s_2}},\tag{2.4}$$

where $\gamma=(s_2-s_1)/(s_2+d/p)$ and

$$C(p;s_1,s_2):=\frac{1+\gamma}{\gamma^\gamma}\left(\frac{2^d}{(d+s_1p)(d-1)!}\right)^{\gamma/p}.$$

For large d, $C(p;s_1,s_2)\sim(2e/d)^{s_2-s_1}$.

The dependence of $C(p;s_1,s_2)$ on d is sharp by substituting $f(x)=e^{-\pi|x|^2}$ into (2.4). The proof of Lemma 2.1 is based on Fourier modes splitting.

Proof. Let $K>0$ be a constant to be determined later on. Decompose

$$\|f\|_{v_{s_1}}=\int_{\|x\|_1\le K}\|x\|_1^{s_1}|\widehat{f}(x)|\mathrm{d}x+\int_{\|x\|_1>K}\|x\|_1^{s_1}|\widehat{f}(x)|\mathrm{d}x.$$

By symmetry, we obtain that for any $s>-d$,

$$\begin{aligned}\int_{\|x\|_1\le K}\|x\|_1^s\mathrm{d}x&=2^d\int_{x_1+\cdots+x_d\le K,x_i\ge 0}(x_1+\cdots+x_d)^s\mathrm{d}x_1\cdots\mathrm{d}x_d\\&=2^d\int_0^\infty\cdots\int_0^\infty\int_{y_1+\cdots+y_{d-1}}^K t^s\mathrm{d}t\mathrm{d}y_1\cdots\mathrm{d}y_{d-1}\\&=2^d\int_0^K t^s\int_{y_1+\cdots+y_{d-1}\le t,y_i\ge 0}\mathrm{d}y_1\cdots\mathrm{d}y_{d-1}\mathrm{d}t\\&=\frac{2^d}{(d-1)!}\frac{K^{d+s}}{d+s}.\end{aligned}\tag{2.5}$$

It follows from the Hausdorff-Young inequality (2.1) and the above identity that

$$\int_{\|x\|_1\le K}\|x\|_1^{s_1}|\widehat{f}(x)|\,\mathrm{d}x\le \left\|\widehat{f}\right\|_{L^{p'}(\mathbb{R}^d)}\left(\int_{\|x\|_1\le K}\|x\|_1^{s_1p}\,\mathrm{d}x\right)^{1/p}$$
$$\le K^{s_1+d/p}\left(\frac{2^d}{(d+s_1p)(d-1)!}\right)^{1/p}\|f\|_{L^p(\mathbb{R}^d)}.$$

On the other hand, it is clear that

$$\int_{\|x\|_1>K}\|x\|_1^{s_1}|\widehat{f}(x)|\,\mathrm{d}x\le K^{s_1-s_2}\|f\|_{v_{s_2}}.$$

Combining the above two inequalities, optimizing with respect to K, and then taking

$$K=\left(\frac{s_2-s_1}{s_2+d/p}\frac{\|f\|_{v_{s_2}}}{\|f\|_{L^p(\mathbb{R}^d)}}\left(\frac{(d+s_1p)(d-1)!}{2^d}\right)^{1/p}\right)^{1/(s_2+d/p)},$$

we obtain (2.4). □

A direct consequence of Lemma 2.1 is the integrability of $\widehat{f}$ for any function $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ with a nonnegative index s.

Corollary 2.2. *Let $s\ge 0$ and $1\le p\le 2$. Then for any $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$, there holds $\widehat{f}\in L^1(\mathbb{R}^d)$.*

Proof. If $s=0$, then $\widehat{f}\in L^1(\mathbb{R}^d)$ because

$$\left\|\widehat{f}\right\|_{L^1(\mathbb{R}^d)}=\|f\|_{v_s}.$$

If $s>0$, then we take $s_1=0$ and $s_2=s$ in (2.4). Thus,

$$\left\|\widehat{f}\right\|_{L^1(\mathbb{R}^d)}=\|f\|_{v_0}\lesssim\|f\|_{L^p(\mathbb{R}^d)}^{\gamma}\|f\|_{v_s}^{1-\gamma}\lesssim\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)},\tag{2.6}$$

where $\gamma=s/(s+d/p)$. □

We are ready to prove Theorem 2.2.

Proof of Theorem 2.2. (i) By (2.4), we obtain

$$\|f\|_{v_{s_1}}\lesssim\gamma\|f\|_{L^p(\mathbb{R}^d)}+(1-\gamma)\|f\|_{v_{s_2}}.$$

This immediately implies

$$\|f\|_{\mathcal{B}_{s_1,p}(\mathbb{R}^d)}\lesssim\|f\|_{\mathcal{B}_{s_2,p}(\mathbb{R}^d)},$$

and hence (2.3) is true.

(ii) Using Corollary 2.2, we have that for any $f\in\mathcal{B}_{s,1}(\mathbb{R}^d)$ with $s\geq 0$, $\widehat{f}\in L^1(\mathbb{R}^d)$. Therefore, by the Fourier inversion theorem and (2.6) with $p=1$, we obtain

$$\|f\|_{L^\infty(\mathbb{R}^d)}\leq\left\|\widehat{f}\right\|_{L^1(\mathbb{R}^d)}\lesssim\|f\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)}.$$

Thus, for any $1<p<2$,

$$\|f\|_{L^p(\mathbb{R}^d)}\leq\|f\|_{L^1(\mathbb{R}^d)}^{1/p}\|f\|_{L^\infty(\mathbb{R}^d)}^{1-1/p}\lesssim\|f\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)}.$$

This proves that $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ and the embedding $\mathcal{B}_{s,1}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d)$.

Next, invoking Corollary 2.2 again, we conclude that for any $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$, $\widehat{f}\in L^1(\mathbb{R}^d)$. Therefore, using the Fourier inversion theorem again and (2.6) with $1<p<2$, we obtain

$$\|f\|_{L^\infty(\mathbb{R}^d)}\leq\left\|\widehat{f}\right\|_{L^1(\mathbb{R}^d)}\lesssim\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}.$$

Thus,

$$\|f\|_{L^2(\mathbb{R}^d)}^2\leq\|f\|_{L^p(\mathbb{R}^d)}^p\|f\|_{L^\infty(\mathbb{R}^d)}^{2-p}.$$

Combining the above two inequalities, we obtain

$$\|f\|_{L^2(\mathbb{R}^d)}^2\lesssim\|f\|_{L^p(\mathbb{R}^d)}^p\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}^{2-p}\lesssim\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}^2. \tag{2.7}$$

This proves that $f\in\mathcal{B}_{s,2}(\mathbb{R}^d)$ and the embedding $\mathcal{B}_{s,p}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,2}(\mathbb{R}^d)$. □

To deal with the endpoint for $s=-d$ and $p=1$ in Theorem 2.2, we recall the definition of real Hardy space $H^1(\mathbb{R}^d)$.

Definition 2.2. [51] The Hardy space $H^1(\mathbb{R}^d)$ is defined by

$$H^1(\mathbb{R}^d):=\left\{f\text{ is a distribution}\mid\varphi^+(f):=\sup_{t>0}|\varphi_t*f|\in L^1(\mathbb{R}^d)\right\},$$

where φ is Schwartz function with $\int_{\mathbb{R}^d}\varphi(x)dx=1$, and for all $y\in\mathbb{R}^d$ and $t\in(0,\infty)$, $\varphi_t(y):=t^{-d}\varphi(y/t)$. Moreover, define

$$\|f\|_{H^1(\mathbb{R}^d)}:=\|\varphi^+(f)\|_{L^1(\mathbb{R}^d)}.$$

Theorem 2.3. *Let $L^1_0(\mathbb{R}^d)$ be the subspace of $L^1(\mathbb{R}^d)$ with zero mean. There holds*

$$H^1(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{-d,1}(\mathbb{R}^d)\hookrightarrow L^1_0(\mathbb{R}^d). \tag{2.8}$$

To prove Theorem 2.3, we need the following Hardy type inequality due to BOURGAIN [5,6], and we refer to [22, Chapter 3, Corollary 7.23 with $p=1$].

Lemma 2.2. *There exists a positive constant C such that for any $f \in H^1(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}^d} \frac{|\widehat{f}(x)|}{|x|^d} \mathrm{d}x \le C \|f\|_{H^1(\mathbb{R}^d)}. \tag{2.9}$$

Proof of Theorem 2.3. We rewrite (2.9) as

$$\|f\|_{v_{-d}} \lesssim \|f\|_{H^1(\mathbb{R}^d)},$$

which together with

$$\|f\|_{L^1(\mathbb{R}^d)} \le \|f\|_{H^1(\mathbb{R}^d)}$$

yields that $H^1(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{-d,1}(\mathbb{R}^d)$.

On the other hand, for any $f \in \mathcal{B}_{-d,1}(\mathbb{R}^d)$, $f \in L^1(\mathbb{R}^d)$, and then $\widehat{f}$ is uniformly continuous. By $\|f\|_{v_{-d}} < \infty$, we conclude that $\widehat{f}(0) = 0$. This implies

$$\int_{\mathbb{R}^d} f(x) \mathrm{d}x = 0.$$

Hence $f \in L^1_0(\mathbb{R}^d)$. This proves the second embedding of (2.8) and completes the proof. □

3 Sufficient conditions for functions in Barron spectrum spaces

In this part, we give sufficient conditions for functions in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ with suitable indices. We start with the characterization in one dimension. To this end, we firstly introduce the L^p-modulus of continuity: for any $f \in L^p(\mathbb{R})$ with $1 \le p \le \infty$ and $t \in \mathbb{R}$, define

$$\omega_p(f;t) := \|f(\cdot + t) - f(\cdot)\|_{L^p(\mathbb{R})}.$$

It is clear that $\omega_p(f;t) \to 0$ as $t \to 0$ for any $f \in L^p(\mathbb{R})$.

Theorem 3.1. *Suppose that $f \in L^p(\mathbb{R})$ with $1 \le p \le 2$ satisfies $\omega_p(f;t) \le C|t|^\alpha$ with $0 < \alpha \le 1$, then $f \in \mathcal{B}_{s,p}(\mathbb{R})$ with $-1/p < s < \alpha - 1/p$.*

The following example shows that Theorem 3.1 fails when $s = \alpha - 1/p$.

Example 3.1. Let $1 < p \le 2$ and $-1/p < \beta < 0$, and let f be an even function with

$$f(x) = \frac{1}{x + x^{-\beta}}, \qquad x > 0.$$

A direct calculation gives

$$\begin{aligned}\int_{\mathbb{R}}|f(x)|^p\,\mathrm{d}x&=2\int_0^\infty (x+x^{-\beta})^{-p}\,\mathrm{d}x\\&=2\int_0^1 (x+x^{-\beta})^{-p}\,\mathrm{d}x+2\int_1^\infty (x+x^{-\beta})^{-p}\,\mathrm{d}x\\&\le 2\int_0^1 x^{p\beta}\,\mathrm{d}x+2\int_1^\infty x^{-p}\,\mathrm{d}x\\&=\frac{2}{1+p\beta}+\frac{2}{p-1}.\end{aligned}$$

This shows $f\in L^p(\mathbb{R})$.

Without loss of generality, we may assume $0<t<1$. For any $x>t$, there exists a constant $0<\theta<1$ such that

$$|f(x+t)-f(x)|=t|f'(x+\theta t)|\le t|f'(x)|,$$

because

$$|f'(x)|=\frac{1-\beta x^{-\beta-1}}{(x+x^{-\beta})^2}$$

is positive and decreasing for $x>0$. This leads to

$$\begin{aligned}\int_t^\infty |f(x+t)-f(x)|^p\,\mathrm{d}x&\le t^p\int_t^\infty |f'(x)|^p\,\mathrm{d}x\\&\le (1-\beta)^p t^p\left(\int_t^1 x^{-(1-\beta)p}\,\mathrm{d}x+\int_1^\infty x^{-2p}\,\mathrm{d}x\right)\\&\le (1-\beta)^p\Big[\frac{t^{1+p\beta}}{(1-\beta)p-1}+\frac{t^p}{2p-1}\Big]\\&\le (1-\beta)^p\Big[\frac{1}{(1-\beta)p-1}+\frac{1}{2p-1}\Big]t^{1+p\beta}.\end{aligned}$$

On the other hand, a direct calculation gives us that for any $x>0$,

$$|f(x+t)-f(x)|=f(x)-f(x+t)\le x^\beta.$$

Hence,

$$\int_0^t |f(x+t)-f(x)|^p\,\mathrm{d}x\le\frac{t^{1+p\beta}}{1+p\beta}.$$

Choose $\alpha=\beta+1/p$. Then, for any $0<t<1$,

$$\omega_p(f;t)\le C|t|^\alpha.$$

By Theorem 3.1, we conclude that $f\in\mathcal{B}_{s,p}(\mathbb{R})$ with $-1/p<s<\beta$.

Using [56, Theorem 126], we obtain the following growth property for $\widehat{f}$:

$$|\widehat{f}(x)| \sim |x|^{-\beta-1} \qquad \text{for} \quad |x| \to \infty.$$

Hence,

$$\|f\|_{v_\beta} \gtrsim \int_1^\infty |x|^{-1} \mathrm{d}x = \infty.$$

We conclude that $f \notin \mathcal{B}_{\beta,p}(\mathbb{R})$. This means the range of the index s in Theorem 3.1 cannot be extended to $s = \alpha - 1/p$.

Proof of Theorem 3.1. By the Hausdorff-Young inequality (2.1), we obtain that for any $k \in \mathbb{N} \cup \{0\}$,

$$\begin{aligned}
&\int_{2^k \le |x| < 2^{k+1}} |x|^s |\widehat{f}(x)| \mathrm{d}x \\
&\le \left(\int_{2^k \le |x| < 2^{k+1}} |x|^{sp} \mathrm{d}x \right)^{1/p} \left(\int_{2^k \le |x| < 2^{k+1}} |\widehat{f}(x)|^{p'} \mathrm{d}x \right)^{1/p'} \\
&\lesssim 2^{k(s+1/p)} \left(\int_{2^k \le |x| < 2^{k+1}} |e^{2\pi i x t} - 1|^{p'} |\widehat{f}(x)|^{p'} \mathrm{d}x \right)^{1/p'} \\
&\lesssim 2^{k(s+1/p)} \left\| \widehat{f(\cdot + t) - f(\cdot)} \right\|_{L^{p'}(\mathbb{R}^d)} \\
&\lesssim 2^{k(s+1/p)} \| f(\cdot + t) - f(\cdot) \|_{L^p(\mathbb{R}^d)} \\
&\lesssim 2^{k(s+1/p-\alpha)},
\end{aligned}$$

where we have chosen $t = 2^{-k-2}$ and have used the fact that for any $|x| \in [2^k, 2^{k+1})$,

$$|e^{2\pi i x t} - 1|^2 = 4\sin^2(\pi x t) \ge 16 x^2 t^2 \ge 1.$$

Summing up for all k, we obtain that for $-1/p < s < \alpha - 1/p$,

$$\begin{aligned}
\|f\|_{v_s} &= \int_{-1}^1 |x|^s |\widehat{f}(x)| \mathrm{d}x + \sum_{k=0}^\infty \int_{2^k \le |x| < 2^{k+1}} |x|^s |\widehat{f}(x)| \mathrm{d}x \\
&\le 2 \left(\int_0^1 x^{sp} \mathrm{d}x \right)^{1/p} \left\| \widehat{f} \right\|_{L^{p'}(\mathbb{R}^d)} + \sum_{k=0}^\infty 2^{-k(\alpha - 1/p - s)} \\
&\le \frac{2}{(1+sp)^{1/p}} \|f\|_{L^p(\mathbb{R}^d)} + \sum_{k=0}^\infty 2^{-k(\alpha - 1/p - s)} \\
&< \infty.
\end{aligned}$$

This implies that $f \in \mathcal{B}_{s,p}(\mathbb{R})$. □

The high dimensional analog of Theorem 3.1 also holds true provided that L^p-modulus is replaced by a more general average operator. Let μ be a finite Borel measure with unit total mass. For any $t\in\mathbb{R}$, we define the average operator: for any $f\in L^p(\mathbb{R}^d)$ and for any $x\in\mathbb{R}^d$,

$$\mathcal{M}_\mu^t f(x):=\int_{\mathbb{R}^d} f(x+ty)\mathrm{d}\mu(y);$$

see [7]. For $\sigma>0$, let $\mathcal{K}_\sigma$ be the set of all μ whose Fourier transform satisfying

$$|1-\widehat{\mu}(\xi)|\sim\min(1,|\xi|^{2\sigma})\qquad\text{for any }\xi\in\mathbb{R}^d. \tag{3.1}$$

Theorem 3.2. *Let $1\le p\le 2$, $d\ge 2$ and $\sigma>d/(2p)$. Suppose that $\mu\in\mathcal{K}_\sigma$ and $f\in L^p(\mathbb{R}^d)$ satisfying*

$$\left\|\mathcal{M}_\mu^t f-f\right\|_{L^p(\mathbb{R}^d)}\le C|t|^\alpha$$

for some $C>0$ and $0<\alpha\le 2\sigma$. Then $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ with $-d/p<s<\alpha-d/p$.

The index σ specifies the possible range of the Lipschitz order α, which limits the range on the upper bound of s. Before proving the above result, we adapt some examples from [7, §2] to show that $\mathcal{K}_\sigma$ may be realized.

Example 3.2. Let $\mathrm{d}\mu(x)=\mathrm{d}\sigma(x)/\omega_{d-1}$, where $\mathrm{d}\sigma(x)$ is the usual surface measure on the unit sphere $\mathbb{S}^{d-1}$. For any $t\in\mathbb{R}$, the average operator $\mathcal{M}_\mu^t$ is the spherical mean $\mathcal{M}^t$ defined for any $x\in\mathbb{R}^d$,

$$\mathcal{M}^t f(x):=\frac{1}{\omega_{d-1}}\int_{\mathbb{S}^{d-1}} f(x+ty)\mathrm{d}\sigma(y).$$

In this case, for any $\xi\in\mathbb{R}^d$,

$$\widehat{\mu}(\xi)=j_{d/2-1}(|\xi|)$$

with j_ν the spherical Bessel function given by

$$j_\nu(x)=2^\nu\Gamma(\nu+1)x^{-\nu}J_\nu(x),\quad x\in\mathbb{R}, \tag{3.2}$$

where J_ν with $\nu>-1$ is the standard Bessel function of the first kind defined by

$$J_\nu(x)=\frac{(x/2)^\nu}{\Gamma(\nu+1/2)\Gamma(1/2)}\int_0^\pi e^{ix\cos\theta}\sin^{2\nu}\theta\mathrm{d}\theta. \tag{3.3}$$

By [8, Corollary 1.4], we have that for any $\xi\in\mathbb{R}^d$,

$$1-j_{d/2-1}(|\xi|)\sim\min(1,|\xi|^2).$$

Hence $\mu\in\mathcal{K}_1$.

The next example concerns how to achieve a bigger index σ via a combination of the spherical mean.

Example 3.3. Define the average operator: For any $x\in\mathbb{R}^d$,

$$V_k^t f(x) := 2\binom{2k}{k}^{-1}\sum_{l=1}^{k}(-1)^{l+1}\binom{2k}{k-l}\mathcal{M}^{lt}f(x).$$

The corresponding measure is

$$\mu = 2\binom{2k}{k}^{-1}\sum_{l=1}^{k}(-1)^{l+1}\binom{2k}{k-l}\omega^{lt},$$

where $\omega=\mathrm{d}\sigma(x)/\omega_{d-1}$ is the normalized surface measure on the unit sphere defined in the above example, and ω^k is its dilation by k. By [7, Example 2.9], we obtain that for any $\xi\in\mathbb{R}^d$,

$$1-\widehat{\mu}(|\xi|)\sim\min(1,|\xi|^{2k}).$$

Hence, $\mu\in\mathcal{K}_k$.

Proof of Theorem 3.2. We distinguish $p=1$ and $1<p\le 2$.

Let $1<p\le 2$ firstly. Noting that for any $\xi\in\mathbb{R}^d$,

$$(\widehat{\mathcal{M}_\mu^t f - f})(\xi) = (\widehat{\mu}(t\xi)-1)\widehat{f}(\xi),$$

and using the Hausdorff-Young inequality, we have

$$\left(\int_{\mathbb{R}^d}|1-\widehat{\mu}(t\xi)|^{p'}|\widehat{f}(\xi)|^{p'}\mathrm{d}\xi\right)^{1/p'} \le \left\|\mathcal{M}_\mu^t f - f\right\|_{L^p(\mathbb{R}^d)},$$

which, together with the estimate (3.1) yields

$$\left(\int_{|\xi|>1/t}|\widehat{f}(\xi)|^{p'}\mathrm{d}\xi\right)^{1/p'} \le \left\|\mathcal{M}_\mu^t f - f\right\|_{L^p(\mathbb{R}^d)}. \tag{3.4}$$

Using Hölder's inequality, we obtain that for any $k\in\mathbb{N}\cup\{0\}$,

$$\begin{aligned}&\int_{2^k\le|x|<2^{k+1}}|x|^s|\widehat{f}(x)|\,\mathrm{d}x\\ &\le\left(\int_{2^k\le|x|<2^{k+1}}|x|^{sp}\mathrm{d}x\right)^{1/p}\left(\int_{2^k\le|x|<2^{k+1}}|\widehat{f}(x)|^{p'}\mathrm{d}x\right)^{1/p'}.\end{aligned}$$

By (2.5) and the fact $d+sp>0$, we have

$$\int_{2^k\le|x|<2^{k+1}}|x|^{sp}\mathrm{d}x \le \int_{|x|<2^{k+1}}|x|^{sp}\mathrm{d}x \le \frac{\omega_{d-1}}{d+sp}2^{(k+1)(d+sp)}.$$

It follows from (3.4) that

$$\int_{2^k\le|x|<2^{k+1}}|\widehat{f}(x)|^{p'}\,\mathrm{d}x\le\left\|\mathcal{M}^{2^{-k}}f-f\right\|_{L^p(\mathbb{R}^d)}^{p'}\le C^{p'}2^{-p'k\alpha}.$$

Using Hölder's inequality, the Hausdorff-Young inequality, and the fact that $-d/p<s<\alpha-d/p$, we get

$$\begin{aligned}\|f\|_{v_s}&\le\max(1,d^{s/2})\left(\int_{|x|\le1}|x|^s|\widehat{f}(x)|\,\mathrm{d}x+\sum_{k=0}^{\infty}\int_{2^k\le|x|<2^{k+1}}|x|^s|\widehat{f}(x)|\,\mathrm{d}x\right)\\&\le\max(1,d^{s/2})\left(\left(\int_{|x|\le1}|x|^{sp}\,\mathrm{d}x\right)^{1/p}\left\|\widehat{f}\right\|_{L^{p'}(\mathbb{R}^d)}+C\sum_{k=0}^{\infty}2^{-k(\alpha-d/p-s)}\right)\\&\le\max(1,d^{s/2})\left(\frac{\omega_{d-1}}{d+sp}\right)^{1/p}\left(\|f\|_{L^p(\mathbb{R}^d)}+C2^{s+d/p}\sum_{k=0}^{\infty}2^{-k(\alpha-d/p-s)}\right)\\&<\infty.\end{aligned}$$

This proves $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$.

The proof for $p=1$ is exactly the same provided that we replace (3.4) by

$$\sup_{|\xi|>1/t}|\widehat{f}(\xi)|\le\left\|\mathcal{M}_\mu^t f-f\right\|_{L^1(\mathbb{R}^d)},$$

because the Hausdorff-Young inequality still holds true for $p=1$. □

Before closing this section, we extend Example 3.1 to high dimension to show that the index α in Theorem 3.2 cannot be extended to $\alpha-d/p$. This example is taken from [7].

Example 3.4. Let $1<p\le2$ and $-d/p<\beta<0$. Define

$$f(x):=\frac{1}{|x|^d+|x|^{-\beta}},\qquad x\in\mathbb{R}^d\backslash\{0\}.$$

Let $\mathcal{M}_\mu^t$ be the spherical mean as in Example 3.2. By [7, eq. 3.5], we see that

$$\left\|\mathcal{M}_\mu^t f-f\right\|_{L^p(\mathbb{R}^d)}\lesssim|t|^\alpha$$

with $\alpha=\beta+d/p$.

By Theorem 3.2, we conclude that $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ for any $-d/p<s<\beta$. A direct calculation gives us that for any $\xi\in\mathbb{R}^d$,

$$\widehat{f}(\xi)=\omega_{d-1}\int_0^\infty(r^d+r^{-s})^{-1}j_{d/2-1}(2\pi r|\xi|)r^{d-1}\,\mathrm{d}r.$$

Proceeding along the same line that leads to [56, Theorem 126], we have that for any $\xi\in\mathbb{R}^d\backslash\{0\}$:

$$|\widehat{f}(\xi)|\sim|\xi|^{-\beta-d}.$$

This implies

$$\|f\|_{v_\beta}\geq\int_{|\xi|\geq 1}|\xi|^\beta|\widehat{f}(\xi)|\mathrm{d}\xi\gtrsim\int_1^\infty r^{-1}\mathrm{d}r=\infty.$$

Hence $f\notin\mathcal{B}_{\beta,p}(\mathbb{R}^d)$. This means the range of the index α in Theorem 3.2 cannot be extended to $s=\alpha-d/p$.

4 The relationship among the Barron spectrum space, the Bessel potential space, the Besov space and the Sobolev space

We start with the embedding between the Bessel potential space and the Barron spectrum space. Let $\alpha\geq 0$ and G_α be the Bessel potential of order α, i.e., for any $x\in\mathbb{R}^d$,

$$G_\alpha(x)=\frac{1}{(4\pi)^{\alpha/2}\Gamma(\alpha/2)}\int_0^\infty t^{\frac{\alpha-d}{2}}e^{-\frac{\pi|x|^2}{t}-\frac{t}{4\pi}}\frac{\mathrm{d}t}{t}. \tag{4.1}$$

Define the operator $\mathscr{S}_\alpha$: for any $g\in L^p(\mathbb{R}^d)$ with $1\leq p\leq\infty$ and for any $x\in\mathbb{R}^d$,

$$\mathscr{S}_\alpha(g)(x):=G_\alpha*g(x)$$

for $\alpha>0$ and $\mathscr{S}_0(g):=g$ for $\alpha=0$.

Definition 4.1. [50] Let $\alpha\geq 0$ and $1\leq p\leq\infty$. The Bessel potential space $\mathscr{L}_\alpha^p(\mathbb{R}^d)$ is the set of the function $f\in L^p(\mathbb{R}^d)$ that can be written as $f=\mathscr{S}_\alpha(g),g\in L^p(\mathbb{R}^d)$. Moreover, the norm of f is defined by

$$\|f\|_{\mathscr{L}_\alpha^p(\mathbb{R}^d)}=\|g\|_{L^p(\mathbb{R}^d)}.$$

Theorem 4.1. *Let $1\leq p\leq 2$ and $\alpha>s+d/p>0$. There holds*

$$\mathscr{L}_\alpha^p(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d). \tag{4.2}$$

The embedding (4.2) fails when $\alpha=s+d/p$, because for $1\leq p\leq 2$, $s=-d/p$ and $\alpha=0$, there holds

$$\mathcal{B}_{-d/p,p}(\mathbb{R}^d)\hookrightarrow L^p(\mathbb{R}^d)=\mathscr{L}_0^p(\mathbb{R}^d).$$

We refer to Remark 4.1 for a further discussion about $\alpha=s+d/p\neq 0$.

Proof of Theorem 4.1. For any $f\in\mathscr{L}_\alpha^p(\mathbb{R}^d)$, write $f=\mathscr{S}_\alpha(g)$ with $g\in L^p(\mathbb{R}^d)$. Recalling the fact that for any $x\in\mathbb{R}^d$, $\widehat{G}_\alpha(x)=(1+4\pi^2|x|^2)^{-\alpha/2}$, we obtain

$$\|f\|_{v_s}=\int_{\mathbb{R}^d}\|x\|_1^s|\widehat{G_\alpha(x)}||\widehat{g}(x)|\mathrm{d}x\leq\max(1,d^{s/2})\int_{\mathbb{R}^d}|x|^s(1+4\pi^2|x|^2)^{-\alpha/2}|\widehat{g}(x)|\mathrm{d}x.$$

A straightforward calculation gives

$$\int_{\mathbb{R}^d} \frac{|x|^{sp}}{(1+4\pi^2|x|^2)^{p\alpha/2}} \mathrm{d}x = \omega_{d-1} \int_0^\infty \frac{r^{sp+d-1}}{(1+4\pi^2 r^2)^{p\alpha/2}} \mathrm{d}r$$
$$= \frac{\omega_{d-1}}{(2\pi)^{d+sp}} \int_0^\infty \frac{t^{d-1+sp}}{(1+t^2)^{p\alpha/2}} \mathrm{d}t.$$

Let $t = \tan\theta$. Rewrite the above integral as

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{|x|^{sp}}{(1+4\pi^2|x|^2)^{p\alpha/2}} \mathrm{d}x &= \frac{\omega_{d-1}}{(2\pi)^{d+sp}} \int_0^{\pi/2} (\sin\theta)^{d-1+sp} (\cos\theta)^{(\alpha-s)p-1-d} \mathrm{d}\theta \\ &= \frac{\omega_{d-1}}{2(2\pi)^{d+sp}} B\left(\frac{d+sp}{2}, \frac{(\alpha-s)p-d}{2}\right),\end{aligned} \tag{4.3}$$

where for any $z,w>0$, $B(z,w)$ is the beta function defined by

$$B(z,w) := \int_0^1 t^{z-1}(1-t)^{w-1} \mathrm{d}t.$$

By Hölder's inequality, the Hausdorff-Young inequality (2.1) and (4.3), we obtain

$$\begin{aligned}\|f\|_{v_s} &\le \max(1,d^{s/2}) \left\| |\cdot|^s (1+4\pi^2|\cdot|^2)^{-\alpha/2} \right\|_{L^p(\mathbb{R}^d)} \|\widehat{g}\|_{L^{p'}(\mathbb{R}^d)} \\ &\le \max(1,d^{s/2}) \left\| |\cdot|^s (1+4\pi^2|\cdot|^2)^{-\alpha/2} \right\|_{L^p(\mathbb{R}^d)} \|g\|_{L^p(\mathbb{R}^d)} \\ &= C_s \|f\|_{\mathscr{L}_\alpha^p(\mathbb{R}^d)}\end{aligned}$$

with

$$C_s = \frac{\max(1,d^{s/2})}{(2\pi)^{s+d/p}} \left(\frac{\omega_{d-1}}{2}\right)^{1/p} B^{1/p}\left(\frac{d+sp}{2}, \frac{(\alpha-s)p-d}{2}\right).$$

The constant C_s is finite provided that $\alpha > s+d/p > 0$.

Next, via Young's inequality and $f(x) = \mathscr{S}_\alpha(g)(x)$, we have

$$\|f\|_{L^p(\mathbb{R}^d)} \le \|G_\alpha\|_{L^1(\mathbb{R}^d)} \|g\|_{L^p(\mathbb{R}^d)} = \|g\|_{L^p(\mathbb{R}^d)} = \|f\|_{\mathscr{L}_\alpha^p(\mathbb{R}^d)}.$$

This gives (4.2). □

Now we study the embedding between the Barron spectrum space and the Besov space, which is defined by

Definition 4.2. [57] Let $\Phi(\mathbb{R}^d)$ be the set of all systems $\phi = \{\phi_j\}_{j=0}^\infty \subset \mathscr{S}(\mathbb{R}^d)$ such that

$$\begin{cases} \text{supp } \phi_0 \subset \Gamma_0 := \left\{ x \in \mathbb{R}^d \mid |x| < 2 \right\}, \\ \text{supp } \phi_j \subset \Gamma_j := \left\{ x \in \mathbb{R}^d \mid 2^{j-1} \le |x| \le 2^{j+1} \right\}, \qquad j = 1,2,3,\cdots. \end{cases}$$

For every multi-index α there exists a positive number c_α such that for all $j=0,1,2,\cdots$, and for all $x\in\mathbb{R}^d$,

$$2^{j|\alpha|}|\nabla^\alpha\phi_j(x)|\le c_\alpha,$$

and for every $x\in\mathbb{R}^d$, $\sum_{j=0}^\infty\phi_j(x)=1$.

Let $s\in\mathbb{R}$ and $1\le p,q\le\infty$. Let $\phi=\{\phi_j\}_{j=0}^\infty\in\Phi(\mathbb{R}^d)$. Define

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

with

$$\|f\|_{B^s_{p,q}(\mathbb{R}^d)}:=\left(\sum_{j=0}^\infty 2^{jsq}\left\|\left(\phi_j\widehat{f}\right)^\vee\right\|^q_{L^p(\mathbb{R}^d)}\right)^{1/q}.$$

Theorem 4.2. (i) *For $1\le p\le 2$, and $s>-d/p$, there holds*

$$B^{s+d/p}_{p,1}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d). \tag{4.4}$$

(ii) *If $1\le p\le 2$ and $\alpha>s+d/p>0$, then for all $1\le q\le\infty$,*

$$B^\alpha_{p,q}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d). \tag{4.5}$$

(iii) *For $1\le p\le 2$ and $s\in\mathbb{R}$, there holds*

$$\mathcal{B}_{s,p}(\mathbb{R}^d)\hookrightarrow B^s_{\infty,1}(\mathbb{R}^d). \tag{4.6}$$

Taking $s=0$ and $p=1$ in (4.4), we find that the Bump algebra $B^d_{1,1}(\mathbb{R}^d)$ [40] embeds into $\mathcal{B}_{0,1}(\mathbb{R}^d)$. The Bump algebra has been used as the target function space in [17] to measure the degree of approximation for the neural network.

Proof. (i) For any $s>-d/p$ and for any $1\le p\le 2$, It follows from the facts [57, Proposition 2 §2.3.2 and Proposition, §2.5.7]

$$B^{s+d/p}_{p,1}(\mathbb{R}^d)\hookrightarrow B^0_{p,1}(\mathbb{R}^d)\hookrightarrow L^p(\mathbb{R}^d).$$

This gives that

$$\|f\|_{L^p(\mathbb{R}^d)}\lesssim\|f\|_{B^{s+d/p}_{p,1}(\mathbb{R}^d)}. \tag{4.7}$$

It remains to bound $\|f\|_{v_s}$.

By definition,

$$\text{supp }\phi_j\cap\Gamma_k\neq\emptyset$$

implies that $|j-k|\le 1$, $j,k=0,1,\cdots$.

A direct calculation gives that for any $j=0,1,\cdots$,

$$\int_{\Gamma_j}\|x\|_1^{sp}\,\mathrm{d}x\le 2^{(j+1)(d+sp)}\frac{\max(1,d^{sp/2})\omega_{d-1}}{d+sp}.$$

By the Hausdorff-Young inequality, we get

$$\begin{aligned}\int_{\Gamma_j}\|x\|_1^s|\widehat{f}(x)|\,\mathrm{d}x&=\int_{\Gamma_j}\sum_{k=0}^{\infty}\|x\|_1^s\phi_k(x)|\widehat{f}(x)|\,\mathrm{d}x\\&=\sum_{k=j-1}^{j+1}\int_{\Gamma_j}\|x\|_1^s\phi_k(x)|\widehat{f}(x)|\,\mathrm{d}x\\&\le\sum_{k=j-1}^{j+1}\int_{\Gamma_j}\|x\|_1^s|\phi_k(x)\widehat{f}(x)|\,\mathrm{d}x\\&\le\sum_{k=j-1}^{j+1}\left(\int_{\Gamma_j}\|x\|_1^{sp}\,\mathrm{d}x\right)^{1/p}\left(\int_{\Gamma_j}|\phi_k(x)\widehat{f}(x)|^{p'}\,\mathrm{d}x\right)^{1/p'}\\&\le\left(\frac{2^{d+sp}\max(1,d^{sp/2})\omega_{d-1}}{d+sp}\right)^{1/p}2^{j(s+d/p)}\sum_{k=j-1}^{j+1}\left\|(\phi_k\widehat{f})^\vee\right\|_{L^p(\mathbb{R}^d)}.\end{aligned}$$

Therefore, summing over j, we obtain

$$\begin{aligned}\|f\|_{v_s}&\le\sum_{j=0}^{\infty}\int_{\Gamma_j}\|x\|_1^s|\widehat{f}(x)|\,\mathrm{d}x\\&\le 3\left(\frac{2^{d+sp}\max(1,d^{sp/2})\omega_{d-1}}{d+sp}\right)^{1/p}\sum_{j=0}^{\infty}2^{j(s+d/p)}\left\|(\phi_j\widehat{f})^\vee\right\|_{L^p(\mathbb{R}^d)}\\&\lesssim\|f\|_{B^{s+d/p}_{p,1}(\mathbb{R}^d)},\end{aligned}\tag{4.8}$$

which together with (4.7) leads to

$$\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}\lesssim\|f\|_{B^{s+d/p}_{p,1}(\mathbb{R}^d)}.$$

This gives (4.4).

(ii) For any $\alpha>s+d/p>0$, $1\le p\le 2$ and $1\le q\le\infty$, we rewrite (4.8) as

$$\begin{aligned}\|f\|_{v_s}&\lesssim\sum_{j=0}^{\infty}2^{j(s+d/p)}\left\|(\phi_j\widehat{f})^\vee\right\|_{L^p(\mathbb{R}^d)}\\&\le\left(\sum_{j=0}^{\infty}2^{qj\alpha}\left\|(\phi_j\widehat{f})^\vee\right\|^q_{L^p(\mathbb{R}^d)}\right)^{1/q}\left(\sum_{j=0}^{\infty}2^{q'j(s+d/p-\alpha)}\right)^{1/q'}\\&\lesssim\|f\|_{B^{\alpha}_{p,q}(\mathbb{R}^d)}.\end{aligned}$$

Moreover,

$$B^{\alpha}_{p,q}(\mathbb{R}^d) \hookrightarrow B^0_{p,1}(\mathbb{R}^d) \hookrightarrow L^p(\mathbb{R}^d).$$

This gives

$$\|f\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{B^{\alpha}_{p,q}(\mathbb{R}^d)}.$$

A combination of the above two inequalities leads to (4.5).

(iii) For any $f \in \mathcal{B}_{s,p}$ with $1 \le p \le 2$, by Corollary 2.2, there holds

$$\begin{aligned}
\|f\|_{B^s_{\infty,1}(\mathbb{R}^d)} &= \sum_{j=0}^{\infty} 2^{js} \left\| \left(\phi_j \widehat{f}\right)^{\vee} \right\|_{L^\infty(\mathbb{R}^d)} \le \sum_{j=0}^{\infty} 2^{js} \left\|\phi_j \widehat{f}\right\|_{L^1(\mathbb{R}^d)} \\
&\le \left\|\phi_0 \widehat{f}\right\|_{L^1(\mathbb{R}^d)} + \sum_{j=1}^{\infty} 2^{js} \left\|\phi_j \widehat{f}\right\|_{L^1(\mathbb{R}^d)} \\
&\lesssim \left\|\widehat{f}\right\|_{L^1(\mathbb{R}^d)} + \sum_{j=0}^{\infty} \int_{\Gamma_j} \|x\|_1^s |\widehat{f}(x)| \mathrm{d}x \\
&\lesssim \|f\|_{\mathcal{B}_{s,p}} + \int_{\mathbb{R}^d} \|x\|_1^s |\widehat{f}(x)| \mathrm{d}x \lesssim \|f\|_{\mathcal{B}_{s,p}}.
\end{aligned}$$

This gives (4.6). □

The endpoint $s=-d/p$ in Theorem 4.2 is subtle. We have the following result.

Corollary 4.1. (i) *There holds*

$$B^0_{1,1}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{-d,1}(\mathbb{R}^d). \tag{4.9}$$

(ii) *For* $1 \le r < p \le 2$, *there holds*

$$B^0_{p,1}(\mathbb{R}^d) \cap L^r(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{-d/p,p}(\mathbb{R}^d). \tag{4.10}$$

Proof. (i) The first embedding (4.9) is a combination of the following well-known fact [57, Remark 2 in §2.5.8]

$$B^0_{1,1}(\mathbb{R}^d) \hookrightarrow \dot{F}^0_{1,2}(\mathbb{R}^d) = H^1(\mathbb{R}^d)$$

and the embedding $H^1(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{-d,1}(\mathbb{R}^d)$ proved in Theorem 2.3.

(ii) To prove (4.10), we firstly claim

$$\int_{\mathbb{R}^d} (1+|x|)^{-d/p} |\widehat{f}(x)| \mathrm{d}x \le 3 \left(\frac{2^d \omega_{d-1}}{d}\right)^{1/p} \|f\|_{B^0_{p,1}(\mathbb{R}^d)}. \tag{4.11}$$

Given the above inequality, using the Hausdorff-Young inequality, we have

$$\begin{aligned}
\|f\|_{v_{-d/p}} &\le \int_{|x|\le 1} |x|^{-d/p} |\widehat{f}(x)| \mathrm{d}x + \int_{|x|>1} |x|^{-d/p} |\widehat{f}(x)| \mathrm{d}x \\
&\le \left(\int_{|x|\le 1} |x|^{-dr/p} \mathrm{d}x \right)^{1/r} \left\|\widehat{f}\right\|_{L^{r'}(\mathbb{R}^d)} + 2^{d/p} \int_{\mathbb{R}^d} (1+|x|)^{-d/p} |\widehat{f}(x)| \mathrm{d}x \\
&\le \left(\frac{p\omega_{d-1}}{d(p-r)} \right)^{1/r} \|f\|_{L^r(\mathbb{R}^d)} + 3 \left(\frac{4^d \omega_{d-1}}{d} \right)^{1/p} \|f\|_{B^0_{p,1}(\mathbb{R}^d)},
\end{aligned}$$

which together with (4.7) gives (4.10).

It remains to prove (4.11). We follow the line that leads to (4.8). A direct calculation gives

$$\int_{\Gamma_0}(1+|x|)^{-d}\mathrm{d}x \le \omega_{d-1}\int_0^2 r^{d-1}\mathrm{d}r = \frac{\omega_{d-1}2^d}{d},$$

and for $j\in\mathbb{N}$, there holds

$$\int_{\Gamma_j}(1+|x|)^{-d}\mathrm{d}x = \omega_{d-1}\int_{2^{j-1}}^{2^{j+1}}\frac{r^{d-1}}{(1+r)^d}\mathrm{d}r \le \omega_{d-1}\int_{2^{j-1}}^{2^{j+1}} r^{-1}\mathrm{d}r = 2\omega_{d-1}\ln 2.$$

For $j=0,1,\cdots$, using the Hausdorff-Young inequality, we obtain

$$\begin{aligned}
\int_{\Gamma_j}(1+|x|)^{-d/p}|\widehat{f}(x)|\mathrm{d}x &= \sum_{k=j-1}^{j+1}\int_{\Gamma_j}(1+|x|)^{-d/p}\phi_k(x)|\widehat{f}(x)|\mathrm{d}x\\
&\le \sum_{k=j-1}^{j+1}\int_{\Gamma_j}(1+|x|)^{-d/p}|\phi_k(x)\widehat{f}(x)|\mathrm{d}x\\
&\le \sum_{k=j-1}^{j+1}\left(\int_{\Gamma_j}(1+|x|)^{-d}\mathrm{d}x\right)^{1/p}\left(\int_{\Gamma_j}|\phi_k(x)\widehat{f}(x)|^{p'}\mathrm{d}x\right)^{1/p'}\\
&\le \left(\frac{2^d\omega_{d-1}}{d}\right)^{1/p}\sum_{k=j-1}^{j+1}\left\|(\phi_k\widehat{f})^{\vee}\right\|_{L^p(\mathbb{R}^d)}.
\end{aligned}$$

Therefore, summing over j, we obtain

$$\begin{aligned}
\int_{\mathbb{R}^d}(1+|x|)^{-d/p}|\widehat{f}(x)|\mathrm{d}x &\le \sum_{j=0}^{\infty}\int_{\Gamma_j}(1+|x|)^{-d/p}|\widehat{f}(x)|\mathrm{d}x\\
&\le 3\left(\frac{2^d\omega_{d-1}}{d}\right)^{1/p}\sum_{j=0}^{\infty}\left\|(\phi_j\widehat{f})^{\vee}\right\|_{L^p(\mathbb{R}^d)}\\
&\le 3\left(\frac{2^d\omega_{d-1}}{d}\right)^{1/p}\|f\|_{B^0_{p,1}(\mathbb{R}^d)}.
\end{aligned}$$

This yields (4.11) and finishes the proof. □

A combination of Theorem 4.2, Corollary 4.1 and embedding (4.6) yields the following chains of embedding.

Corollary 4.2. *For $1\le p\le 2$ and $s>-d/p$,*

$$B^{s+d/p}_{p,1}(\mathbb{R}^d)\hookrightarrow \mathcal{B}_{s,p}(\mathbb{R}^d)\hookrightarrow B^s_{\infty,1}(\mathbb{R}^d),$$

and

$$B^0_{1,1}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{-d,1}(\mathbb{R}^d)\hookrightarrow B^{-d}_{\infty,1}(\mathbb{R}^d),$$
$$B^0_{p,1}(\mathbb{R}^d)\cap L^r(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{-d/p,p}(\mathbb{R}^d)\hookrightarrow B^{-d/p}_{\infty,1}(\mathbb{R}^d),\qquad 1\le r<p\le 2.$$

As a consequence of Theorem 4.2, we establish the relationship between the Sobolev space and the Barron spectrum space.

Definition 4.3. Let $s>0$ and $1\le p<\infty$. The Sobolev space $W^{s,p}(\mathbb{R}^d)$ is defined as a class of functions that together with all the distributional derivatives of order less than s are in $L^p(\mathbb{R}^d)$ with finite norm

$$\|f\|_{W^{s,p}(\mathbb{R}^d)}:=\sum_{\|\beta\|_1\le k}\left\|\nabla^\beta f\right\|_{L^p(\mathbb{R}^d)}+\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{|\nabla^k f(x)-\nabla^k f(y)|^p}{|x-y|^{d+(s-k)p}}\mathrm{d}x\mathrm{d}y\right)^{1/p},$$

where $k=\lfloor s\rfloor$ is the integer part of s. The double integral is dropped when s is an integer, and we make the obvious modification when $p=\infty$.

Theorem 4.3. (i) *If $1\le p\le 2$ and $\alpha>s+d/p>0$, then*

$$W^{\alpha,p}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d). \tag{4.12}$$

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

$$W^{s+d,1}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,1}(\mathbb{R}^d). \tag{4.13}$$

The inclusion (4.12) essentially fails for $\alpha=s+d/p$ with $1<p\le 2$; see Remark 4.1.

To prove Theorem 4.3, we need the following Hardy type inequality proved in [32].

Lemma 4.1. *Let $k\in\mathbb{N}$ and $d\ge 2$. There exists C such that for any $f\in W^{k,1}(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}^d}|x|^{k-d}|\widehat{f}(x)|\,\mathrm{d}x\le C\left\|\nabla^k f\right\|_{L^1(\mathbb{R}^d)}.$$

Proof of Theorem 4.3. (i) To prove (4.12), we distinguish the following two cases.

Case a). If α is not an integer and $\alpha>s+d/p$ with $1\le p\le 2$, then the embedding (4.12) follows from (4.5) with $q=p$ and the equivalence between the Sobolev space and the Besov space, i.e.,

$$W^{\alpha,p}(\mathbb{R}^d)=B^\alpha_{p,p}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d).$$

Case b). If α is an integer and $\alpha>s+d/p$ with $1\le p\le 2$, then (4.12) follows from (4.5) with $q=\infty$ and the embedding $W^{\alpha,p}(\mathbb{R}^d)\hookrightarrow B^\alpha_{p,\infty}(\mathbb{R}^d)$, i.e.,

$$W^{\alpha,p}(\mathbb{R}^d)\hookrightarrow B^\alpha_{p,\infty}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{s,p}(\mathbb{R}^d).$$

(ii) When $s+d$ is not an integer, we obtain (4.13) by taking $p=1$ in (4.4) and the equivalence $B^{s+d}_{1,1}(\mathbb{R}^d)=W^{s+d,1}(\mathbb{R}^d)$.

If $s+d$ is an integer, then we take $k=s+d>0$ in Lemma 4.1 and obtain

$$\|f\|_{v_s} \lesssim \left\|\nabla^{s+d} f\right\|_{L^1(\mathbb{R}^d)}.$$

This immediately implies

$$\|f\|_{\mathcal{B}_{s,1}(\mathbb{R}^d)} \lesssim \|f\|_{W^{s+d,1}(\mathbb{R}^d)}.$$

Hence (4.13) is valid. □

We also study the relationship between the Barron spectrum space and the Sobolev space in another direction. The starting point is Corollary 2.2, which shows that Fourier inversion theorem is valid for any $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ with $s\geq 0$, and we may study the derivative of f. We firstly introduce the Hölder space $C^s(\mathbb{R}^d)$; see [57].

Definition 4.4. Let $s>0$, $k=\lfloor s\rfloor$ and $\gamma=s-\lfloor s\rfloor$. The space $C^s(\mathbb{R}^d)$ is defined as a class of $k-$th order differentiable functions f with finite norm

$$\begin{aligned}\|f\|_{C^s(\mathbb{R}^d)} := &\sup_{x\in\mathbb{R}^d}|f(x)| + \max_{\|\alpha\|_1=k}\sup_{x\in\mathbb{R}^d}|\nabla^\alpha f(x)|\\ &+\max_{\|\alpha\|_1=k}\sup_{x,y\in\mathbb{R}^d,x\neq y}\frac{|\nabla^\alpha f(x)-\nabla^\alpha f(y)|}{|x-y|^\gamma}.\end{aligned}$$

Theorem 4.4. *Let $1\leq p\leq 2$ and $s\geq 0$. Then*

$$\mathcal{B}_{s,p}(\mathbb{R}^d)\hookrightarrow C^s(\mathbb{R}^d) \tag{4.14}$$

with

$$\|\nabla^s f\|_{C^0(\mathbb{R}^d)} \leq (2\pi)^s\|f\|_{v_s} \qquad \text{when } s\in\mathbb{N}\cup\{0\}. \tag{4.15}$$

and

$$|\nabla^k f(x)-\nabla^k f(y)|\leq 2(2\pi)^s(1+d^{\gamma/2})\|f\|_{v_s}|x-y|^\gamma \qquad \text{a.e. } x,y\in\mathbb{R}^d \quad \text{and} \quad k=\lfloor s\rfloor. \tag{4.16}$$

Theorem 4.4 shows that any function in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ with a nonnegative index s is a smooth function. In particular, any function in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ with $s\in(0,1)$ is a Hölder function with a Hölder constant $2(2\pi)^s(1+d^{s/2})\|f\|_{v_s}$.

Proof of Theorem 4.4. (i) By Corollary 2.2, we obtain $\widehat{f}\in L^1(\mathbb{R}^d)$ for any $f\in\mathcal{B}_{1,p}(\mathbb{R}^d)$. Therefore, the Fourier inversion holds true: For a.e. $x\in\mathbb{R}^d$,

$$f(x)=\int_{\mathbb{R}^d}\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi.$$

This implies that f is a bounded and continuous function and

$$\|f\|_{C^0(\mathbb{R}^d)} \lesssim \|f\|_{\mathcal{B}_{1,p}(\mathbb{R}^d)}.$$

Next, for any $j\in\{1,\cdots,d\}$ and $-\infty<h<\infty$, we write, for a.e. $x\in\mathbb{R}^d$,

$$\frac{f(x+he_j)-f(x)}{h}=\int_{\mathbb{R}^d}\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\frac{e^{i2\pi h\xi_j}-1}{h}\mathrm{d}\xi.$$

By the fact that $\|f\|_{v_1}<\infty$, we conclude that for any $j\in\{1,\cdots,d\}$,

$$\frac{\partial f}{\partial x_j}(x)=2\pi i\int_{\mathbb{R}^d}\xi_j\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\mathrm{d}\xi \qquad \textit{a.e. } x\in\mathbb{R}^d, \tag{4.17}$$

and hence ∇f is continuous and

$$\|\nabla f\|_{C^0(\mathbb{R}^d)}\le 2\pi\|f\|_{v_1}.$$

This immediately implies the embedding (4.14) with $s=1$.

Suppose (4.14) holds true with $s=k$. This means that for any $f\in\mathcal{B}_{k,p}(\mathbb{R}^d)$, $f\in C^k(\mathbb{R}^d)$. We are ready to prove that (4.14) is also valid for $s=k+1$. By Theorem 2.2, we obtain that for any $f\in\mathcal{B}_{k+1,p}(\mathbb{R}^d)$, $f\in\mathcal{B}_{k,p}(\mathbb{R}^d)$ and then

$$\partial^\alpha f(x)=(2\pi i)^k\int_{\mathbb{R}^d}\xi^\alpha\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\mathrm{d}\xi \qquad \textit{a.e. } x\in\mathbb{R}^d,$$

where $\xi^\alpha=\xi_1^{\alpha_1}\cdots\xi_d^{\alpha_d}$ with nonnegative integers α_i and $\|\alpha\|_1=\|(\alpha_1,\cdots,\alpha_d)\|_1=k$. By the inductive hypothesis, we have

$$\left\|\nabla^k f\right\|_{C^0(\mathbb{R}^d)}\le(2\pi)^k\|f\|_{\mathcal{B}_{k,p}(\mathbb{R}^d)}. \tag{4.18}$$

Proceeding along the same line that leads to (4.17), we write that for any $j\in\{1,\cdots,d\}$, $h\in\mathbb{R}$ and $\|\alpha\|_1=k$,

$$\frac{\partial^\alpha f(x+he_j)-\partial^\alpha f(x)}{h}=(2\pi i)^k\int_{\mathbb{R}^d}\xi^\alpha\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\frac{e^{i2\pi h\xi_j}-1}{h}\mathrm{d}\xi \qquad \textit{a.e. } x\in\mathbb{R}^d.$$

From the fact that $\|f\|_{v_{k+1}}<\infty$, it follows that for any $j\in\{1,\cdots,d\}$,

$$\frac{\partial(\partial^\alpha)f}{\partial x_j}(x)=(2\pi i)^{k+1}\int_{\mathbb{R}^d}\xi^\alpha\xi_j\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\mathrm{d}\xi, \qquad \textit{a.e. } x\in\mathbb{R}^d.$$

In particular, we obtain that for any multi-index α with $\|\alpha\|_1=k+1$,

$$\partial^\alpha f(x)=(2\pi i)^{k+1}\int_{\mathbb{R}^d}\xi^\alpha\widehat{f}(\xi)e^{i2\pi x\cdot\xi}\mathrm{d}\xi, \qquad \textit{a.e. } x\in\mathbb{R}^d. \tag{4.19}$$

This immediately gives us that for $\alpha=(\alpha_1,\cdots,\alpha_d)$ with $\|\alpha\|_1=k+1$,

$$\|\nabla^\alpha f\|_{C^0(\mathbb{R}^d)}\le(2\pi)^{k+1}\|f\|_{v_{k+1}},$$

which together with (4.18) gives that the embedding (4.14) is valid with $s=k+1$. By induction, the embedding (4.14) is valid for any $s\in\mathbb{N}$. The estimate (4.15) follows from (4.19).

(ii) Recall that $\mathcal{B}_{s,p}(\mathbb{R}^d)\hookrightarrow\mathcal{B}_{k,p}(\mathbb{R}^d)$ by Theorem 2.2. Then for any $s>0$ with $k=\lfloor s\rfloor$ and $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$, we conclude $f\in\mathcal{B}_{k,p}(\mathbb{R}^d)$. Hence the identity (4.19) is true with $\|\alpha\|_1=k$. Thus, for any a.e. $x,y\in\mathbb{R}^d$,

$$\partial^\alpha f(x)-\partial^\alpha f(y)=(2\pi\imath)^k\int_{\mathbb{R}^d}\xi^\alpha\widehat{f}(\xi)\left(e^{i2\pi x\cdot\xi}-e^{i2\pi y\cdot\xi}\right)\mathrm{d}\xi.$$

By the inequality that for any $x,y,\xi\in\mathbb{R}^d$,

$$\left|e^{i2\pi x\cdot\xi}-e^{i2\pi y\cdot\xi}\right|\le\min(2,2\pi|x-y||\xi|),$$

we write, for any a.e. $x,y\in\mathbb{R}^d$,

$$\begin{aligned}|\partial^\alpha f(x)-\partial^\alpha f(y)|&\le(2\pi)^k\int_{\mathbb{R}^d}|\xi|^k|\widehat{f}(\xi)|\left|e^{i2\pi x\cdot\xi}-e^{i2\pi y\cdot\xi}\right|\mathrm{d}\xi\\&\le(2\pi)^k\int_{\mathbb{R}^d}\|\xi\|_1^k|\widehat{f}(\xi)|\min(2,2\pi|x-y||\xi|)\mathrm{d}\xi\\&\le(2\pi)^{k+1}|x-y|\int_{|\xi|\le\pi^{-1}|x-y|^{-1}}\|\xi\|_1^k|\xi||\widehat{f}(\xi)|\mathrm{d}\xi\\&\quad+2(2\pi)^k\int_{|\xi|>\pi^{-1}|x-y|^{-1}}\|\xi\|_1^k|\widehat{f}(\xi)|\mathrm{d}\xi.\end{aligned}$$

A direct calculation gives us that

$$\begin{aligned}\int_{|\xi|\le\pi^{-1}|x-y|^{-1}}\|\xi\|_1^k|\xi||\widehat{f}(\xi)|\mathrm{d}\xi&\le\int_{|\xi|\le\pi^{-1}|x-y|^{-1}}\|\xi\|_1^s|\xi|^{1-\gamma}|\widehat{f}(\xi)|\mathrm{d}\xi\\&\le\pi^{\gamma-1}|x-y|^{\gamma-1}\int_{|\xi|\le\pi^{-1}|x-y|^{-1}}\|\xi\|_1^s|\widehat{f}(\xi)|\mathrm{d}\xi\\&\le\pi^{\gamma-1}|x-y|^{\gamma-1}\|f\|_{v_s},\end{aligned}$$

and

$$\begin{aligned}\int_{|\xi|>\pi^{-1}|x-y|^{-1}}\|\xi\|_1^k|\widehat{f}(\xi)|\mathrm{d}\xi&=\int_{|\xi|>\pi^{-1}|x-y|^{-1}}\|\xi\|_1^s\|\xi\|_1^{-\gamma}|\widehat{f}(\xi)|\mathrm{d}\xi\\&\le(\pi\sqrt{d})^\gamma|x-y|^\gamma\|f\|_{v_s}.\end{aligned}$$

Combining the above three inequalities, we conclude that for any a.e. $x,y\in\mathbb{R}^d$,

$$|\partial^\alpha f(x)-\partial^\alpha f(y)|\le2^{k+1}\pi^s(1+d^{\gamma/2})\|f\|_{v_s}|x-y|^\gamma. \tag{4.20}$$

This together with (4.18) immediately implies

$$\|f\|_{C^s(\mathbb{R}^d)} \lesssim \|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}.$$

Hence the embedding (4.14) holds true when s is positive but it is not an integer.

The estimate (4.16) follows from (4.20). □

The proof of Theorem 4.4 is elementary and it yields the precise estimates (4.15) and (4.16). If we give up these two estimates, then the relation (4.14) may be proved in a much simpler way. Resorting to the relation (4.6), and using $B^s_{\infty,1}(\mathbb{R}^d) \hookrightarrow B^s_{\infty,\infty}(\mathbb{R}^d)$, we get $\mathcal{B}_{s,p}(\mathbb{R}^d) \hookrightarrow B^s_{\infty,\infty}(\mathbb{R}^d)$. This immediately implies (4.14).

Remark 4.1. It follows from Corollary 4.2, Theorem 4.3 and Theorem 4.4 that the Barron spectrum space is an intermediate space among certain Besov spaces and certain Sobolev spaces. We may also exploit these results to show that the embedding (4.12) is sharp.

(i) If s is a non negative integer and $\alpha > s+d/p$ with $1 \le p \le 2$, then

$$W^{\alpha,p}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,p}(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d) \hookrightarrow W^{s,\infty}(\mathbb{R}^d). \tag{4.21}$$

(ii) The embedding $W^{s+d/p,p}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,p}(\mathbb{R}^d)$ fails when $s \ge 0$ and $1 < p \le 2$. Otherwise, we would have

$$W^{s+d/p,p}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,p}(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d).$$

Hence $W^{s+d/p,p}(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d)$ for $1 < p \le 2$. This is absurd.

The exceptional case is $p = 1$, i.e.,

$$W^{s+d,1}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,1}(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d) \hookrightarrow W^{s,\infty}(\mathbb{R}^d) \qquad s \in \mathbb{N} \cup \{0\}, d \ge 2,$$
$$W^{s+d,1}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,1}(\mathbb{R}^d) \hookrightarrow C^s(\mathbb{R}^d) \qquad s \ge 0, s \notin \mathbb{N} \cup \{0\}.$$

(iii) The inclusion $\mathscr{L}^p_{s+d/p}(\mathbb{R}^d) \hookrightarrow \mathcal{B}_{s,p}(\mathbb{R}^d)$ essentially fails for $1 < p \le 2$ because $\mathscr{L}^p_{s+d/p}(\mathbb{R}^d) = W^{s+d/p,p}(\mathbb{R}^d)$ when $s+d/p$ is a positive integer.

5 The radial functions in $\mathcal{B}_{s,p}(\mathbb{R}^d)$

The functions with rotation symmetry in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ frequently appear in the radial basis functions neural network; see [23, 41]. Denote by $\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)$ the set consisting of radial functions in $\mathcal{B}_{s,p}(\mathbb{R}^d)$. The theorem below shows that the functions in $\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)$ with certain p are smooth, and they become smoother as increasing of the dimension, while at a cost of an exponentially large embedding constant with respect to the dimension.

Theorem 5.1. *Let $d>1, 1\le p<2d/(d+1)$ and $s\ge 0$. Then*

$$\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)\hookrightarrow W^{\frac{1}{2}(s+d/p'),2}(\mathbb{R}^d) \tag{5.1}$$

with

$$\|f\|_{W^{\frac{1}{2}(s+d/p'),2}(\mathbb{R}^d)}\le\left(\frac{\sqrt{A_d}}{2}+C(p;0,s)^{1-p/2}\right)\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)},$$

where for $d\ge 3$,

$$A_d=5\left(\frac{(d-1)p}{2d-(d+1)p}\right)^{1/p'}(\Gamma(d/2))^{\frac{d+1}{(d-1)p'}}\pi^{-\frac{d}{2p'}},$$

and for $d=2$,

$$A_2=\left(\frac{p}{\pi^2(4-3p)}\right)^{1/p'}.$$

Remark 5.1. The remarkable point of this theorem is that the smoothness index of the functions in $\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)$ may exceed s when $1<p<2d/(d+1)$ and d is sufficiently large, i.e., $d>sp'$. Tracking the constant in (5.1), we find it is exponential large with respect to d, i.e.,

$$\text{Imbedding constant}\sim\left(\frac{d}{2\pi e}\right)^{d/(4p')}d^{1/[4p']}\qquad\text{as}\quad d\to\infty.$$

It follows from (5.1) that $\mathcal{B}^{\mathrm{rad}}_{s,1}(\mathbb{R}^d)\hookrightarrow W^{s/2,2}(\mathbb{R}^d)$. This means that the radial functions in $\mathcal{B}_{s,1}(\mathbb{R}^d)$ do not become smoother in high dimension. This may be due to the fact that the Fourier transform of the radial function in $L^1(\mathbb{R}^d)$ does not have a uniform decaying rate at infinity; see [52].

To prove Theorem 5.1, we recall the following representation formula for the Fourier transform acting on the radial function, which may be found in [52, Theorem 3.3 in Chapter IV and p. 155, footnote 8].

Lemma 5.1. *Let $d\ge 2$ and Bessel function $J_{(d-2)/2}$ be as in* (3.3). *Suppose f is a radial function in $L^p(\mathbb{R}^d)$ with $1\le p\le 2$, thus $f(x)=f_0(|x|)$ for a.e. $x\in\mathbb{R}^d$. Then the Fourier transform $\widehat{f}$ is also radial and has the form $\widehat{f}(x)=F_0(|x|)$ for all $x\in\mathbb{R}^d$, where*

$$F_0(|x|)=F_0(r)=\frac{2\pi}{r^{[(d-2)/2]}}\int_0^\infty f_0(s)J_{(d-2)/2}(2\pi rs)s^{d/2}\mathrm{d}s.$$

We also need a pointwise estimate for the spherical Bessel function.

Lemma 5.2. *Let j_ν be the spherical Bessel function as in* (3.2).

(i) *If $\nu\ge 1/2$, then for any $x>0$,*

$$|j_\nu(x)|\le 5\nu^{1/6}2^\nu\Gamma(\nu+1)|x|^{-1/2-\nu}. \tag{5.2}$$

(ii) *There holds*

$$\sup_{x>0}|x^{1/2}j_0(x)|\le\sqrt{\frac{2}{\pi}}. \tag{5.3}$$

The estimate (5.2) may be proved by checking the constant in [43, Theorem 2.1], and (5.3) may be found in [55, Theorem 7.31.2].

Proof of Theorem 5.1. We prove $p=1$ firstly. Using Hölder's inequality and the Hausdorff-Young inequality, we obtain

$$\int_{\mathbb{R}^d}\|x\|_1^s|\widehat{f}(x)|^2\mathrm{d}x\le\|f\|_{v_s}\left\|\widehat{f}\right\|_{L^\infty(\mathbb{R}^d)}\le\|f\|_{v_s}\|f\|_{L^1(\mathbb{R}^d)}\le\frac{1}{4}\|f\|^2_{\mathcal{B}_{s,1}}.$$

It follows from (2.7) that

$$\|f\|^2_{L^2(\mathbb{R}^d)}\le C(1;0,s)\|f\|^2_{\mathcal{B}_{s,1}}.$$

A combination of the above two inequalities gives (5.1) with $p=1$.

It remains to consider $1<p<2d/(d+1)$. For any $f\in\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)$, let $f(x)=f_0(|x|)$. By Lemma 5.1, we have

$$\widehat{f}(x)=\omega_{d-1}\int_0^\infty j_{d/2-1}(2\pi r|x|)f_0(r)r^{d-1}\mathrm{d}r. \tag{5.4}$$

We firstly assume $d\ge3$. It follows from Hölder's inequality that

$$\begin{aligned}&\left|\int_0^\infty j_{d/2-1}(2\pi r|x|)f_0(r)r^{d-1}\mathrm{d}r\right|\\ \le&\left(\int_0^\infty|f_0(r)|^pr^{d-1}\mathrm{d}r\right)^{1/p}\left(\int_0^\infty|j_{d/2-1}(2\pi r|x|)|^{p'}r^{d-1}\mathrm{d}r\right)^{1/p'}\\ =&\frac{1}{\omega_{d-1}^{1/p}(2\pi|x|)^{d/p'}}\|f\|_{L^p(\mathbb{R}^d)}\left(\int_0^\infty|j_{d/2-1}(s)|^{p'}s^{d-1}\mathrm{d}s\right)^{1/p'}.\end{aligned}$$

Note that for $\nu\ge0$,

$$j_\nu(x)=\frac{\Gamma(\nu+1)}{\Gamma(\nu+1/2)\Gamma(1/2)}\int_0^\pi e^{ix\cos\theta}\sin^{2\nu}\theta\mathrm{d}\theta,$$

which immediately implies

$$|j_\nu(x)|\le\frac{\Gamma(\nu+1)}{\Gamma(\nu+1/2)\Gamma(1/2)}\int_0^\pi\sin^{2\nu}\theta\mathrm{d}\theta=1. \tag{5.5}$$

Let $K>0$ be a constant to be determined later on. By (5.5), we get

$$\int_0^K|j_{d/2-1}(s)|^{p'}s^{d-1}\mathrm{d}s\le\int_0^Ks^{d-1}\mathrm{d}s=\frac{K^d}{d}.$$

Using (5.2) and noting $1<p<2d/(d+1)$, we obtain

$$\int_K^\infty |j_{d/2-1}(s)|^{p'} s^{d-1}\mathrm{d}s \le [5(d/2-1)^{1/6}2^{d/2-1}\Gamma(d/2)]^{p'}\int_K^\infty s^{d-1-(d-1)p'/2}\mathrm{d}s$$
$$\le [5d^{1/6}2^{d/2-7/6}\Gamma(d/2)]^{p'}\frac{K^{d-(d-1)p'/2}}{[(d-1)p'/2]-d}.$$

The above two estimates give us that

$$\int_0^\infty |j_{d/2-1}(s)|^{p'} s^{d-1}\mathrm{d}s \le \frac{K^d}{d}+[5d^{1/6}2^{d/2-7/6}\Gamma(d/2)]^{p'}\frac{K^{d-(d-1)p'/2}}{[(d-1)p'/2]-d}.$$

Optimizing with respect to K, choosing

$$K=\Big(5d^{1/6}2^{d/2-7/6}\Gamma(d/2)\Big)^{2d/[d-1]},$$

and noting $1<p<2d/(d+1)$, we obtain

$$\sup_{|x|>0}|x|^{d/p'}|\widehat{f}(x)| \le \frac{5^{2d/[(d-1)p']}2^{-\frac{d+3}{3(d-1)p'}}}{(1-2d/[(d-1)p'])^{1/p'}}[\Gamma(d/2)]^{\frac{d+1}{(d-1)p'}}\pi^{-\frac{d}{2p'}}\|f\|_{L^p(\mathbb{R}^d)}$$
$$\le A_d\|f\|_{L^p(\mathbb{R}^d)}.$$

Using the above inequality, we obtain, for $s\ge 0$, there holds

$$\int_{\mathbb{R}^d}|x|^{s+d/p'}|\widehat{f}(x)|^2\mathrm{d}x \le \sup_{|\xi|>0}|x|^{d/p'}|\widehat{f}(x)|\int_{\mathbb{R}^d}\|x\|_1^s|\widehat{f}(x)|^2\mathrm{d}x$$
$$\le A_d\|f\|_{L^p(\mathbb{R}^d)}\|f\|_{v_s}\le \frac{A_d}{4}\|f\|^2_{\mathcal{B}_{s,p}(\mathbb{R}^d)}.$$

Proceeding along the same line that leads to (2.7), we obtain

$$\|f\|^2_{L^2(\mathbb{R}^d)}\le C(p;0,s)^{2-p}\|f\|^2_{\mathcal{B}_{s,p}(\mathbb{R}^d)}.$$

A combination of the above two inequalities gives that $f\in W^{s/2+d/(2p'),2}(\mathbb{R}^d)$.

The proof for $d=2$ is the same provided that we use (5.3) in lieu of (5.2). □

In what follows, we study the decaying behavior of the functions in $\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)$. The decaying behavior of the radial function in various function spaces dates back to the so-called Strauss' inequality [53]; see also [36].

Theorem 5.2. *Let $d\ge 2$, $1\le p\le 2$ and $s\ge 0$. Every function $f\in\mathcal{B}^{\mathrm{rad}}_{s,p}(\mathbb{R}^d)$ is almost everywhere equal to a function f_0, which is a continuous function except $x\ne 0$ such that*

$$\sup_{x\in\mathbb{R}^d\setminus\{0\}}|x|^{(d-1)/2}|f_0(x)|\le C\|f\|^\gamma_{L^p(\mathbb{R}^d)}\|f\|^{1-\gamma}_{v_s}, \tag{5.6}$$

where $C>0$ *is a constant depending on* d,p *and* s*, and*

$$\gamma=\frac{s+(d-1)/2}{s+d/p}.$$

Proof. It follows from Corollary 2.2 that Fourier inversion theorem is valid for any $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ with $s\geq 0$. Denote $f_0(r)=f(|x|)$. The representation formula (5.4) is also valid by exchanging the role of f and its Fourier transform, i.e., for any $r>0$,

$$f_0(r)=\omega_{d-1}\int_0^\infty j_{d/2-1}(2\pi r\rho)\widehat{f_0}(\rho)\rho^{d-1}\mathrm{d}\rho.$$

We firstly consider $d\geq 3$. By (5.2), we have that for any $r>0$,

$$|f_0(r)|\leq 5\sqrt{2\pi}(d/2-1)^{1/6}r^{(1-d)/2}\int_0^\infty|\widehat{f_0}(\rho)|\rho^{(d-1)/2}\mathrm{d}\rho.$$

Let $K>0$ be a constant to be specified later on. By Hölder's inequality and the Hausdorff-Young inequality, we obtain

$$\begin{aligned}\int_0^K|\widehat{f_0}(\rho)|\rho^{(d-1)/2}\mathrm{d}\rho&\leq\left(\int_0^K\rho^{(d-1)(p/2-p/p')}\mathrm{d}\rho\right)^{1/p}\left(\int_0^K|\widehat{f_0}(\rho)|^{p'}\rho^{d-1}\mathrm{d}\rho\right)^{1/p'}\\&\leq\omega_{d-1}^{-1/p'}\left(\int_0^K\rho^{(d-1)(p/2-p/p')}\mathrm{d}\rho\right)^{1/p}\left\|\widehat{f}\right\|_{L^{p'}(\mathbb{R}^d)}\\&\leq\frac{K^{[d/p-(d-1)/2]}}{\omega_{d-1}^{1/p'}(d-(d-1)p/2)^{1/p}}\|f\|_{L^p(\mathbb{R}^d)}.\end{aligned}$$

It follows from (5.2) and $s\geq 0$ that

$$\int_K^\infty|\widehat{f_0}(\rho)|\rho^{(d-1)/2}\mathrm{d}\rho\leq\omega_{d-1}^{-1}K^{(1-d)/2-s}\|f\|_{v_s}.$$

Combining the above two inequalities and optimize with respect to K, i.e.,

$$K=\left(\frac{\|f\|_{v_s}}{\|f\|_{L^p(\mathbb{R}^d)}}\right)^{1/(s+d/p)}\left(\frac{2s+d-1}{2(d/p-(d-1)/2)^{1/p'}}\right)^{1/(s+d/p)}\left(\frac{p}{\omega_{d-1}}\right)^{1/(d+sp)},$$

we obtain (5.6).

The proof for $d=2$ is the same provided that we replace (5.3) by (5.2). □

6 Application to neural network approximation

As an application of $\mathcal{B}_{s,p}(\mathbb{R}^d)$, we prove L^q approximation rate for two-layer neural network (NN) with $\mathcal{B}_{s,p}(\mathbb{R}^d)$ as the target function space. We shall not give a literature review on the approximation of NN, and refer to [45] and [14] for surveys.

Let σ be an activation function on $\mathbb{R}$ and a two-layer neural network is defined as

$$\mathcal{M}_n(\sigma):=\left\{h(x)=\sum_{i=1}^n c_i\sigma(a_i\cdot x+b_i)\mid b_i,c_i\in\mathbb{R},a_i,x\in\mathbb{R}^d\right\}.$$

We consider the approximation over the hypercube $D:=(-1/2,1/2)^d$, and define $\gamma(q)=\min(1-1/q,1/2)$ for any $1\le q<\infty$. The main result reads

Theorem 6.1. *For any $1\le q<\infty$ and $1\le p\le 2$, suppose that $f\in\mathcal{B}_{1-1/q,p}(\mathbb{R}^d)$. If there exist $\alpha>q/(q-1)$ and $A>0$ such that*

$$|\sigma(t)|\le A(1+|t|)^{-\alpha}, \tag{6.1}$$

then there exists $f_n\in\mathcal{M}_n(\sigma)$ such that

$$\|f-f_n\|_{L^q(D)}\le C\sqrt{q}d^{1/[2q]}n^{-\gamma(q)}\left(\|f\|_{\upsilon_{1-1/q}}+\|f\|_{\upsilon_{-1/q}}\right), \tag{6.2}$$

where C depends on σ,x_0,A and α, while is independent of q,d and n.

If $f\in\mathcal{B}_{s,p}(\mathbb{R}^d)$ with $s\ge 1-1/q$, then (6.2) *changes to*

$$\|f-f_n\|_{L^q(D)}\le C\sqrt{q}d^{1-1/[2q]-s}n^{-\gamma(q)}\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}. \tag{6.3}$$

The above theorem proves L^q approximation bound for the target function belonging to the Barron spectrum space for commonly used activation functions. The rate $\gamma(q)$ is the same with [15], which may be improved when $1\le q<2$; cf., Corollary 6.1.

MAKOVOZ [38, Theorem 3] proved L^q-error when σ is a sigmoid with the rate $\gamma(q)=1/2+1/[q^*d]$, where q^* is the smallest even integer satisfying $q^*\ge q$, the extra term $1/[q^*d]$ is significant for small d while vanishes when $d\to\infty$. This rate is better than $\gamma(q)$ in (6.2), while the finiteness of $\|f\|_{\upsilon_1}$ has been assumed, and the dependence of the error bound on d and q is unclear.

Taking $s=1/2$ and $q=2$ in (6.3), we obtain

$$\|f-f_n\|_{L^2(D)}\le Cd^{1/4}n^{-1/2}\|f\|_{\mathcal{B}_{1/2,p}(\mathbb{R}^d)}.$$

This recovers [48, $m=0$ in Theorem 2], in which the authors have assumed that $\|f\|_{\upsilon_0}+\|f\|_{\upsilon_1}$ is finite. Recently, the same authors improved this result in [49, Theorem 5] when σ is a sigmoid under weaker regularity on f, i.e., a finite $\|f\|_{\upsilon_0}+\|f\|_{\upsilon_{1/2}}$ suffices for the above error bound, while this improvement is achieved at the cost of a constant that depends on the dimension exponentially. We proved the algebraical dependence of the error on the dimension, which has been ignored in [48].

If the activation function is ReLU, i.e., $\sigma(x)=\max(x,0)$, then Breiman [9] proved

$$\|f-f_n\|_{L^2(D)}\le Cn^{-1/2}\|f\|_{\upsilon_2},$$

where the dependence of the error bound on d is unknown. We may infer L^q-estimate from the uniform approximation result in [31, Theorem 2]: there exists a universal constant C such that

$$\|f-f_n\|_{L^q(D)} \le \|f-f_n\|_{L^\infty(D)} \le C\max\Big(\sqrt{\log n},\sqrt{d}\Big)\, n^{-1/2-1/d}\|f\|_{v_2}.$$

If we take $s=1$ and $p=2$ in (6.3), then

$$\|f-f_n\|_{L^q(D)} \le C\sqrt{q}d^{-1/[2q]}n^{-1/2}\|f\|_{\mathcal{B}_{1,2}(\mathbb{R}^d)}.$$

The convergence rate is less sharp than that in [31, Theorem 2], while our result requires a smaller index on the spectrum norm with an algebraic decay on the dimension.

The estimate (6.2) is invalid for $q=\infty$, while we believe similar estimate remains true if we exploit the deep sampling theorem of Dudley as in [38,39] and [31], which will be left for further pursuit.

Remark 6.1. As observed in [48, Corollary 1], the decaying condition (6.1) need not be satisfied by σ per se, a finite translation of σ suffices. Moreover, this condition is stronger than the one in [48], i.e., they have assumed $\alpha=1$. Nevertheless, this condition is satisfied by a finite shift of the commonly used activation functions such as sigmoid, arctan, hyperbolic tangent, ReLU, and $k-$th power of ReLU for $k=0,1,2,\cdots$; We refer to [48, Table 1] for an elaboration on this point.

Our proof is based on the representation (6.4), which has been *assumed* in the proof of several approximation results for NN; see, e.g., [30]. While this formula is valid for functions in $\mathcal{B}_{s,p}(\mathbb{R}^d)$ with $s\ge 0$ as proved in Corollary 2.2. Another tool in proving Theorem 6.1 is the cube slicing theorem of Ball [1, Theorem 4], which helps us to clarify the dependence of the estimate on the dimension.

Lemma 6.1. *Define the hyperplane*

$$H_t := \Big\{ x\in\mathbb{R}^d \mid \omega\cdot x = t,\ |\omega|=1 \Big\}$$

for any $t\in\mathbb{R}$. *Then*

$$|H_t\cap D| \le \sqrt{2},$$

where $D=(-1/2,1/2)^d$ *is the hypercube.*

Proof of Theorem 6.1. By the growth condition on σ, we conclude that $\sigma\in L^1(\mathbb{R})$, and there exists $x_0\neq 0$ such that $\widehat{\sigma}(x_0)\neq 0$. For any $f\in\mathcal{B}_{1-1/q,p}(\mathbb{R}^d)$ with $1\le q<\infty$, using Corollary 2.2, we have $\widehat{f}\in L^1(\mathbb{R}^d)$, hence Fourier inversion theorem is true. Invoking the fact that $\sigma\in L^1(\mathbb{R})$ again, using Fubini's theorem and changing of variables, we obtain

$$f(x) = \int_{\mathbb{R}^d}\int_{\mathbb{R}} \frac{\sigma((\xi/x_0)\cdot x+s)}{\widehat{\sigma}(x_0)}\widehat{f}(\xi)e^{-i2\pi s x_0}\mathrm{d}s\mathrm{d}\xi. \tag{6.4}$$

We rewrite the above representation as an expectation as [2]. Let $z=(\xi,s)\in\mathcal{Z}$ with $\mathcal{Z}:=\mathbb{R}^d\times\mathbb{R}$. For any $1<q\le 1/(1-s)$ when $0<s<1$ and $1<q<\infty$ for $s\ge 1$, we define

$$F(x;z):=\left(\frac{\|\xi\|_1}{|x_0|}\right)^{1/q}\frac{\sigma((\xi/x_0)\cdot x+s)}{\widehat{\sigma}(x_0)}\cos[2\pi(sx_0+\theta(\xi)]\widetilde{h}^{-1}(z),$$
$$h(z):=\left(\frac{|x_0|}{\|\xi\|_1}\right)^{1/q}|\widehat{f}(\xi)|\widetilde{h}(z)\quad\text{with}\quad \widetilde{h}(z):=\left(1+\max\left(0,|s|-\frac{\|\xi\|_1}{2|x_0|}\right)\right)^{-\alpha(1-1/q)},$$
$$w(z):=\frac{h(z)}{Q}\qquad\text{with}\quad Q:=\int_{\mathcal{Z}}h(z)\mathrm{d}z.$$

Hence, we write f as

$$f=Q\mathbb{E}_w[F(x;z)]\qquad\text{with}\qquad \mathbb{E}_wF(x;z):=\int_{\mathcal{Z}}F(x;z)w(z)\mathrm{d}z.$$

The neural network approximation may be viewed as

$$f_n(x;\overline{z})=\frac{Q}{n}\sum_{i=1}^n F(x;z_i),$$

where $\overline{z}=(z_1,\cdots,z_n)$, and it is clear that $f_n\in\mathcal{M}_n(\sigma)$.

Let

$$\overline{w}=\underbrace{w\otimes\cdots\otimes w}_{n}\qquad\text{and}\qquad \overline{\mathcal{Z}}=\underbrace{\mathcal{Z}\otimes\cdots\otimes\mathcal{Z}}_{n},$$

and denote $\overline{w}(\overline{z})=\prod_{i=1}^n w(z_i)$. For any measurable function G defined over $\overline{\mathcal{Z}}$, we define the expectation of G with respect to $\overline{w}$ as

$$\mathbb{E}_{\overline{w}}(G):=\int_{\overline{\mathcal{Z}}}G(\overline{z})\overline{w}(\overline{z})\mathrm{d}\overline{z}.$$

Using Fubini's theorem again, we obtain

$$\mathbb{E}_{\overline{w}}\left(\|f-f_n(\cdot;\overline{z})\|^q_{L^q(D)}\right)=\int_D\mathbb{E}_{\overline{w}}|f(x)-f_n(x;\overline{z})|^q\mathrm{d}x.$$

By definition, we write

$$f(x)-f_n(x;\overline{z})=\frac{Q}{n}\sum_{i=1}^n(\mathbb{E}_wF(x;z)-F(x;z_i)).$$

For $i=1,\cdots,n$, we denote $X_i(x)=\mathbb{E}_wF(x;z)-F(x;z_i)$, which is independent and $\mathbb{E}_{\overline{w}}X_i(x)=0$ for all $x\in D$. By the Marcinkiewicz-Zygmund inequality [13, Theorem 2, p. 386], for any $1\le q<\infty$,

$$\mathbb{E}_{\overline{w}}\left[\left|\sum_{i=1}^n X_i\right|^q\right]\le C^q(q)\mathbb{E}_{\overline{w}}\left[\left(\sum_{i=1}^n X_i^2\right)^{q/2}\right],\tag{6.5}$$

where

$$C(q)=\begin{cases} 1, & 1\le q\le 2,\\ \sqrt{2}\left(\dfrac{\Gamma((q+1)/2)}{\sqrt{\pi}}\right)^{1/q}, & 2<q<\infty. \end{cases}$$

The constant $C(q)$ is optimal; see e.g., [21,25].

Hence, there holds

$$\mathbb{E}_{\overline{w}}\Big(\|f-f_n(\cdot;\overline{z})\|^q_{L^q(D)}\Big)\le\left(\frac{QC(q)}{n}\right)^q\int_D\mathbb{E}_{\overline{w}}\left[\left(\sum_{i=1}^n X_i^2(x)\right)^{q/2}\right]\mathrm{d}x.$$

By Hölder's inequality, for any $a\in\mathbb{R}^n$ and $1<q<\infty$, there holds

$$\|a\|_2\le n^{(1/2-1/q)_+}\|a\|_q,$$

where $(1/2-1/q)_+=1/2-1/q$ if $q\ge 2$, and 0 otherwise.

Note the elementary inequality:

$$(a+b)^q\le 2^{q-1}(a^q+b^q)\qquad a,b\ge 0,\quad 1<q<\infty.$$

Combining the above three inequalities, we obtain

$$\mathbb{E}_{\overline{w}}\Big(\|f-f_n(\cdot;\overline{z})\|^q_{L^q(D)}\Big)\le\left(\frac{2QC(q)}{n}\right)^q n^{1+q(1/2-1/q)_+}\int_D\mathbb{E}_w|F(x;z)|^q\mathrm{d}x. \tag{6.6}$$

It remains to bound Q and $\mathbb{E}_w|F|^q$. A direct calculation gives that for any $\xi\in\mathbb{R}^d$,

$$\begin{aligned}\int_{\mathbb{R}}\widetilde{h}(\xi,s)\mathrm{d}s&=\int_{|s|\le\|\xi\|_1/[2|x_0|]}\mathrm{d}s+\int_{|s|>\|\xi\|_1/[2|x_0|]}(1+|s|-\|\xi\|_1/[2|x_0|])^{-\alpha(1-1/q)}\mathrm{d}s\\&=\frac{\|\xi\|_1}{|x_0|}+\frac{2q}{\alpha(q-1)-q}\\&\le\max(2q/[\alpha(q-1)-q],1/|x_0|)(1+\|\xi\|_1).\end{aligned}$$

This immediately implies

$$Q\le\max(2q/(\alpha(q-1)-q),1/|x_0|)|x_0|^{1/q}\Big(\|f\|_{v_{-1/q}}+\|f\|_{v_{1-1/q}}\Big).$$

For any $x\in D$,

$$|s+\xi\cdot x/x_0|\ge\max(0,|s|-|\xi\cdot x|/|x_0|)\ge\max(0,|s|-\|\xi\|_1/[2|x_0|]),$$

which immediately implies

$$|\sigma|^{q-1}\widetilde{h}^{-q}(z)\le A^{q-1}(1+|s+\xi\cdot x/x_0|)^{-\alpha(q-1)}(1+\max(0,|s|-\|\xi\|_1/[2|x_0|]))^{\alpha(q-1)}\le A^{q-1}.$$

Using Fubini's theorem and the above inequality, we write

$$\begin{aligned}\int_D \mathbb{E}_w|F(x;z)|^q \mathrm{d}x &\le \int_{\mathcal{Z}} \frac{\|\xi\|_1}{|\widehat{\sigma}(x_0)\widetilde{h}(z)|^q|x_0|}\int_D |\sigma((\xi/x_0)\cdot x+s)|^q \mathrm{d}x\, w(z)\mathrm{d}z\\ &\le \frac{A^{q-1}\sqrt{d}}{|\widehat{\sigma}(x_0)|^q}\int_{\mathcal{Z}}\frac{|\xi|}{|x_0|}\int_D |\sigma((\xi/x_0)\cdot x+s)|\mathrm{d}x\, w(z)\mathrm{d}z.\end{aligned}$$

Using the cube slicing Lemma 6.1, we get

$$\begin{aligned}\int_D |\sigma((\xi/x_0)\cdot x+s)|\mathrm{d}x &\le \int_{-\sqrt{d}/2}^{\sqrt{d}/2}\int_{H_t\cap D}|\sigma((|\xi|/x_0)t+s)|\mathrm{d}x\mathrm{d}t\\ &= \int_{-\sqrt{d}/2}^{\sqrt{d}/2}|\sigma((|\xi|/x_0)t+s)||H_t\cap D|\mathrm{d}t\\ &\le \sqrt{2}\int_{-\sqrt{d}/2}^{\sqrt{d}/2}|\sigma((|\xi|/x_0)t+s)|\mathrm{d}t\\ &\le \sqrt{2}\frac{|x_0|}{|\xi|}\int_{\mathbb{R}}|\sigma(t)|\mathrm{d}t\\ &\le \frac{2\sqrt{2}A|x_0|}{|\xi|(\alpha-1)}.\end{aligned}$$

A combination of the above two inequalities gives

$$\int_D \mathbb{E}_w|F(x;z)|^q \mathrm{d}x \le \frac{2\sqrt{2d}A^q}{|\widehat{\sigma}(x_0)|^q(\alpha-1)} \le \frac{2\sqrt{2d}qA^q}{|\widehat{\sigma}(x_0)|^q}.$$

Substituting the bounds for Q and $\mathbb{E}_w|F|^q$ into (6.6), we get

$$\|f-f_n\|_{L^q(D)} \le Cn^{-\gamma(q)}\left(\|f\|_{v_{-1/q}}+\|f\|_{v_{1-1/q}}\right),$$

with $C=C_1(x_0)C(q)d^{1/[2q]}$ and

$$C_1=16A(|x_0|q)^{1/q}\max(q/[\alpha(q-1)-q],1/|x_0|)|\widehat{\sigma}(x_0)|^{-1}.$$

Using the fact $C(q)\simeq\sqrt{q}$ for large q, we obtain (6.2).

Using the interpolation inequality in Lemma 2.1, the above estimate changes to

$$\|f-f_n\|_{L^q(D)} \le Cn^{-\gamma(q)}\big(C(p;1-1/q,s)+C(p;-1/q,s)\big)\|f\|_{\mathcal{B}_{s,p}(\mathbb{R}^d)}.$$

Noting that

$$C(p,1-1/q,s)\simeq d^{1-1/q-s} \qquad \text{and} \qquad C(p,-1/q,s)\simeq d^{-1/q-s}$$

for large d. This gives (6.3) and completes the proof. □

Theorem 6.1 may be extended to domain $D_R := [-R/2, R/2]^d$. The estimate (6.2) changes to

$$\|f - f_n\|_{L^q(D_R)} \le C R^{1+(d-1)/q} \sqrt{q} d^{1/[2q]} n^{-\gamma(q)} \Big(\|f\|_{v_{1-1/q}} + \|f\|_{v_{-1/q}} \Big), \tag{6.7}$$

where C depends on σ, x_0, A and α while independent of q, d, R and n. The proof is the same with that leads to (6.2) except that the constant in the cube slicing Lemma 6.1 is replaced by $\sqrt{2} R^{d-1}$.

For $1 \le q < 2$, the rate $\gamma(q)$ in (6.7) may be slightly improved at a cost of a lager prefactor concerning R, d and a smaller target function space.

Corollary 6.1. *Under the same condition of Theorem* 6.1*, for* $1 \le q < 2$*, there holds*

$$\|f - f_n\|_{L^q(D_R)} \le C R^{d/q+1/2} d^{1/4} n^{-1/2} \|f\|_{\mathcal{B}_{1/2,p}(\mathbb{R}^d)}, \tag{6.8}$$

where C *depends on* σ, x_0, A, p *and* α *while independent of* q, d, R *and* n*.*

Proof. For $1 \le q < 2$, one use Hölder's inequality to obtain

$$\|f - f_n\|_{L^q(D_R)} \le \text{mes } (D_R)^{1/q-1/2} \|f - f_n\|_{L^2(D_R)}.$$

Taking $q = 2$ in (6.7), we obtain

$$\|f - f_n\|_{L^2(D_R)} \le C R^{(d+1)/2} d^{1/4} n^{-1/2} \Big(\|f\|_{v_{1/2}} + \|f\|_{v_{-1/2}} \Big).$$

A combination of the above two inequalities and Lemma 2.1 give (6.8). □

7 Conclusion

We introduce a new function space $\mathcal{B}_{s,p}$ that may be viewed as a classical realization of the Barron function class. The embedding between Barron spectrum space and the Besov space has been established, which leads to the embedding among Barron spectrum space, the Sobolev space and the Bessel potential space. These embeddings also show that $\mathcal{B}_{s,p}$ is smooth with a non-negative index s. The dimension independent approximation rate has been proved for two-layer neural network with $\mathcal{B}_{s,p}$ as a target space. The connection among this space and other Baron type spaces will be discussed in a forthcoming work, partial results in this direction may be found in [11]. The fine properties of the Barron spectrum space and other Barron type spaces may be useful for study the approximation and the estimation of the neural network models [47] as well as the convergence behavior of certain NN based methods for partial differential equations; see; e.g., [18,28,35].

Acknowledgments

The authors thank Professor Weinan E in Peking university and Princeton university for many enlightening discussion. It was him who suggested the term "Barron Spectrum Space". The authors thank the anonymous referees for the inspiring remarks that improve Theorem 6.1.

The work of Meng was supported by the National Natural Science Foundation of China under the grant 11971058, and the work of Ming was supported by the National Natural Science Foundation of China under the grant 11971467.

References

[1] K. Ball, *Cube slicing in R^n*, Proceed. AMS **97** (1986), 465–473.
[2] A.R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory **39** (1993), 930–945.
[3] A.R. Barron, A. Cohen, W. Dahmen, and R.A. DeVore, *Approximation and learning by greedy algorithms*, The Annals of Statistics **36** (2008), 64–94.
[4] R. Bellman, *On the theory of dynamical programming*, Proc. National Acad. Sci. **38(8)** (1952), 716.
[5] J. Bourgain, *A Hardy inequality in Sobolev spaces*, 1981, Vrjie University, Brussels.
[6] ———, *Some examples of multipliers in Sobolev spaces*, 1985, IHES.
[7] W.O. Bray, *Growth and integrability of Fourier transforms on Euclidean spaces*, J. Fourier Anal. Appl. **20** (2014), 1234–1256.
[8] W.O. Bray and M. Pinsky, *Growth properties of the Fourier transforms*, Filomat **26** (2012), 755–760.
[9] L. Breiman, *Hinging hyperplanes for regression, classification and function approximation*, IEEE Trans. Info. Theory **30** (1993), 999–1013.
[10] E.J. Candes, *Ridgelets: Theory and Applications*, Ph.D. thesis, Stanford University, Stanford, 1998.
[11] A. Caragea, P. Petersen, and F. Voigtlaender, *Neural network approximation and estimation of classifiers with classification boundary in Barron class*, 2022, arXiv:2011.09363v2.
[12] N. Chikami, *On Gagliardo-Nirenberg type inequalities in Fourier-Herz spaces*, J. Func. Anal. **275** (2018), 1138–1172.
[13] Y.S. Chou and H. Teicher, *Probability Theory*, Springer-Verlag, 1997, 3rd eds.
[14] R. DeVore, B. Hanin, and G. Petrova, *Neural network approximation*, Acta Numer. **30** (2021), 327–444.
[15] M. J. Donahue, L. Gurvits, C. Darken, and E. Sontag, *Rate of convex approximation in non-Hilbert spaces*, Constr. Approx. **13** (1997), 187220.
[16] D.L. Donoho, *High-dimensional data analysis: The curses and blessings of dimensionality*, AMS Conference on Mathematical Challenges of the 21st Century, AMS, Providence, Rhode Island, 2000.
[17] ———, *Ridge functions and orthonormal ridgelets*, J. Approx. Theory **111** (2001), 143–179.
[18] C.G. Duan, Y.L. Jiao, Y.M. Lai, D.W. Li, X.L. Lv, and Jerry Z.J. Yang, *Convergence rate analysis for deep Ritz method*, Commun. Comput. Phys. **31** (2022), 1020–1048.

[19] W. E, C. Ma, and L. Wu, *The Barron space and the flow-induced function spaces for neural network models*, Constr. Approx. **55** (2022), 369–406, Special Issue on Deep Networks in Approximation Theory; Issue Editors: W. Dahmen, R. A. DeVore, and P. Grohs.

[20] W. E and S. Wojtowytsch, *Representation formula and pointwise properties for Barron functions*, Cal. Var. **61** (2022), no. 46.

[21] D. Ferger, *Optimal constants in the Marcinkiewicz-Zygmund inequalities*, Statis. Probab. Lett. **84** (2014), 96–101.

[22] J. Garcia-Cuerva and J.L. Rubio De Francia, *Weighted Norm Inequalities and Related Topics*, North-Holland-Amsterdam, New York, Oxford, 1985.

[23] F. Girosi and G. Anzellotti, *Rates of convergence for radial basis functions and neural networks*, Artifical Neural Networks for Speech and Visions, R. J. Mammone (Editor), Chapman & Hall, London, 1993, pp. 96–113.

[24] R. Gribonval, G. Kutyniok, M. Nielsen, and F. Voigtlaender, *Approximation spaces of deep neural networks*, Constr. Approx. **55** (2022), 252–367, Special Issue on Deep Networks in Approximation Theory; Issue Editors: W. Dahmen, R. A. DeVore, and P. Grohs.

[25] U. Haagerup, *The best constants in the Khinchine inequality*, Studia Math. **70** (1982), 231–283.

[26] K. Hornik, M. Stinchcombe, H. White, and P. Auer, *Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives*, Neural. Comput. **6** (1994), 1262–1275.

[27] T. Iwabuchi, *Global well-posedness for Keller-Segel system in Besov type spaces*, J. Math. Anal. Appl. **379** (2011), 930–948.

[28] Y.L. Jiao, Y.M. Lai, D.W. Li, X.L. Lv, F.R. Wang, Y. Wang, and J.Z.J. Yang, *A rate of convergence of physics informed neural networks for the linear second order elliptic pdes*, Commun. Comput. Phys. **31** (2022), 1272–1295.

[29] J.K. Jones, *A simple lemma on greedy approximation in Hilbert space and convergence rates for Projection Pursuit Regression and neural network training*, Ann. Statis. **20** (1992), 608–613.

[30] J.M. Klusowski and A.B. Barron, *Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ_1 and ℓ_0 controls*, IEEE Trans. Info. Theory **64** (2018), 7649–7656.

[31] J.M. Klusowski and A.R. Barron, *Uniform approximation by neural networks activated by first order and second order ridge splines*, 2016, arXiv:1607.07819v2.

[32] V.L. Kolyada, *Estimates of Fourier transforms in Sobolev spaces*, Studia Math. **125** (1997), 67–74.

[33] P. Konieczny and T. Yoneda, *On the dispersive effect of the Coriolis force for the stationary Navier-Stokes equations*, J. Diff. Eqn. **250** (2011), 3859–3873.

[34] Z. Lei and F.H. Lin, *Global mild solutions of Navier-Stokes equations*, Comm. Pure and Appl. Math. **LXIV** (2011), 1297–1304.

[35] Y.L. Liao and P.B. Ming, *Deep Nitsche method: Deep Ritz-method with essential boundary conditions*, Commun. Comput. Phys. **29** (2021), 1365–1384.

[36] P.L. Lions, *Symétrie et compacité ans les espaces de sobolev*, J. Func. Anal. **49** (1982), 315–334.

[37] J. 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), 1–46.

[38] Y. Makovoz, *Random approximations and neural networks*, J. Approx. Theory **85** (1996), 98–109.

[39] ______, *Uniform approximation by neural networks*, J. Approx. Theory **95** (1998), 215–228.

[40] Y. Meyer, *Ondeletts et Opérateurs*, Hermann, Paris, 1990, English translation: "Wavelets and Operators", Cambridge University Press, Cambridge, UK, 1992.

[41] H.N. Mhaskar and C.A. Micchelli, *Approximation by superposition of sigmoidal and radial basis functions*, Adv. Appl. Math. **13** (1992), 350–373.

[42] P. Niyogi and F. Girosi, *Generalization bounds for function approximation from scattered noisy data*, Adv. Comput. Math. **10** (1999), 51–80.
[43] A.Y. Olenko, *Upper bound on* $\sqrt{x}J_v(x)$ *and its applications*, Integral Transforms and Special Functions **17** (2006), 455–467.
[44] P.P. Petrushev, *Approximation by ridge functions and neural networks*, SIAM J. Math. Anal. **30** (1998), 155–189.
[45] A. Pinkus, *Approximation theory of the MLP model in neural networks*, Acta Numer. **8** (1999), 143–198.
[46] U. Sgaham, A. Cloninger, and R.R. Coifman, *Provable approximation properties for deep neural networks*, Appl. Comput. Harmon. Anal. **44** (2018), 537–557.
[47] Z.W. Shen, H.Z. Yang, and S.J. Zhang, *Deep network approximation characterized by number of neurons*, Commun. Comput. Phys. **28** (2020), 1768–1811.
[48] J. Siegel and J. Xu, *Approximation rates for neural networks with general activation functions*, Neural Networks **128** (2020), 313–321.
[49] ———, *High-order approximation rates for shallow neural networks with cosine and ReLU*^{k} *activation functions*, Appl. Comput. Harmon. Anal. **58** (2022), 1–26.
[50] E.M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.
[51] ———, *Harmonic Analysis, Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, 1993.
[52] E.M. Stein and W. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton N.J., 1971.
[53] W.A. Strauss, *Existence of solitary waves in high dimensions*, Commun. Math. Phys. **55** (1977), 149–165.
[54] T. Suzuki, *Adaptivity of deep ReLu network for learning in Besov and mixed smooth Besov spaces: Optimal rate and curse of dimensionality*, ICLR (2019).
[55] G. Szego, *Orthogonal Polynomials*, American Mathematical Society, Providence, Rhode Island, 4th eds., 1975.
[56] E.C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford at the Clarendon Press, 2nd eds., 1948.
[57] H. Triebel, *Theory of Function Spaces*, Birkhäuser/Springer Basel AG, Basel, 2010, Reprint of the 1983 Edition.