AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(5): 10852–10865.
DOI: 10.3934/math.2025493
Received: 29 December 2024
Revised: 04 March 2025
Accepted: 06 May 2025
Published: 12 May 2025

Research article

Relaxed conditions for universal approximation by radial basis function neural networks of Hankel translates

Isabel Marrero*

Departamento de Análisis Matemático and Instituto de Matemáticas y Aplicaciones (IMAULL), Universidad de La Laguna (ULL), 38200 La Laguna, Spain

* **Correspondence:** Email: imarrero@ull.es.

Abstract: Radial basis function neural networks (RBFNNs) of Hankel translates of order $\mu > -1/2$ with varying widths whose activation function σ is a.e. continuous, such that $z^{-\mu-1/2}\sigma(z)$ is locally essentially bounded and not an even polynomial, are shown to enjoy the universal approximation property (UAP) in appropriate spaces of continuous and integrable functions. In this way, the requirement that σ be continuous for this kind of networks to achieve the UAP is weakened, and some results that hold true for RBFNNs of standard translates are extended to RBFNNs of Hankel translates.

Keywords: universal approximation property; RBF; nonpolynomiality; neural network; Hankel translation; esssup-norm; density; activation function
Mathematics Subject Classification: 41A30, 46F12

1. Introduction

1.1. Radial basis function neural networks (RBFNNs)

Many complex problems are nowadays modeled and solved by means of neural networks (NNs), which have become a fundamental tool in machine learning and artificial intelligence. While NNs admit many possible architectures, radial basis function neural networks (RBFNNs) may be classified as single hidden layer, feedforward nonlinear NNs. In fact, they consist of three sequential layers: the first or *input layer*, the last or *output layer*, and an intermediate one, referred to as the *hidden layer*. Information flows only in one direction, from the input layer to the output one. Each layer is composed of several *nodes*, which act as neurons in the network. Once an input is received by the neurons in the first layer, it is processed by the neurons in the hidden layer by means of a locally biased *activation function*, thus producing partial outputs that are linearly combined by the neurons in the last layer to render a final output. The nonlinearity of the model comes from the activation function, which, in the case of RBFNNs, is some radial kernel, often a Gaussian.

More specifically, given $d \in \mathbb{N}$, an RBFNN is any function $v : \mathbb{R}^d \to \mathbb{R}$ expressible as

$$v(\mathbf{x}) = \sum_{i=1}^{N} w_i h\left(\frac{\|\mathbf{x} - \mathbf{z}_i\|}{\theta_i}\right), \tag{1.1}$$

where $h : [0, \infty) \to \mathbb{R}$ represents the activation function; $\mathbf{x} \in \mathbb{R}^d$ is the input; $N \in \mathbb{N}$ is the quantity of hidden layer nodes; $(w_1, \ldots, w_N) \in \mathbb{R}^N$ is the N-tuple of *weights* connecting the i-th node to the output layer; and $\mathbf{z}_i \in \mathbb{R}^d$, $\theta_i > 0$ respectively denote the *centroid* and *width* of the kernel at the i-th node ($1 \leq i \leq N$). The kernel widths can either remain uniform across all nodes or vary individually for each node.

1.2. The universal approximation property (UAP) of RBFNNs

Soon after their introduction by Broomhead and Lowe [1] in the 1980s, RBFNNs were applied to supervised learning tasks like classification, pattern recognition, regression, and time series prediction [2, 3]. Their theoretical appeal relies on their capacity of being dense in appropriate spaces of integrable or continuous functions, which, in NNs terminology, is referred to as the *universal approximation property* (UAP). A substantial corpus of literature has been devoted to studying this property in terms of the activation function h. For instance, Park and Sandberg [4, 5] demonstrated that relatively soft conditions on h (such as being integrable with a nonzero integral, bounded, and a.e. continuous) are sufficient to guarantee this property in $L^p(\mathbb{R}^d)$ ($1 \leq p < \infty$). Later on, Liao et al. [6] established that RBFNNs can uniformly approximate any continuous function provided that h is a.e. continuous, locally essentially bounded, and not a polynomial. Moreover, for $1 \leq p < \infty$, any function in an L^p space with respect to a finite measure can be approximated by some RBFNN with an essentially bounded activation function h that is not a polynomial. For further insights on p-mean approximation capabilities of RBFNNs, see [7] and references therein. Although the nonpolynomiality of h is clearly necessary, it has also been shown to suffice for other classes of networks to achieve the UAP [8, 9].

1.3. RBFNNs of Hankel translates

The Hankel transformation, being particularly well-suited to handle radial functions, motivated Arteaga and Marrero [10] to propose and study a radial basis function (RBF) interpolation scheme where the interpolants are given by

$$u(x) = \sum_{i=1}^{n} \alpha_i (\tau_{a_i} \phi)(x) + \sum_{j=0}^{m-1} \beta_j p_{\mu,j}(x) \quad (x \in I).$$

Here, $I = (0, \infty)$, ϕ is a complex basis function on I, $\mu \geq -1/2$, and $\tau_z = \tau_{\mu,z}$ stands for the operator of Hankel translation with order μ and symbol $z \in I$, while, for $1 \leq i \leq n$ and $0 \leq j \leq m-1$, $a_i \in I$ are the interpolation nodes, $p_{\mu,j}(x) = x^{2j+\mu+1/2}$ are monomials of Müntz type, and α_i, β_j are complex coefficients.

Details on the Hankel transformation and its associated translation and convolution operators will be provided in Section 2 below, as the results in the present paper will delve into this approach in the framework of NNs. In fact, by replacing the standard translation with the Hankel translation τ_z $(z \in I)$ in (1.1), we give the next

Definition 1.1 ([11, 12]). *An* RBFNN of Hankel translates *is any real function v on I that can be expressed as*

$$v(x) = \sum_{i=1}^{N} w_i \tau_{z_i}(\lambda_{\sigma_i}\phi)(x) \quad (x \in I),$$

where ϕ is the activation function*, $N \in \mathbb{N}$ accounts for the quantity of* nodes *in the hidden layer, and $w_i \in \mathbb{R}$ stands for the* weight *from the i-th node to the output one, while $z_i, \sigma_i \in I$ represent the* centroid *and* width*, respectively, of the i-th node* $(1 \leq i \leq N)$*. Also, $(\lambda_r\phi)(t) = \phi(rt)$ $(t \in I)$ is a homothety of ratio $r \in I$.*

The class of all RBFNNs of Hankel translates will be denoted by $\mathcal{S}_1(\phi) = \mathcal{S}_{\mu,1}(\phi)$.

It should be remarked that the UAP of closely related structures (termed *RBFNNs of Delsarte translates*) was investigated by Arteaga and the author in a series of papers, beginning with [13]. By considering RBFNNs of Hankel (or Delsarte) translates, a new parameter μ is introduced, which provides the practitioner with a greater variety of manageable kernels. This might be useful in handling mathematical models built upon a class of RBFs depending on the order μ [14, 15], as network performance can be improved just by finely tuning this extra parameter, without increasing the number of centroids. Indeed, numerical and graphical examples illustrating the effect of μ in the approximation of functions can be found in [12, Section 5].

1.4. A brief glossary on function spaces

Unless otherwise stated, henceforth we let $\mu > -1/2$. The following function spaces are to be considered:

- $L^\infty_{\mu,c} = z^{\mu+1/2}L^\infty([0, c], z^{2\mu+1}dz)$ $(c \in I)$. The usual norm of this space will be denoted by $\|\cdot\|_{\mu,\infty,c}$.
- $L^\infty_{\mu,\ell}$ is the space of functions belonging to $L^\infty_{\mu,c}$ for all $c \in I$, topologized by the sequence of seminorms $\{\|\cdot\|_{\mu,\infty,n}\}_{n\in\mathbb{N}}$.
- $C_{\mu,c}$ $(c \in I)$ is the space of functions u, continuous on $(0, c]$, for which
$$\lim_{z\to 0+} z^{-\mu-1/2}u(z) \tag{1.2}$$
exists and is finite, normed by $\|\cdot\|_{\mu,\infty,c}$. The correspondence $u \mapsto z^{-\mu-1/2}u(z)$ sets up an isometric isomorphism between $C_{\mu,c}$ and the Banach space $C[0, c]$ of the functions that are continuous on the interval $[0, c]$, with the supremum norm. Therefore, $C_{\mu,c}$ is Banach, too.
- C_μ is the space of functions u, continuous on I, for which (1.2) exists and is finite. Topologized by the sequence of seminorms $\{\|\cdot\|_{\mu,\infty,n}\}_{n\in\mathbb{N}}$, C_μ becomes Fréchet.

1.5. Structure and main results

In [12], Marrero proved the following: When $\phi \in C_\mu$, the class $\mathcal{S}_1(\phi)$ is dense in C_μ if, and only if, $\phi \notin \pi_\mu$, where

$$\pi_\mu = \operatorname{span}\left\{t^{2r+\mu+1/2} : r \in \mathbb{N}_0\right\}. \tag{1.3}$$

This generalizes to RBFNNs of Hankel translates a result of Pinkus [9, Theorem 12] for standard translates. Here we aim to extend to the Hankel setting the results in [6] as well: We will show that the density of $\mathcal{S}_1(\phi)$ in C_μ (in the sense that the closure of $\mathcal{S}_1(\phi)$ as a subspace of $L^\infty_{\mu,\ell}$ contains C_μ) can be achieved under relaxed conditions on ϕ, namely, membership in $L^\infty_{\mu,\ell} \setminus \pi_\mu$ and a.e. continuity, instead of membership in C_μ.

The structure and main results of the paper are as follows: After gathering in Section 2 the basic preliminaries on the translation and convolution operators associated with the Hankel transformation, the UAP is addressed. In Section 3, we recall from [12] the UAP for the case of activation functions in C_μ (Theorem 3.2) along with an auxiliary lemma, which gets slightly improved. In Section 4, the UAP for a.e. continuous activation functions in $L^\infty_{\mu,\ell}$ is established (Theorems 4.6 and 4.7). We remark that, at any event, nonpolynomiality of the activation function in the hidden layer, understood as exclusion from the class (1.3), has a pivotal role.

2. Preliminaries: the Hankel translation and the Hankel convolution

Let $\mu \in \mathbb{R}$, let J_μ denote the well-known Bessel function of the first kind and order μ, and let $\mathcal{J}_\mu(z) = z^{1/2} J_\mu(z)$ $(z \in I)$. Whenever the involved integral exists, the Hankel transform of a function $\phi = \phi(x)$ $(x \in I)$ is typically defined as

$$(h_\mu \phi)(x) = \int_0^\infty \phi(t) \mathcal{J}_\mu(xt)\, dt \quad (x \in I).$$

Zemanian extended the Hankel transformation to spaces of distributions by adapting the ideas that led Schwartz [16] to produce a distributional theory of the Fourier transformation. In fact, the Zemanian class $\mathcal{H}_\mu$ [17, 18] of all complex functions $\phi \in C^\infty(I)$ such that

$$\nu_{\mu,r}(\phi) = \max_{0 \le k \le r} \sup_{x \in I} \left| (1 + x^2)^r (x^{-1} D)^k x^{-\mu-1/2} \phi(x) \right| < \infty \quad (r \in \mathbb{N}_0),$$

where $D = d/dx$, plays in the Hankel transformation setting the same role as the Schwartz space of rapidly decreasing functions with respect to the Fourier transformation. When $\mu \ge -1/2$, the sequence of norms $\{\nu_{\mu,r}\}_{r \in \mathbb{N}_0}$ makes $\mathcal{H}_\mu$ into a Fréchet space, and h_μ a self-isomorphism of $\mathcal{H}_\mu$. Hence, its adjoint h'_μ is also a self-isomorphism of the dual $\mathcal{H}'_\mu$ when either its weak* or strong topologies are considered.

Zemanian [19] further introduced the class $\mathcal{B}_\mu$, which plays with respect to the Hankel transformation the same role as the test space of infinitely differentiable, compactly supported functions in the context of the Fourier transformation. Given $a \in I$, the space $\mathcal{B}_{\mu,a}$ consists of all complex functions $\phi \in C^\infty(I)$ satisfying $\phi(x) = 0$ for $x > a$, and

$$\delta_{\mu,r}(\phi) = \sup_{x \in I} \left| (x^{-1} D)^r x^{-\mu-1/2} \phi(x) \right| < \infty \quad (r \in \mathbb{N}_0).$$

Topologized by means of the seminorms $\{\delta_{\mu,r}\}_{r \in \mathbb{N}_0}$, this space is Fréchet. The strict inductive limit $\mathcal{B}_\mu$ of $\{\mathcal{B}_{\mu,a}\}_{a \in I}$ is a dense subspace of $\mathcal{H}_\mu$; consequently, its dual $\mathcal{B}'_\mu$ can be viewed as a superspace of $\mathcal{H}'_\mu$.

Sousa Pinto [20] pioneered in the study of the distributional Hankel convolution, although focusing on distributions of compact support, with $\mu = 0$. Betancor and the author [21–23] subsequently extended this theory to wider distribution spaces for any $\mu > -1/2$. The definition of the Hankel #-convolution of $\varphi, \phi \in \mathcal{H}_\mu$, in the classical sense, is as follows:

$$(\varphi\#\phi)(x) = \int_0^\infty \varphi(y)\,(\tau_x\phi)(y)\,dy \quad (x \in I),$$

where

$$(\tau_x\phi)(y) = \int_0^\infty \phi(z)\,D_\mu(x,y,z)\,dz \quad (y \in I) \tag{2.1}$$

is the Hankel translate of ϕ, with symbol $x \in I$. For $x, y, z \in I$, the nonnegative function

$$\begin{aligned} D_\mu(x,y,z) &= \int_0^\infty t^{-\mu-1/2}\mathcal{J}_\mu(xt)\,\mathcal{J}_\mu(yt)\,\mathcal{J}_\mu(zt)\,dt \\ &= \begin{cases} \dfrac{[z^2-(x-y)^2]^{\mu-1/2}\,[(x+y)^2-z^2]^{\mu-1/2}}{2^{3\mu-1}\pi^{1/2}\Gamma(\mu+1/2)\,(xyz)^{\mu-1/2}}, & |x-y| < z < x+y \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

occurring in (2.1) is known as the *Delsarte kernel*. It is symmetric in its variables and satisfies the *duplication formula*

$$\int_0^\infty \mathcal{J}_\mu(zt)\,D_\mu(x,y,z)\,dz = \mathcal{J}_\mu(xt)\mathcal{J}_\mu(yt) \quad (x, y, t \in I)$$

along with the integrability property

$$\int_0^\infty D_\mu(x,y,z)\,z^{\mu+1/2}\,dz = c_\mu^{-1}(xy)^{\mu+1/2} \quad (x, y \in I), \tag{2.2}$$

where $c_\mu = 2^\mu\Gamma(\mu+1)$. In particular,

$$(\tau_x\phi)(y) = (\tau_y\phi)(x) \quad (\phi \in \mathcal{H}_\mu,\ x, y \in I).$$

Other key results include the *shifting formula*

$$h_\mu(\tau_y\phi)(x) = x^{-\mu-1/2}\mathcal{J}_\mu(xy)(h_\mu\phi)(x) \quad (\phi \in \mathcal{H}_\mu,\ x, y \in I),$$

and the *exchange formula*

$$h_\mu(\varphi\#\phi)(x) = x^{-\mu-1/2}(h_\mu\varphi)(x)(h_\mu\phi)(x) \quad (\varphi, \phi \in \mathcal{H}_\mu,\ x \in I).$$

The translation operator extends up to $\mathcal{H}'_\mu$ by transposition. Given $f \in \mathcal{H}'_\mu$ and $\phi \in \mathcal{H}_\mu$, their Hankel convolution $f\#\phi \in \mathcal{H}'_\mu$ is

$$(f\#\phi)(x) = \langle f, \tau_x\phi\rangle \quad (x \in I) \ \text{[23, Definition 3.1]}.$$

The shifting and exchange formulas

$$h'_\mu(\tau_y f)(x) = x^{-\mu-1/2}\mathcal{J}_\mu(xy)(h'_\mu f)(x)$$

and

$$h'_\mu(f\#\phi)(x) = x^{-\mu-1/2}(h_\mu\phi)(x)(h'_\mu f)(x)$$

are valid in the distributional sense (cf. [23, Proposition 3.5]). The interested reader is especially referred to [18, 21–23] for a more extensive study of the generalized Hankel transformation and its associated translation and convolution.

3. Uniform approximation with continuous activation functions

Except for the a.e. pointwise convergence stated in part (i), the next lemma is contained in [12, Lemma 2.1].

Lemma 3.1. *For $z \in I$ and $\phi \in L^\infty_{\mu,\ell}$, let $\tau_z\phi$ be as in (2.1), and define*

$$(T_z\phi)(x) = \phi_z(x) = c_\mu z^{-\mu-1/2}(\tau_z\phi)(x) \quad (x \in I).$$

Then, the following holds:

(i) The function $x \mapsto (\tau_z\phi)(x)$ is well defined and continuous on I. Both operators T_z and τ_z are linear and continuous from $L^\infty_{\mu,\ell}$ into itself. If, moreover, ϕ is a.e. continuous, then $\lim_{z\to 0+}\phi_z(x) = \phi(x)$ *a.e. $x \in I$.*

(ii) When restricted to C_μ, both T_z and τ_z define continuous linear operators into C_μ. Also, if $\phi \in C_\mu$, then $\lim_{z\to 0+}\phi_z = \phi$ *in C_μ.*

Proof. As said above, it only remains to show that $\lim_{z\to 0+}\phi_z(x) = \phi(x)$ a.e. $x \in I$ whenever $\phi \in L^\infty_{\mu,\ell}$ is a.e. continuous, that is, the measure of the set of its discontinuity points is null.

Assume $x \in I$ is a continuity point of ϕ; then, given any $\varepsilon > 0$, for some $\delta = \delta(x,\varepsilon) > 0$, the conditions $t \in I$ and $|t-x| < \delta$ imply

$$\left|t^{-\mu-1/2}\phi(t) - x^{-\mu-1/2}\phi(x)\right| < \varepsilon.$$

Furthermore, if $0 < z < \delta$ and $t \in I$ with $|t-x| \geq \delta > z$, then $D_\mu(x,z,t) = 0$. Thus, using (2.2), we may write

$$\begin{aligned}
&\left|x^{-\mu-1/2}\phi_z(x) - x^{-\mu-1/2}\phi(x)\right| \\
&\quad = \left|c_\mu(xz)^{-\mu-1/2}(\tau_z\phi)(x) - x^{-\mu-1/2}\phi(x)\right| \\
&\quad = \left|c_\mu(xz)^{-\mu-1/2}\int_0^\infty \phi(t)\,D_\mu(x,z,t)\,dt - c_\mu(xz)^{-\mu-1/2}x^{-\mu-1/2}\phi(x)\int_0^\infty D_\mu(x,z,t)\,t^{\mu+1/2}dt\right| \\
&\quad \leq c_\mu(xz)^{-\mu-1/2}\int_{|t-x|<\delta}\left|t^{-\mu-1/2}\phi(t) - x^{-\mu-1/2}\phi(x)\right| D_\mu(x,z,t)\,t^{\mu+1/2}\,dt \\
&\quad < \varepsilon \quad (0 < z < \delta),
\end{aligned}$$

which settles the lemma. □

We end this section with a main result from [12] and some comments about its proof.

Theorem 3.2 ([12, Theorem 3.3])**.** *Let $\phi \in C_\mu \setminus \pi_\mu$. Then, $\mathcal{S}_1(\phi) = \operatorname{span}\{\tau_s(\lambda_r\phi) : s, r \in I\} \subset C_\mu$ is dense in C_μ, i.e., for any $f \in C_\mu$, $c \in I$ and $\varepsilon > 0$, some $g \in \mathcal{S}_1(\phi)$ satisfies $\|f-g\|_{\mu,\infty,c} < \varepsilon$.*

Conversely, if $\phi \in \pi_\mu$, then $\mathcal{S}_1(\phi)$ has finite dimension, which prevents it from being dense in C_μ.

Proof. The description of $\mathcal{S}_1(\phi)$ is clear. A proof of the converse part was given in [12, Theorem 2.5]; however, we include it here for completeness. Let

$$S_\mu = x^{-\mu-1/2}Dx^{2\mu+1}Dx^{-\mu-1/2}$$

denote the Bessel differential operator of order μ. Given $m \in \mathbb{N}_0$, a distribution $f \in \mathcal{H}'_\mu$ solves the differential equation $S_\mu^{m+1} f = 0$ if, and only if, $f \in \pi_\mu$ and the degree of the even polynomial $t^{-\mu-1/2} f(t)$ is not greater than $2m$ [10, Theorem 2.19]. Assume $\phi \in \pi_\mu$ and $z^{-\mu-1/2}\phi(z)$ has degree $2m$, so that $S_\mu^{m+1}\phi = 0$. The commutativity of S_μ with Hankel translations (cf. [24]), followed by a simple computation, yields

$$S_\mu^{m+1}[\tau_s(\lambda_r\phi)] = r^{2(m+1)}\tau_s[\lambda_r(S_\mu^{m+1}\phi)] = 0 \quad (s, r \in I).$$

This means that the dimension of the linear space $\mathcal{S}_1(\phi)$ is at most $2m$. Being finite-dimensional and hence closed, $\mathcal{S}_1(\phi)$ cannot be dense in infinite-dimensional spaces. □

4. Uniform approximation with locally essentially bounded, a.e. continuous activation functions

In this section, a series of lemmas will lead us to our main result. We begin with the following basic fact.

Lemma 4.1. *Assume $A \subset X_\mu$, where $X_\mu = L^\infty_{\mu,\ell}$ or $X_\mu = C_\mu$, and let $\overline{A}$, respectively $\overline{A}^c$, denote the closure of A in the topology of X_μ, respectively in the norm of $X_{\mu,c}$, where, for any $c \in I$, $X_{\mu,c} = L^\infty_{\mu,c}$ or $X_{\mu,c} = C_{\mu,c}$. Then,*

$$\overline{A} = \bigcap_{c\in I} \overline{A}^c.$$

Proof. The inclusion map $X_\mu \hookrightarrow X_{\mu,c}$ being continuous, it is evident that $\overline{A} \subset \overline{A}^c$ for all $c \in I$.

Conversely, suppose $g \in \overline{A}^c$ whenever $c \in I$. Then, in particular, for every $n \in \mathbb{N}$, there exists $g_n \in A$ such that $\|g - g_n\|_{\mu,\infty,n} < n^{-1}$. Given $b \in I$ and $\varepsilon > 0$, choose $m \in \mathbb{N}$ with $m \geq \max\{b, \varepsilon^{-1}\}$. We have

$$\begin{aligned}\|g - g_n\|_{\mu,\infty,b} &\leq \|g - g_n\|_{\mu,\infty,m} \\ &\leq \|g - g_n\|_{\mu,\infty,n} < \frac{1}{n} \leq \frac{1}{m} \leq \varepsilon \quad (n \geq m).\end{aligned}$$

The arbitrariness of $b \in I$ shows that $\lim_{n\to\infty} g_n = g$ in the topology of X_μ, so that $g \in \overline{A}$. □

Lemma 4.2. *Let $\sigma \in L^\infty_{\mu,\ell}$ be a.e. continuous, and let $b, c \in I$. Then, given $\rho \in \mathcal{B}_{\mu,b}$, the convolution*

$$(\sigma\#\rho)(x) = \int_0^\infty (\tau_x\sigma)(t)\rho(t)\,dt \quad (x \in I) \tag{4.1}$$

lies in $C_{\mu,c}$ and can be approximated from $\operatorname{span}\{\tau_s\sigma : s \in I\}$ *in the norm of $L^\infty_{\mu,c}$. In other words, for any $\rho \in \mathcal{B}_\mu$ we have that $\sigma\#\rho$ lies in C_μ and belongs to the closure of* $\operatorname{span}\{\tau_s\sigma : s \in I\}$ *in $L^\infty_{\mu,\ell}$.*

Proof. It can be adapted from that of [12, Lemma 3.1]. Fix $\rho \in \mathcal{B}_{\mu,b}$. By virtue of Lemma 3.1(i), $\tau_x\sigma \in L^\infty_{\mu,\ell}$ for each $x \in I$; consequently, the function (4.1) is well defined.

We begin by showing the continuity of $\sigma\#\rho$ on $(0, c]$. With this purpose, pick $x_0 \in (0, c]$. We have

$$\begin{aligned}&|(\sigma\#\rho)(x) - (\sigma\#\rho)(x_0)| \\ &\quad\leq \int_0^\infty \left|(\tau_x\sigma)(z) - (\tau_{x_0}\sigma)(z)\right| \left|\rho(z)\right| dz\end{aligned}$$

$$\leq b^{\mu+1/2}\int_0^b \left|(\tau_x\sigma)(z)-(\tau_{x_0}\sigma)(z)\right|\,\left|z^{-\mu-1/2}\rho(z)\right|dz$$
$$\leq b^{\mu+1/2}\sup_{z\in I}\left|z^{-\mu-1/2}\rho(z)\right|\int_0^b \left|(\tau_x\sigma)(z)-(\tau_{x_0}\sigma)(z)\right|dz \quad (x\in(0,c]).$$

Moreover, for each $z\in(0,b]$, using (2.2) we may write

$$\begin{aligned}
&\left|(\tau_x\sigma)(z)-(\tau_{x_0}\sigma)(z)\right|\\
&\quad\leq \operatorname*{ess\,sup}_{t\in[0,b+c]}\left|t^{-\mu-1/2}\sigma(t)\right|\int_0^{b+c}\left|D_\mu(x,z,t)-D_\mu(x_0,z,t)\right|t^{\mu+1/2}dt\\
&\quad\leq c_\mu^{-1}z^{\mu+1/2}\left(x^{\mu+1/2}+x_0^{\mu+1/2}\right)\operatorname*{ess\,sup}_{t\in[0,b+c]}\left|t^{-\mu-1/2}\sigma(t)\right|\\
&\quad\leq 2\,c_\mu^{-1}(bc)^{\mu+1/2}\operatorname*{ess\,sup}_{t\in[0,b+c]}\left|t^{-\mu-1/2}\sigma(t)\right| \quad (x\in(0,c]).
\end{aligned}$$

Lemma 3.1(i) guarantees that

$$\lim_{x\to x_0}\left|(\tau_x\sigma)(z)-(\tau_{x_0}\sigma)(z)\right|=0 \quad (z\in(0,b])\,.$$

The desired continuity now follows from an application of the Lebesgue theorem of dominated convergence.

Similarly, because of Lemma 3.1(i), the estimate

$$\begin{aligned}
&\left|c_\mu x^{-\mu-1/2}(\sigma\#\rho)(x)-\int_0^\infty\sigma(z)\rho(z)\,dz\right|\\
&\quad=\left|\int_0^b c_\mu x^{-\mu-1/2}(\tau_x\sigma)(z)\rho(z)\,dz-\int_0^b\sigma(z)\rho(z)\,dz\right|\\
&\quad\leq\int_0^b\left|c_\mu(xz)^{-\mu-1/2}(\tau_x\sigma)(z)-z^{-\mu-1/2}\sigma(z)\right|\,\left|\rho(z)\right|z^{\mu+1/2}dz\\
&\quad=\int_0^b\left|z^{-\mu-1/2}\sigma_x(z)-z^{-\mu-1/2}\sigma(z)\right|\,\left|z^{-\mu-1/2}\rho(z)\right|z^{2\mu+1}dz\\
&\quad\leq\sup_{z\in I}\left|z^{-\mu-1/2}\rho(z)\right|\int_0^b\left|z^{-\mu-1/2}\sigma_x(z)-z^{-\mu-1/2}\sigma(z)\right|\,z^{2\mu+1}dz \quad (x\in I),
\end{aligned}$$

and dominated convergence:

$$\begin{aligned}
&\left|z^{-\mu-1/2}\sigma_x(z)-z^{-\mu-1/2}\sigma(z)\right|\\
&\quad\leq\left|z^{-\mu-1/2}\sigma_x(z)\right|+\left|z^{-\mu-1/2}\sigma(z)\right|\\
&\quad\leq\left|c_\mu(xz)^{-\mu-1/2}\int_0^{b+x}D(x,z,t)\sigma(t)\,dt\right|+\left|z^{-\mu-1/2}\sigma(z)\right|\\
&\quad\leq 2\operatorname*{ess\,sup}_{t\in[0,b+c]}\left|t^{-\mu-1/2}\sigma(t)\right| \quad (x\in(0,c],\ z\in(0,b])\,,
\end{aligned}$$

we arrive at

$$\lim_{x\to0+}x^{-\mu-1/2}(\sigma\#\rho)(x)=c_\mu^{-1}\int_0^\infty\sigma(z)\rho(z)\,dz.$$

Thus, $\sigma\#\rho \in C_{\mu,c}$.

Next, fix $x \in (0, c]$. For each $n \in \mathbb{N}$, consider the partition $\{t_i = ib/n : 0 \le i \le n\}$ of $[0, b]$, and let $\varepsilon > 0$. The following estimate is easily obtained:

$$\begin{aligned}
&\left|(\sigma\#\rho)(x) - \sum_{i=1}^{n} \frac{b\rho(t_i)}{n} (\tau_{t_i}\sigma)(x)\right| \\
&\quad \le \left|\int_0^\infty (\tau_x\sigma)(t)\,\rho(t)\,dt - \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} t_i^{-\mu-1/2} (\tau_x\sigma)(t_i)\, t^{\mu+1/2}\rho(t)\,dt\right| \\
&\qquad + \left|\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} t_i^{-\mu-1/2} (\tau_x\sigma)(t_i)\, t^{\mu+1/2}\rho(t)\,dt - \frac{b}{n}\sum_{i=1}^{n} (\tau_x\sigma)(t_i)\,\rho(t_i)\right|. \qquad (4.2)
\end{aligned}$$

As $z^{2\mu+1}$ and $z^{-\mu-1/2}\rho(z)$ are uniformly continuous on $[0, b]$ (cf. [18, Lemma 5.2-1]), for large enough n, the second term on the right-hand side of (4.2) can be bounded by

$$\begin{aligned}
&\left|\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} t_i^{-\mu-1/2} (\tau_x\sigma)(t_i)\, t^{\mu+1/2}\rho(t)\,dt - \frac{b}{n}\sum_{i=1}^{n} (\tau_x\sigma)(t_i)\,\rho(t_i)\right| \\
&\le x^{\mu+1/2} c_\mu^{-1} \operatorname*{ess\,sup}_{z\in[0,b+c]} \left|z^{-\mu-1/2}\sigma(z)\right| \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \left|t^{\mu+1/2}\rho(t) - t_i^{\mu+1/2}\rho(t_i)\right| dt \\
&\le x^{\mu+1/2} c_\mu^{-1} \operatorname*{ess\,sup}_{z\in[0,b+c]} \left|z^{-\mu-1/2}\sigma(z)\right| \\
&\quad \times \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \left[\sup_{t\in I}\left|t^{-\mu-1/2}\rho(t)\right|\left|t^{2\mu+1} - t_i^{2\mu+1}\right| + \left|t^{-\mu-1/2}\rho(t) - t_i^{-\mu-1/2}\rho(t_i)\right| t_i^{2\mu+1}\right] dt \\
&< x^{\mu+1/2}\,\frac{\varepsilon}{2}. \qquad (4.3)
\end{aligned}$$

Concerning the first term on the right-hand side of (4.2), recall that σ is a.e. continuous and note that the representation (2.1), jointly with Lemma 3.1, renders the map $(x, t) \mapsto (xt)^{-\mu-1/2}(\tau_x\sigma)(t)$ continuous on $(I \setminus U) \times [0, \infty)$, where U is some open set containing the points of discontinuity of σ, with measure less than a given $\lambda > 0$. Therefore, this map is uniformly continuous over compacta: To every $\alpha, \beta > 0$, there corresponds $N \in \mathbb{N}$, independent of $x \in [\alpha, c] \setminus U$, such that $n \ge N$ implies

$$\left|(xt)^{-\mu-1/2}(\tau_x\sigma)(t) - (xt_i)^{-\mu-1/2}(\tau_x\sigma)(t_i)\right| < \beta \quad (t \in [t_{i-1}, t_i],\ 1 \le i \le n).$$

In particular, given $\alpha, \eta > 0$, we may arrange for

$$\begin{aligned}
&\left|\int_0^\infty (\tau_x\sigma)(t)\rho(t)\,dt - \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} t_i^{-\mu-1/2}(\tau_x\sigma)(t_i)\, t^{\mu+1/2}\rho(t)\,dt\right| \\
&\le \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \left|t^{-\mu-1/2}(\tau_x\sigma)(t) - t_i^{-\mu-1/2}(\tau_x\sigma)(t_i)\right| \left|t^{-\mu-1/2}\rho(t)\right| t^{2\mu+1} dt \\
&\le x^{\mu+1/2}\sup_{t\in I}\left|t^{-\mu-1/2}\rho(t)\right| \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \left|(xt)^{-\mu-1/2}(\tau_x\sigma)(t) - (xt_i)^{-\mu-1/2}(\tau_x\sigma)(t_i)\right| t^{2\mu+1} dt \\
&< x^{\mu+1/2}\,\eta \quad (x \in [\alpha, c] \setminus U), \qquad (4.4)
\end{aligned}$$

provided that n is large enough. This way, given $\eta, \delta > 0$, there exists $N \in \mathbb{N}$ such that, whenever $n \geq N$, the measure of the set of points $x \in (0, c]$ for which the left-hand side of (4.4), weighted by $x^{-\mu-1/2}$, is greater than or equal to η, does not exceed δ; that is, the sequence of such measures converges to zero, or, in other words, the corresponding functional sequence converges to zero in measure. By passing to a subsequence if necessary, a.e. convergence is achieved; thus, we obtain

$$\left| \int_0^\infty (\tau_x\sigma)(t)\rho(t)\,dt - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} t_i^{-\mu-1/2} (\tau_x\sigma)(t_i)\, t^{\mu+1/2}\rho(t)\,dt \right| < x^{\mu+1/2}\,\frac{\varepsilon}{2} \tag{4.5}$$

for a.e. $x \in [0, c]$ and sufficiently large n. A combination of (4.2), (4.3), and (4.5) results in the estimate

$$\left\| \sigma\#\rho - \sum_{i=1}^n \frac{b\rho(t_i)}{n}\,\tau_{t_i}\sigma \right\|_{\mu,\infty,c} = \operatorname*{ess\,sup}_{x\in[0,c]} \left| x^{-\mu-1/2}(\sigma\#\rho)(x) - x^{-\mu-1/2} \sum_{i=1}^n \frac{b\rho(t_i)}{n}\,(\tau_{t_i}\sigma)(x) \right| < \varepsilon$$

being valid for large n, which accomplishes the first part of the proof.

Now, for any $\rho \in \mathcal{B}_\mu$, we have that $\sigma\#\rho \in C_\mu$ lies in the closure of span $\{\tau_s\sigma : s \in I\}$ in $L^\infty_{\mu,c}$ whenever $c \in I$. Since, by Lemma 3.1(i), span $\{\tau_s\sigma : s \in I\} \subset L^\infty_{\mu,\ell}$, a direct application of Lemma 4.1 reveals that $\sigma\#\rho$ belongs to the closure of span $\{\tau_s\sigma : s \in I\}$ in $L^\infty_{\mu,\ell}$. The proof is complete. □

Remark 4.3. *Observe that, in the notation and conditions of Lemma 4.2, both*

$$\left\{ \sum_{i=1}^n \frac{b\rho(t_i)}{n}\,\tau_{t_i}\sigma \right\}_{n\in\mathbb{N}}$$

and

$$\left\{ \sum_{i=1}^n \left[t_i^{-\mu-1/2} \int_{t_{i-1}}^{t_i} t^{\mu+1/2}\rho(t)\,dt \right] \tau_{t_i}\sigma \right\}_{n\in\mathbb{N}}$$

are approximating sequences to $\sigma\#\rho$ from span $\{\tau_s\sigma : s \in I\}$.

Lemma 4.4. *Assume $\sigma \in L^\infty_{\mu,\ell}$ is a.e. continuous and does not lie in π_μ. Then, some $\rho \in \mathcal{B}_\mu$ is such that $\sigma\#\rho$ does not lie in π_μ, either.*

Proof. Lemma 4.2 allows us to argue as in the proof of [12, Lemma 3.2]. □

Lemma 4.5. *If $\sigma \in L^\infty_{\mu,\ell}$, $\rho \in \mathcal{B}_\mu$ and $a \in I$, then $\tau_a(\sigma\#\rho) = \sigma\#\tau_a\rho$.*

Proof. Defined as in (4.1), the convolution $\sigma\#\tau_a\rho$ makes sense, because $\mathcal{B}_\mu$ is stable under Hankel translations [21, Corollary 3.3].

Let $b \in I$ be such that $\rho(t) = 0$ for $t > b$. There holds:

$$\begin{aligned}
&\int_0^\infty D_\mu(a, x, z)\,dz \int_0^\infty |\rho(s)|\,ds \int_0^\infty |\sigma(t)|\,D_\mu(z, s, t)\,dt \\
&\leq \int_0^{x+a} D_\mu(a, x, z)\,dz \int_0^b |\rho(s)|\,ds \int_0^{x+a+b} |\sigma(t)|\,D_\mu(z, s, t)\,dt \\
&\leq \sup_{s\in I} \left| s^{-\mu-1/2}\rho(s) \right| \int_0^\infty D_\mu(a, x, z)\,dz \int_0^{x+a+b} |\sigma(t)|\,dt \int_0^\infty D_\mu(z, s, t)\, s^{\mu+1/2} ds
\end{aligned}$$

$$= c_\mu^{-1} \sup_{s\in I} \left|s^{-\mu-1/2}\rho(s)\right| \int_0^\infty D_\mu(a,x,z)\, z^{\mu+1/2} dz \int_0^{x+a+b} \left|t^{-\mu-1/2}\sigma(t)\right| t^{2\mu+1} dt$$
$$\leq c_\mu^{-2}(ax)^{\mu+1/2} \operatorname*{ess\,sup}_{t\in[0,x+a+b]} \left|t^{-\mu-1/2}\sigma(t)\right| \sup_{s\in I} \left|s^{-\mu-1/2}\rho(s)\right| \int_0^{x+a+b} t^{2\mu+1} dt < \infty \quad (x \in I).$$

Thus, the Fubini theorem may be applied to obtain

$$\begin{aligned}
\tau_a(\sigma\#\rho)(x) &= \int_0^\infty (\sigma\#\rho)(z)\, D_\mu(a,x,z)\, dz \\
&= \int_0^\infty D_\mu(a,x,z)\, dz \int_0^\infty \rho(s)\, ds \int_0^\infty \sigma(t)\, D_\mu(z,s,t)\, dt \\
&= \int_0^\infty \sigma(t)\, dt \int_0^\infty \rho(s)\, ds \int_0^\infty D_\mu(a,x,z)\, D_\mu(z,s,t)\, dz \\
&= \int_0^\infty \sigma(t)\, dt \int_0^\infty \rho(s)\, ds \int_0^\infty D_\mu(a,z,s)\, D_\mu(x,z,t)\, dz \\
&= \int_0^\infty dz \int_0^\infty \sigma(t)\, D_\mu(x,z,t)\, dt \int_0^\infty \rho(s)\, D_\mu(a,z,s)\, ds \\
&= \int_0^\infty (\tau_x\sigma)(z)\, (\tau_a\rho)(z)\, dz = (\sigma\#\tau_a\rho)(x) \quad (x \in I),
\end{aligned}$$

as claimed. □

Theorem 4.6. *Let $\sigma \in L^\infty_{\mu,\ell} \setminus \pi_\mu$ be a.e. continuous. Then,*

$$\mathcal{S}_1(\sigma) = \operatorname{span}\{\tau_s(\lambda_r\sigma) : s, r \in I\} \subset L^\infty_{\mu,\ell}$$

is dense in C_μ, i.e., for any $f \in C_\mu$, $c \in I$ and $\varepsilon > 0$, some $g \in \mathcal{S}_1(\sigma)$ satisfies $\|f - g\|_{\mu,\infty,c} < \varepsilon$.

Conversely, if $\sigma \in \pi_\mu$, then $\mathcal{S}_1(\sigma)$ has finite dimension, which prevents it from being dense in C_μ.

Proof. The converse statement is contained in Theorem 3.2.

For the direct one, use Lemmas 4.2 and 4.4 to get some $\rho \in \mathcal{B}_\mu$ such that $\sigma\#\rho \in C_\mu \setminus \pi_\mu$. The identity

$$\lambda_r(\tau_q\sigma) = r^{\mu+1/2}\tau_{q/r}(\lambda_r\sigma) \quad (r, q \in I) \tag{4.6}$$

can be derived by simple changes of variables. A combination of Theorem 3.2 with (4.6) and Lemma 4.5 yields the density of

$$\mathcal{S}_1(\sigma\#\rho) = \operatorname{span}\{\lambda_r(\sigma\#\tau_q\rho) : r, q \in I\}$$

in C_μ. Recalling that $\mathcal{B}_\mu$ is stable under Hankel translations, invoke Lemma 4.2 again, this time to approximate $\sigma\#\tau_q\rho$ from $\operatorname{span}\{\tau_s\sigma : s \in I\}$ in the topology of $L^\infty_{\mu,\ell}$. After a new application of (4.6), we are done. □

As a consequence of Theorem 4.6, the hypotheses imposed on the activation function in [12, Theorem 4.1] can be weakened.

Theorem 4.7. *Let $\sigma \in L^{\infty}_{\mu,\ell}$ be a.e. continuous, and let $1 \leq p < \infty$. Given $c \in I$, let γ be a Radon measure on $[0, c]$ satisfying*

$$\int_0^c t^{\mu+1/2} d|\gamma|(t) < \infty.$$

Then, for $\mathcal{S}_1(\sigma) = \operatorname{span}\{\tau_s(\lambda_r\sigma) : s, r \in I\}$ to be dense in $L^p([0, c], d\gamma)$, it is necessary and sufficient that $\sigma \notin \pi_\mu$.

Proof. If $\sigma \in \pi_\mu$ then, as shown above, $\mathcal{S}_1(\sigma)$ has finite dimension, which prevents it from being dense in $L^p([0, c], d\gamma)$.

Conversely, if $\sigma \notin \pi_\mu$ then, from Theorem 4.6, $\mathcal{S}_1(\sigma)$ is dense in $C_{\mu,c}$, and hence in $L^p([0, c], d\gamma)$. □

5. Conclusions

The universal approximation property (UAP) of three-layered radial basis function neural networks of Hankel translates with varying widths has been studied. The requirement on the activation function σ in the hidden layer for such networks to approximate continuous functions locally in the esssup-norm has been satisfactorily weakened from continuity to local essential boundedness and a.e. continuity, provided that $z^{-\mu-1/2}\sigma(z)$ ($z \in I$) is not an even polynomial. The UAP in p-mean ($1 \leq p < \infty$) with respect to a suitable finite measure can therefore be attained under the same relaxed condition.

Use of Generative-AI tools declaration

The author declares she has not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The author wants to express her gratitude to the anonymous reviewers for valuable comments that helped improve the presentation of the paper.

Conflict of interest

There is no conflict of interest to disclose.

References

1. D. S. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks, *Complex Syst.*, **2** (1988), 321–355.

2. R. P. Lippmann, Pattern classification using neural networks, *IEEE Commun. Mag.*, **27** (1989), 47–64. https://doi.org/10.1109/35.41401

3. S. Renals, R. Rohwer, Phoneme classification experiments using radial basis functions, *International 1989 Joint Conference on Neural Networks*, Washington DC (USA), 1989, 461–467. https://doi.org/10.1109/IJCNN.1989.118620

4. J. Park, I. W. Sandberg, Universal approximation using Radial-Basis-Function networks, *Neural Comput.*, **3** (1991), 246–257. https://doi.org/10.1162/neco.1991.3.2.246

5. J. Park, I. W. Sandberg, Approximation and radial-basis-function networks, *Neural Comput.*, **5** (1993), 305–316. https://doi.org/10.1162/neco.1993.5.2.305

6. Y. Liao, S. C. Fang, H. L. W. Nuttle, Relaxed conditions for radial-basis function networks to be universal approximators, *Neural Netw.*, **16** (2003), 1019–1028. https://doi.org/10.1016/S0893-6080(02)00227-7

7. D. Nan, W. Wu, J. L. Long, Y. M. Ma, L. J. Sun, L^p approximation capability of RBF neural networks, *Acta Math. Sin.-Engl. Ser.*, **24** (2008), 1533–1540. https://doi.org/10.1007/s10114-008-6423-x

8. M. Leshno, V. Y. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Netw.*, **6** (1993), 861–867. https://doi.org/10.1016/S0893-6080(05)80131-5

9. A. Pinkus, TDI-subspaces of $C(\mathbb{R}^d)$ and some density problems from neural networks, *J. Approx. Theory*, **85** (1996), 269–287. https://doi.org/10.1006/jath.1996.0042

10. C. Arteaga, I. Marrero, A scheme for interpolation by Hankel translates of a basis function, *J. Approx. Theory*, **164** (2012), 1540–1576. https://doi.org/10.1016/j.jat.2012.08.005

11. I. Marrero, The role of nonpolynomiality in uniform approximation by RBF networks of Hankel translates, *J. Funct. Spaces*, **2019** (2019), 1845491 https://doi.org/10.1155/2019/1845491

12. I. Marrero, Radial basis function neural networks of Hankel translates as universal approximators, *Anal. Appl. (Singap.)*, **17** (2019), 897–930. https://doi.org/10.1142/S0219530519500064

13. C. Arteaga, I. Marrero, Universal approximation by radial basis function networks of Delsarte translates, *Neural Netw.*, **46** (2013), 299–305. https://doi.org/10.1016/j.neunet.2013.06.011

14. H. Corrada, K. Lee, B. Klein, R. Klein, S. Iyengar, G. Wahba, Examining the relative influence of familial, genetic, and environmental covariate information in flexible risk models, *Proc. Natl. Acad. Sci. USA*, **106** (2009), 8128–8133. https://doi.org/10.1073/pnas.0902906106

15. S. Hamzehei Javaran, N. Khaji, A. Noorzad, First kind Bessel function (J-Bessel) as radial basis function for plane dynamic analysis using dual reciprocity boundary element method, *Acta Mech.*, **218** (2011), 247–258. https://doi.org/10.1007/s00707-010-0421-7

16. L. Schwartz, *Théorie des distributions*, Vols. I, II, Publications de l'Institut de Mathématique de l'Université de Strasbourg, Paris: Hermann & Cie, 1950-1951.

17. A. H. Zemanian, A distributional Hankel transformation, *SIAM J. Appl. Math.*, **14** (1966), 561–576. https://doi.org/10.1137/0114049

18. A. H. Zemanian, *Generalized integral transformations*, Pure and Applied Mathematics, Vol. 18, New York: John Wiley & Sons, 1968.

19. A. H. Zemanian, The Hankel transformation of certain distributions of rapid growth, *SIAM J. Appl. Math.*, **14** (1966), 678–690. https://doi.org/10.1137/0114056

20. J. de Sousa Pinto, A generalised Hankel convolution, *SIAM J. Math. Anal.*, **16** (1985), 1335–1346. https://doi.org/10.1137/0516097

21. J. J. Betancor, I. Marrero, The Hankel convolution and the Zemanian spaces $\mathcal{B}_\mu$ and $\mathcal{B}'_\mu$, *Math. Nachr.*, **160** (1993), 277–298. https://doi.org/10.1002/mana.3211600113

22. J. Betancor, I. Marrero, Structure and convergence in certain spaces of distributions and the generalized Hankel convolution, *Math. Japon.*, **38** (1993), 1141–1155.

23. I. Marrero, J. J. Betancor, Hankel convolution of generalized functions, *Rend. Mat. Ser. VII*, **15** (1995), 351–380.

24. J. J. Betancor, A new characterization of the bounded operators commuting with Hankel translation, *Arch. Math.*, **69** (1997), 403–408. https://doi.org/10.1007/s000130050138