Asia Pac. J. Math. 2026 **13**:145

APPROXIMATION SPACES OF NEURAL NETWORKS WITH DIRECT AND INVERSE THEOREMS

MOHAMMED MAJID NEJM*, HAWRAA ABBAS ALMURIEB

Department of Mathematics, College of Education for Pure Sciences, University of Babylon, Hillah, Iraq

*Corresponding author: bsclec.mohammed.majid@uobabylon.edu.iq

Received Apr. 25, 2026

Abstract. Direct and inverse are fundamental theorems in approximation theory, explaining the relationship between the smoothness of a function and the accuracy of its approximation in functional spaces. In this study, multi-channel convolution and the tensor product with the adaptive piecewise linear activation (APL) are used to build the neural network. On the other hand, B-spline functions are important both as activation functions and as fundamental elements of the hidden layer in neural network construction. They provide the model with flexibility in controlling the degree of smoothness across the spline order while ensuring the regularity of the derivatives up to a certain order is maintained. This strikes a balance between power and generalization, providing a flexible and numerically stable approximation basis for the neural network. Therefore, the concept of generalized $B-$spline APL neural networks is introduced, relying on the APL and B-spline functions. Finally, this study proves the direct and inverse theorems, thus providing a mathematical explanation for the efficiency and capability of neural networks within approximation theory, thereby the proof revealing the link between functional analysis and contemporary neural network architecture.

2020 Mathematics Subject Classification. 41A25; 41A27; 46A16; 68T07.

Key words and phrases. approximation space; quasi-banach space; modulus of smoothness; neural networks; direct theorem; inverse theorem.

1. Introduction

Approximation spaces are important for neural networks because they provide a fundamental mathematical framework that helps us to understand the theoretical basis of their design. They help demonstrate the ability of neural networks (NNs) to represent complex functions and operate with high efficiency and accuracy. With the rapid development of applied mathematics, these spaces have been studied in numerous fields. Researchers are continuing to develop new approaches to understanding NNs through functional approximation, both theoretically and practically. The concept of NNs first emerged in the 1940s as an attempt to mimic how the human brain works. In 1943, McCulloch and

DOI: 10.28924/APJM/13-145

Pitts presented a simplified mathematical model of a biological neuron within the context of artificial intelligence. They demonstrated that NNs can calculate any logical function [1]. This work formed the starting point for subsequent research aimed at establishing the theoretical principles underlying the approximation capabilities of these networks. In the framework of developing approximation theory, Pietsch introduced the concept of approximation number in 1981 and discussed mechanisms for including approximation spaces within larger Banach spaces. This expanded the mathematical framework associated with approximation operations [2]. In 1993, DeVore and Lorentz presented a constructivist approach to approximation, focusing on functions with one real variable and achieving notable results by studying approximation classes and their properties [3]. The most significant shift in linking approximation theory to NNs occurred in 1988, when Cybenko introduced the main Universal Approximation Theorem (UAT), which demonstrates the ability of NNs to approximate a wide range of functions under specific conditions [4]. In 1989, Funahashi validated the continuous function approximation realization theorem of Rumelhart Hinton-Williams using multilayer NNs with a sigmoidal activation function. The Kolmogorov-Arnold-Speizer theorem was also used to demonstrate the case of two hidden layers [5]. In 1995, the research of Ditzian, Hristov, and Ivanov was considered an important study in approximation theory, because it addressed the relationship between modules of smoothness and K-functionals within L_p spaces when $0 < p < 1$, which is a more complex mathematical case than the traditional case of $p \geq 1$ [6].The direct theorem and its inverse are fundamental mathematical results in approximation theory. These theories are important because they demonstrate of the approximation capabilities of NNs and provide a theoretical framework for estimating the efficiency and complexity of models. This helps us to understand how many parameters are required to achieve a specific level of accuracy, thus supporting the optimal design of NNs architectures and the analysis of their performance. George A. Anastassiou has focused on this subject since 1997. The convergence rate of a NNs was measured by its operators, namely the squashing and Cardaliaguet operators. The value of the Jackson inequality is often equivalent to the modulus of continuity in most of his researches, as in [7]. Splines are considered an alternative to traditional approximation methods where it is particularly important in practical image processing applications. In 1999, Michael Unser presented a study on splines, demonstrating the fundamental properties that make them suitable for reducing computational complexity and facilitating their application in image processing [8].In 2003, Almira and Uwe presented a study on generalized approximation spaces and their applications, focusing on extending classical theory to include broader spaces defined by the best degree of approximation. These spaces were then generalized to describe relatively compact sets within Banach spaces [9]. In 2014, Chenglong, et al. conducted an approximation analysis of convolutional NNs, emphasizing the important of convolutional structures in enhancing the representational capacity of networks [10]. In

2014, Agostinelli, et al. introduced a new methodology for improving the performance of NNs by learning activation functions rather than pre-fixing them. Their study relied on developing adaptive sectoral linear activation functions that are independently trained using a downward gradient algorithm. This allowed for the replacement of traditional functions like ReLU with more flexible ones, contributing to improved efficiency and performance of deep models [11]. In 2020, Gribonval et al. presented a conceptualization of approximation spaces in deep NNs. They constructed an approximation space based on clusters of networks, taking into account the complexity of connections and neurons [12]. In 2021 DeVore, et al. studied the approximation properties of NNs outputs. They highlighted features that are not found in traditional numerical analysis approximation methods, particularly when using the ReLU activation function. This function produces multi-defined linear functions on complex partitions [13]. Also in the same year, Zuwei Chen, et al. investigated the approximation capability of NNs, demonstrating that the presence of three hidden layers can suffice to achieve highly efficient approximation [14]. Again, in the same year, Elbrächter, Dennis, et al. developed fundamental limits for deep NNs learning by defining what can be achieved if no constraints are imposed on the learning algorithms or training data size [15]. In 2021, Adcock and Dexter discussed the discrepancy between theoretical underpinnings and practical applications of function approximation using deep NNs, drawing attention to a significant discrepancy between theoretical results and the actual performance of these models [16]. The following year, Dmitry presented a study on the global approximation of static maps using NNs, proposing generalizations of global approximation theory to encompass static or equal representations of linear sets [17]. In 2025, George A. Anastassiou presented research on Banach space-valued NNs focusing on the approximating continuous univariate functions quantitively on the whole real line using quasi-interpolation operators. A single hidden layer is included to link Banach space-valued feed-forward NNs [18]. The approximation abilities of NNs with this activation function have been studied by many researchers, for example [19–22]. In 2026, Nejm and Almurieb defined the APL-tensor product neural networks with multichannel convolution. A generalization by using a quasi-Banach space was given to define approximation spaces for neural networks based on connection and neuronal complexities [23]. This study aims to construct a mathematical framework for a generalized B-spline APL NNs based on tensor multiplication and multi-channel convolution. It also introduces the concept of approximation spaces in terms of connection complexity and neuronal complexity. Direct and inverse theorems have been reformulated to show the approximation rate of a NNs is related to the smoothness of the function. This provides a mathematical explanation for the efficiency and capability of NNs within approximation theory. The outline of this research paper is structured as follows. Section 2 introduces the fundamental concepts upon which the concept of NNs is built. Section 3 reviews the concept of approximation spaces, focusing on approximation spaces for

NNs within quasi-Banach spaces. This is followed by Section 4, which focuses on direct and its inverse theorems via generalized B-spline APL NNs. Finally, Section 5 presents the conclusions.

2. Preliminaries

Definition 2.1 (Convolution). [9]

- *Let $u \in \mathbb{R}^n, v \in \mathbb{R}^k$, the cyclic convolution $* : \mathbb{R}^{n\times k} \longrightarrow \mathbb{R}^n$ is given by*

$$(u * v)_i = \sum_{k=1}^{r} \mathrm{u}_{i\ominus k}\, \mathrm{v}_k,$$

  where $i \ominus k = (i-k) \mod n$.

- *Let $U \in \mathbb{R}^{n\times s}, V \in \mathbb{R}^{k\times s}$ the multi-channel convolution $\circledast : \mathbb{R}^{n\times s} \times \mathbb{R}^{k\times s} \to \mathbb{R}^n$ is given by*

$$U \circledast V = \sum_{i=1}^{s} U_i * V_i.$$

Definition 2.2 (Adaptive Piecewise Linear). [10] *Let $N \in \mathbb{N}$, be a prior set of hyper parameters, that refers to the number of hinges, for each unit $i = 1, \ldots, N$, $\varsigma_i : \mathbb{R} \to \mathbb{R}$ be an Adaptive Piecewise Linear (APL) activation function. It is defined as follows*

$$\varsigma_i(x) = \max(0, x) + \sum_{j=1}^{N} a_{i,j} \max(0, -x + b_{i,j}).$$

- *Note that $a_{i,j}$ variables control the slopes of the linear segments*
- *Note that $b_{i,j}$ variables determine the hinges' locations.*
- *Note that when $N = 0$, then APL ς_i is the standard ReLU denoted by ϱ.*

Definition 2.3 (APL Tensor Product). *Let $N \in \mathbb{N}$, then the APL tensor product of the collection of function $\{\varsigma_i^{\mu} : \mathbb{R} \longrightarrow \mathbb{R}\}_{i=1}^{N}$, is define by*

$$\bigotimes_{i=1}^{\aleph_\mu} \varsigma_i^{\mu}(\mathcal{X}_{ij}) : \mathcal{X}_{11} \times \ldots \times \mathcal{X}_{NN} \to \mathtt{Y}_{11} \times \ldots \times \mathtt{Y}_{NN},$$

with

$$\begin{bmatrix} \mathcal{X}_{11} & \ldots & \mathcal{X}_{1N} \\ \vdots & \ddots & \vdots \\ \mathcal{X}_{N1} & \ldots & \mathcal{X}_{NN} \end{bmatrix} \longmapsto \begin{bmatrix} \varsigma_1^{\mu}(\mathcal{X}_{11}) & \ldots & \varsigma_1^{\mu}(\mathcal{X}_{1N}) \\ \vdots & \ddots & \vdots \\ \varsigma_N^{\mu}(\mathcal{X}_{N1}) & \ldots & \varsigma_N^{\mu}(\mathcal{X}_{NN}) \end{bmatrix}.$$

Definition 2.4 (APL Neural Networks (APL NNs)). *Let $\mathrm{Ł} \in \mathbb{N}$, and let $\{\aleph_\mu\}_{\mu=0}^{\mathrm{Ł}} \subseteq \mathbb{N}$. An APL NNs with activation function ς_i^{μ} is a tuple*

$$\underline{N} = \left((T_1^1, \alpha_1^1), \cdots, \left(T_{\aleph-1}^{\mathrm{Ł}-1}, \alpha_{\aleph-1}^{\mathrm{Ł}-1}\right), \left(T_{\aleph}^{\mathrm{Ł}-1}, I_{\mathbb{R}\aleph_{\mathrm{Ł}}}\right)\right),$$

where the functions $\underset{\sim}{T} : \mathbb{R}^{\aleph_{\mu-1}\times\aleph_{\mu-1}} \to \mathbb{R}^{\aleph_\mu\times\aleph_\mu}$ *are affine linear maps,* $1 \leq \mu < Ł$ *and* $\alpha_i^\mu : \mathbb{R}^{\aleph_\mu\times\aleph_\mu} \to \mathbb{R}^{\aleph_\mu\times\aleph_\mu}$ *are given by*

$$\alpha_i^\mu = \bigotimes_{i=1}^{\aleph_\mu} \varsigma_i^\mu,$$

for $1 \leq \mu < Ł$.

Definition 2.5 (APL Generalized Neural Networks (APL GNNs))**.** *An APL NNs is called generalized if the linear maps* $\underset{\sim}{T}_i^\mu(x)$ *is given by*

$$\underset{\sim}{T}_i^\mu(x) = \varsigma_i^\mu \left(A_i^\mu x \circledast K + b_i^\mu \right),$$

so that

$$\underset{\sim}{T}_i^\mu(x) = \max\left(0, A_i^\mu x \circledast K + b_i^\mu\right) + \sum_{j=1}^{N} a_{i,j} \max\left(\left(0, -A_{i,j}^\mu x \circledast K + b_{i,j}^\mu\right)\right),$$

where the matrices $A_i^\mu \in \mathbb{R}^{\aleph_{\mu-1}\times\aleph_{\mu-1}}$ *are the weight matrices, the kernel* $K \in \mathbb{R}^{\aleph_m\times\aleph_m}$ *and the vectors* $b_i^\mu \in \mathbb{R}^{\aleph_\mu}$ *are the bias vectors.*

Definition 2.6 (APL Strict Neural Networks (APL SNN))**.** *An APL NNs is called APL SNN, if and only if* $\varsigma_i^\mu = \varrho, \forall\, 1 \leq \mu < Ł$ *and* $1 \leq i \leq N$.

Definition 2.7 (Realization of APL Neural Networks)**.** *Let* $\underline{N}_i$ *be an APL NNs as Definition 2.4. Then* $\underline{R}_i : \mathbb{R}^{\aleph_0} \times \mathbb{R}^{\aleph_0} \longrightarrow \mathbb{R}^{\aleph_\mu}$ *is called the realization of* $\underline{N}$ *and is defined by*

$$\underline{R}_i\left(\underline{N}\right) = \underset{\sim}{T}_{\aleph}^{Ł} \circledast \alpha_{\aleph}^{Ł-1} \circledast \underset{\sim}{T}_{\aleph-1}^{Ł-1} \circledast \cdots \circledast \alpha_i^1 \circledast \underset{\sim}{T}_i^1,$$

where for $1 \leq \mu < Ł$, $\underset{\sim}{T}_i^\mu : \mathbb{R}^{\aleph_{\mu-1}} \times \mathbb{R}^{\aleph_{\mu-1}} \longrightarrow \mathbb{R}^{\aleph_\mu} \times \mathbb{R}^{\aleph_\mu}$, *and*

$$\circledast_\mu : \mathbb{R}^{\aleph_{\mu-1}} \times \mathbb{R}^{\aleph_{\mu-1}} \times \mathbb{R}^{\aleph_{\mu-1}\times\aleph_{\mu-1}} \longrightarrow \mathbb{R}^{\aleph_\mu}.$$

Definition 2.8 (Classes of APL NNs)**.** *Let* $\aleph_0, \aleph_Ł \in \mathbb{N}$, *let* $W, N, Ł \in \mathbb{N}\bigcup\{\infty\}$. *Then*

$$\mho_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} = \left\{ \underline{N} \, is \, an \, APL \, NN : \begin{matrix} d_{in}\underline{N} = \aleph_0 \times \aleph_0, d_{out}\underline{N} = \aleph_Ł \\ W(\underline{N}) \leq W, Ł(\underline{N}) \leq Ł \end{matrix} \right\},$$

$$S\mho_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} = \left\{ \underline{N} \in \mho_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} : \underline{N} \, is \, an \, APL \, NN \right\},$$

$$NN_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} = \left\{ \underline{R}_i(\underline{N}) : \underline{N} \in \mho_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} \right\},$$

and

$$SNN_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} = \left\{ \underline{R}_i(\underline{N}) : \underline{N} \in S\mho_{W,Ł,N}^{\varsigma_i,\, \aleph_0\times\aleph_0,\aleph_Ł} \right\}.$$

Definition 2.9 (B-splines [8,11]). *Let $\underline{n} \in \mathbb{N}_{\mathbf{0}}$, the B-spline of degree 0 is given by $\beta_{+}^{(0)} = \chi\left[0,1\right)$ The B-spline of degree $\underline{n}$ is given by the $\underline{n}+1$ fold convolution of $\beta_{+}^{(0)}$, that is*

$$\beta_{+}^{(\underline{n})} = \beta_{+}^{(0)} \circledast \cdots \circledast \beta_{+}^{(0)}.$$

Now the Irwin-Hall distribution concept is introduced due to its importance in proving the direct theory.

Definition 2.10 (Irwin-Hall Distribution [8,11]). *Let $n, \underline{n} \in \mathbb{N}_0$, and let $\beta_{+}^{(\underline{n}-1)}$ be a B-spline of degree $\underline{n}$. Then $\beta_{+}^{(\underline{n}-1)} \in SNN_{2(n+1),2,n+1}^{\varrho_{n-1},1,1}$ and it can be decomposed as*

$$\beta_{+}^{(\underline{n}-1)}\left(f,x\right) = \frac{1}{\underline{n}!}\sum_{k=0}^{\underline{n}} \binom{\underline{n}}{k} \left(-1\right)^{\underline{n}-k} \varrho_{n-1}\left(x - \frac{\underline{n}h}{2} + kh\right).$$

Definition 2.11 (Generalized B-spline APL Neural Networks). *An APL NNs is called generalized if the linear maps $\Im_{\underline{n}-1,N}^{\mu}$ is given by*

$$\Im_{\underline{n}-1,N}^{\mu}\left(f,x\right) = \varsigma_{N}^{\mu}\left(\beta_{+}^{(\underline{n}-1)} f\left(\frac{k}{\underline{n}}\right) A_{N}^{\mu} x \circledast K + b_{N}^{\mu}\right),$$

so that

$$\begin{aligned}\Im_{\underline{n}-1,N}^{\mu}\left(f,x\right) = &\max\left(0, \beta_{+}^{(\underline{n}-1)} f\left(\frac{k}{\underline{n}}\right) A_{N}^{\mu} x \circledast K + b_{N}^{\mu}\right)\\ &+ \sum_{j=1}^{N} a_{N,j} \max\left(0, -\beta_{+}^{(\underline{n}-1)} f\left(\frac{k}{\underline{n}}\right) A_{N,j}^{\mu} x \circledast K + b_{N,j}^{\mu}\right),\end{aligned}$$

where $1 \leq \mu < Ł$ and $0 < k < \underline{n}$.

3. Approximation Spaces

This section discusses the most important basic concepts related to approximation spaces in terms of the degree of approximation and approximation classes. It also introduces the concepts of approximation spaces for NNs related to connection complexity and the neuron complexity.

Definition 3.1 (The Degree of Best Approximation [12]). *Let X be a quasi-Banach space with the quasi-norm $\left\|\cdot\right\|_{X}$. Let $f \in X$ and let $\sum \subseteq X$ be non-empty. The degree of best approximation of f from $\sum$ is given by*

$$E\left(f, \sum\right)_{X} = \inf_{g \in \sum} \left\|f - g\right\|_{X}.$$

Definition 3.2 (Approximation Classes). *Let X be a quasi-Banach space with the quasi-norm $\left\|\cdot\right\|_{X}$, and let*

$$X^{d} = \left\{\boldsymbol{f} = \left(f_1, \cdots, f_d\right) : \ f_i \in X\right\},$$

and let $\boldsymbol{f} \in X^{d}, \alpha \in (0, \infty)$ and $q \in (0, \infty]$. Define the approximation class by

$$A_{q}^{\alpha}\left(X^{d}, \sum\right) = \left\{\boldsymbol{f} \in X^{d}, \left\|\boldsymbol{f}\right\|_{A_{q}^{\alpha}\left(X^{d}, \sum\right)} < \infty\right\},$$

and the associated norm is given by

$$\|f\|_{A_q^\alpha}\left(X^d,\sum\right)=\left\{\begin{array}{cc}\sum_{i=1}^{d}\left(\sum_{n=1}^{\infty}\frac{1}{n}\left(n^\alpha\cdot E\left(f_i,\sum_{n-1}\right)_X\right)^q\right)^{\frac{1}{q}}, & \text{if } 0<q<\infty\\ \sup_{1\le i\le d,n\in\mathbb{N}}\left(n^\alpha\cdot E\left(f_i,\sum_{n-1}\right)_X\right), & \text{if } q=\infty\end{array}\right\}.$$

Definition 3.3 (Continuous Embedding [12]). *Let $X_1,\ X_2$ be quasi-Banach spaces, under quasi-norms $\|\cdot\|_{X_1}$ and $\|\cdot\|_{X_2}$, respectively. Then $X_1\hookrightarrow X_2$ denotes a continuous embedding between X_1 and X_2, that is there exist some $C\in\mathbb{R}: X_1\subseteq X_2$ and $C\|x\|_{X_1}\ge\|y\|_{X_2}$. Moreover, X_1 is said to be continuously embedded in X_2.*

Lemma 3.4. [11] *Let X be a quasi-Banach space, let $\sum_n\subseteq X$ and $n\in\mathbb{N}_0$, then the axioms are given by*

i) $\sum_0=\{0\}$.

ii) $\sum_n\subseteq\sum_{n+1}$.

iii) $\sum_n=C.\sum_n,\forall C\in\mathbb{R}/\{0\}$.

iv) $\exists\, c\in\mathbb{N}:\sum_n+\sum_n\subseteq\sum_{cn}$.

v) $\sum_\infty$ *is dense in* X, *where* $\sum_\infty$ *is given by* $\sum_\infty=\bigcup_{n\in\mathbb{N}_0}\sum_n$.

In the following proposition, we show that the classes of Approximation Spaces are quasi-Banach spaces satisfying that the continuous embedding in another spaces. The proof is omitted since it is similar to proof in Proposition VI.7 in [11].

Proposition 3.5 ($A_q^\alpha\left(X^d,\sum\right)$is an Approximation Spaces). *Assume that the properties (i)-(v) of Lemma 3.4 hold, then the spaces $\left(A_q^\alpha\left(X^d,\sum\right),\|\cdot\|_{A_q^\alpha(X^d,\sum)}\right)$ are quasi-Banach spaces that satisfy $A_q^\alpha\left(X^d,\sum\right)\hookrightarrow X^d$. Moreover, if $\alpha=\beta$ and $q\le s$ or if $\alpha>\beta$, then $A_q^\alpha\left(X^d,\sum\right)\hookrightarrow A_s^\beta\left(X^d,\sum\right)$.*

Definition 3.6 (Modulus of Smoothness). [6] *Modulus of Smoothness(MS), is one of the best measures of the approximation rate. This is because it evaluates how accurately the approximation reflects the function. Let X be a quasi-Banach space, and let $f\in X^d$. The norm is defined by*

$$\|f\|_{A_q^\alpha(X^d,\sum)}=\sum_{i=1}^{d}\left(\sum_{n=1}^{\infty}\frac{1}{n}\left(n^\alpha,E\left(f_i,\sum\nolimits_{n-1}\right)_X\right)^q\right)^{\frac{1}{q}},\ \text{if } 0<q<\infty$$

The $\underline{n}$th MS is given by

$$\omega_{\underline{n}}(f,\delta)_p=\sup_{0<h\le\delta}\left\|\Delta_h^{\underline{n}}\right\|_{A_q^\alpha(X^d,\sum)}.$$

Also, the $\underline{n}$ th symmetric difference is as follows

$$\Delta_h^{\underline{n}}(f,x) = \left\{ \binom{\underline{n}}{k} (-1)^{\underline{n}-k} f\left(x - \frac{\underline{n}h}{2} + kh\right) \right\},$$

for $x - \frac{\underline{n}h}{2} \in I$.

Definition 3.7 (Approximation Spaces for APL GNNs)**.** *Let $n \in \mathbb{N}_0$, define the approximation spaces associated with the quasi-norms*

$$W_q^\alpha\left(X^d, \varsigma_i, Ł\right) = A_q^\alpha\left(X^d, \sum\right),$$

and

$$\|\cdot\|_{W_q^\alpha\left(X^d,\varsigma_i,Ł\right)} = \|\cdot\|_{A_q^\alpha\left(X^d,\sum\right)},$$

with $\sum_n = W_q^\alpha\left(X^d, \varsigma_i, Ł\right)$. Similarly, define

$$N_q^\alpha\left(X^d, \varsigma_i, Ł\right) = A_q^\alpha\left(X^d, \sum\right),$$

and

$$\|\cdot\|_{N_q^\alpha\left(X^d,\varsigma_i,Ł\right)} = \|\cdot\|_{A_q^\alpha\left(X^d,\sum\right)},$$

with $\sum_n = N_q^\alpha\left(X^d, \varsigma_i, Ł\right)$.

The space $W_q^\alpha\left(X^d, \varsigma_i, Ł\right)$ is known as the approximation space equipped with the complexity of connection. Similarly, $N_q^\alpha\left(X^d, \varsigma_i, Ł\right)$ is known as the approximation space equipped with the complexity of neurons.

Definition 3.8 (Approximation Space for APL SNNs)**.** *Let $n \in \mathbb{N}_0$, define the approximation spaces associated with the quasi-norms*

$$SW_q^\alpha\left(X^d, \varsigma_i, Ł\right) = A_q^\alpha\left(X^d, \sum\right),$$

and

$$\|\cdot\|_{SW_q^\alpha\left(X^d,\varsigma_i,Ł\right)} = \|\cdot\|_{A_q^\alpha\left(X^d,\sum\right)},$$

with $\sum_n = SW_n\left(X^d, \varsigma_i, Ł\right)$. Similarly, define

$$SN_q^\alpha\left(X^d, \varsigma_i, Ł\right) = A_q^\alpha\left(X^d, \sum\right),$$

and

$$\|\cdot\|_{SN_q^\alpha\left(X^d,\varsigma_i,Ł\right)} = \|\cdot\|_{A_q^\alpha\left(X^d,\sum\right)},$$

with $\sum_n = SN_n\left(X^d, \varsigma_i, Ł\right)$.

4. Direct and Inverse Theorems via Generalized B-Spline APL Neural Networks

This section includes proof of the direct and inverse theorems based on the B-Spline within quasi-Banach spaces.

Theorem 4.1. *Let X be a quasi-Banach space, and for $f \in X$, then there is an $\underline{n} \in \mathbb{N}$, with $\Im^{\mu}_{(\underline{n}-1,N)} \in SNN^{\varsigma_i,1,1}_{2(n+2),2,n+2}$ that satisfies*

$$E_{\underline{n}}(f) \leq \omega_{\underline{n}}(f,\delta)_X .$$

Proof. Let $\Im_{(\underline{n}-1,N)} \in SNN^{\varsigma_i,1,1}_{2(n+2),2,n+2}$. For $f \in X$, then there is some $\underline{n} \in \mathbb{N}$, that satisfies

$$\begin{aligned}\left\|\Im^{Ł}_{(\underline{n}-1,N)}(f,x)-f(x)\right\|_X &\leq \left\|f(x)-\Im^{1}_{(\underline{n}-1,N)}(f,x)\right\|_X+\left\|\Im^{1}_{(\underline{n}-1,N)}(f,x)-\Im^{2}_{(\underline{n}-1,N)}(f,x)\right\|_X\\ &+\cdots+\left\|\Im^{\mu}_{(\underline{n}-1,N)}(f,x)-\Im^{\mu-1}_{(\underline{n}-1,N)}(f,x)\right\|_X\\ &+\cdots+\left\|\Im^{Ł}_{(\underline{n}-1,N)}(f,x)-\Im^{Ł-1}_{(\underline{n}-1,N)}(f,x)\right\|_X .\end{aligned}$$

Since

$$\begin{aligned}\left\|\Im^{Ł}_{(\underline{n}-1,N)}(f,x)-\Im^{Ł-1}_{(\underline{n}-1,N)}(f,x)\right\|_X &\leq \left\|c^{\mu}_N\left(\beta^{(\underline{n}-1)}_{+}f\left(\frac{k}{\underline{n}}\right)A^{\mu-1}_N x\circledast K+b^{\mu-1}_N\right)\right.\\ &\left.-\left(\beta^{(\underline{n}-1)}_{+}f\left(\frac{k}{\underline{n}}\right)A^{\mu-1}_N x\circledast K+b^{\mu-1}_N\right)\right\|_X\\ &\leq \beta^{(\underline{n}-1)}_{+}f\left(\frac{k}{\underline{n}}\right)A^{\mu-1}_N x\circledast K+b^{\mu}_N\\ &+\left\|\beta^{(\underline{n}-1)}_{+}f\left(\frac{k}{\underline{n}}\right)A^{\mu-1}_N x\circledast K+b^{\mu-1}_N\right\|_X\\ &-\beta^{(\underline{n}-1)}_{+}f\left(\frac{k}{\underline{n}}\right)A^{\mu-1}_N x\circledast K+b^{\mu}_N\\ &+\left\|\beta^{(\underline{n}-1)}_{+}f\left(\frac{k}{\underline{n}}\right)A^{\mu-1}_N x\circledast K+b^{\mu-1}_N\right\|_X\\ &\leq \frac{1}{\underline{n}!}\sum_{k=0}^{\underline{n}}\binom{\underline{n}}{k}(-1)^{\underline{n}-k}\varrho_{(n-1)}\left(x-\frac{\underline{n}\underline{h}}{2}+kh\right)\left\|A^{\mu-1}_N x\circledast K+b^{\mu}_N\right\|_X\\ &\leq \frac{1}{\underline{n}!}\omega_{\underline{n}}(f,\delta)\left\|A^{\mu-1}_N x\circledast K+b^{\mu}_N\right\|_X\\ &\leq C_{\mu,\underline{n}}\omega_{\underline{n}}(f,\delta).\end{aligned}$$

□

Lemma 4.2. *Let $I=\cup I_{\underline{m}}, \underline{m}=1,2,\ldots,\underline{n}$ where $I_{\underline{m}}$ be a partition of I with $|I_{\underline{m}}| \sim \frac{1}{\underline{n}}$. Then*

$$\sum_{\underline{m}=1}^{\underline{n}} E_{\underline{n}} \leq C E_{\underline{m}}(f).$$

Lemma 4.3. *For $f \in X$, then for $\Im^{\mu}_{(n-1,N)}(f,x)$, we have*

$$\omega_{\underline{n}}\left(\Im^{\mu}_{(\underline{n}-1,N)}(f,x)\right) \leq C\left\|\Im^{\mu}_{(\underline{n}-1,N)}(f,x)\right\|_X.$$

Proof.

$$\begin{aligned}
\omega_{\underline{\mathfrak{n}}}\left(\Im^{\mu}_{(\underline{\mathfrak{n}}-1,N)}(f,x)\right) &= \left\|\varsigma^{\mu}_{N}\left(\beta^{(\underline{\mathfrak{n}}-1)}_{+} f\left(\frac{k}{\underline{\mathfrak{n}}}\right)\right)\right\|_X \\
&\leq C\left\|\beta^{(\underline{\mathfrak{n}}-1)}_{+} f\left(\frac{k}{\underline{\mathfrak{n}}}\right) A^{\mu}_{N} x \circledast K + b^{\mu}_{N}\right\|_X \\
&+ \left\|-\beta^{(\underline{\mathfrak{n}}-1)}_{+} f\left(\frac{k}{\underline{\mathfrak{n}}}\right) A^{\mu}_{N} x \circledast K\right\|_X \\
&\leq C\left\|\Im^{\text{Ł}}_{(\underline{\mathfrak{n}}-1,N)}(f,x)\, A^{\mu}_{N} x \circledast K + b^{\mu}_{N}\right\|_X \\
&\leq C\left\|\Im^{\mu}_{(\underline{\mathfrak{n}}-1,N)}(f,x)\right\|_X.
\end{aligned}$$

□

Theorem 4.4. *Let X be a quasi-Banach space, then for $f \in X$, get that*

$$\omega_{\underline{n}}(f,\delta) \leq E_{\underline{n}}(f).$$

Proof. Let $\underline{\mathfrak{m}} < \underline{\mathfrak{n}}$, assume that

$$\left\|\Im^{\mu}_{\underline{\mathfrak{n}}-1,N}(f,x) - \Im^{\mu}_{\underline{\mathfrak{m}}-1,N}(f,x)\right\|_X \leq C E_{\underline{\mathfrak{m}}}(f),$$

then by Lemma 4.2, Lemma 4.3 and by Theorem 4.1, get that

$$\begin{aligned}
\omega_{\underline{\mathfrak{n}}}(f,\delta) &\leq C\left(\omega_{\underline{\mathfrak{n}}}\left(\Im^{\text{Ł}}_{\underline{\mathfrak{n}}-1,N}(f,\delta) - f(x)\right) + \omega_{\underline{\mathfrak{n}}}\left(\Im^{\text{Ł}}_{\underline{\mathfrak{n}}-1,N}(f,\delta)\right)\right) \\
&\leq C\left(\left\|\Im^{\text{Ł}}_{\underline{\mathfrak{n}}-1,N} - f(x)\right\|_X + E_{\underline{\mathfrak{m}}}(f)\right) \\
&\leq C E_{\underline{\mathfrak{n}}}(f).
\end{aligned}$$

□

5. Conclusions

This paper discusses approximation spaces of neural networks by constructing a generalized B-spline APL neural network. The direct and inverse theorems for estimating the approximation are also proven by fundamental mathematical properties of B-spline and APL activation function are investigated to obtain an approximation equivalent to the smoothness coefficient, with upper and lower limits approaching zero, due to their relationship to the smoothness coefficient.

Authors' Contributions. All authors have read and approved the final version of the manuscript. The authors contributed equally to this work.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] W.S. McCulloch, W. Pitts, A Logical Calculus of the Ideas Immanent in Nervous Activity, Bull. Math. Biophys. 5 (1943), 115–133. https://doi.org/10.1007/bf02478259.

[2] A. Pietsch, Approximation Spaces, J. Approx. Theory 32 (1981), 115–134. https://doi.org/10.1016/0021-9045(81)90109-x.

[3] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Springer, Berlin, 1993. https://doi.org/10.1007/978-3-662-02888-9.

[4] G. Cybenko, Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient, Technical Report No. 31, Department of Computer Science, Tufts University, 1988.

[5] K.-I. Funahashi, On the Approximate Realization of Continuous Mappings by Neural Networks, Neural Netw. 2 (1989), 183–192. https://doi.org/10.1016/0893-6080(89)90003-8.

[6] Z. Ditzian, V.H. Hristov, K.G. Ivanov, Moduli of Smoothness and K-Functionals in L_p, $0 < p < 1$, Constr. Approx. 11 (1995), 67–83. https://doi.org/10.1007/bf01294339.

[7] G.A. Anastassiou, Rate of Convergence of Some Neural Network Operators to the Unit-Univariate Case, J. Math. Anal. Appl. 212 (1997), 237–262. https://doi.org/10.1006/jmaa.1997.5494.

[8] M. Unser, Splines: a Perfect Fit for Signal and Image Processing, IEEE Signal Process. Mag. 16 (1999), 22–38. https://doi.org/10.1109/79.799930.

[9] J.M. Almira, U. Luther, Generalized Approximation Spaces and Applications, Math. Nachr. 263–264 (2004), 3–35. https://doi.org/10.1002/mana.200310121.

[10] C. Bao, Q. Li, Z. Shen, C. Tai, L. Wu, X. Xiang, Approximation Analysis of Convolutional Neural Networks, East Asian J. Appl. Math. 13 (2023), 524–549. https://doi.org/10.4208/eajam.2022-270.070123.

[11] F. Agostinelli, M. Hoffman, P. Sadowski, P. Baldi, Learning Activation Functions to Improve Deep Neural Networks, arXiv:1412.6830, 2014. https://doi.org/10.48550/arXiv.1412.6830.

[12] R. Gribonval, G. Kutyniok, M. Nielsen, F. Voigtlaender, Approximation Spaces of Deep Neural Networks, Constr. Approx. 55 (2022), 259–367. https://doi.org/10.1007/s00365-021-09543-4.

[13] R. DeVore, B. Hanin, G. Petrova, Neural Network Approximation, Acta Numer. 30 (2021), 327–444. https://doi.org/10.1017/s0962492921000052.

[14] Z. Shen, H. Yang, S. Zhang, Neural Network Approximation: Three Hidden Layers Are Enough, Neural Netw. 141 (2021), 160–173. https://doi.org/10.1016/j.neunet.2021.04.011.

[15] D. Elbrächter, D. Perekrestenko, P. Grohs, H. Bölcskei, Deep Neural Network Approximation Theory, IEEE Trans. Inf. Theory 67 (2021), 2581–2623. https://doi.org/10.1109/tit.2021.3062161.

[16] B. Adcock, N. Dexter, The Gap between Theory and Practice in Function Approximation with Deep Neural Networks, SIAM J. Math. Data Sci. 3 (2021), 624–655. https://doi.org/10.1137/20m131309x.

[17] D. Yarotsky, Universal Approximations of Invariant Maps by Neural Networks, Constr. Approx. 55 (2022), 407–474. https://doi.org/10.1007/s00365-021-09546-1.

[18] G.A. Anastassiou, Parametrized Arctangent Sigmoid Function Based Banach Space Valued Neural Network Approximation, Neural Parallel Sci. Comput. 33 (2025), 25–37. https://doi.org/10.46719/npsc2025.33.02.

[19] H.A. Almurieb, A.A. Hamody, Best Neural Network Approximation by Using Bernstein Polynomials with GRNN Learning Application, Al-Bahir 1 (2022), 2. https://doi.org/10.55810/2313-0083.1003.

[20] H.A. Almurieb, Z.A. Sharba, M. A. Kareem, Trigonometric Jackson Integrals Approximation by a k-Generalized Modulus of Smoothness, Nonlinear Funct. Anal. Appl. 27 (2022), 807–812. https://doi.org/10.22771/nfaa.2022.27.04.09.

[21] R. Abu-Gdairi, Analytical Solution of Nonlinear Fractional Gradient-Based System Using Fractional Power Series Method, Int. J. Anal. Appl. 20 (2022), 51. https://doi.org/10.28924/2291-8639-20-2022-51.

[22] S. Volosivets, Approximation in Orlicz-Stepanets Spaces Defined by Vilenkin-Fourier Coefficients, J. Math. Sci. 290 (2025), 287–300. https://doi.org/10.1007/s10958-025-07607-5.

[23] M.M. Nejm, H.A. Almurieb, Approximation Spaces of Neural Networks with Adaptive Piecewise Linear Activation Functions, Al-Nahrain J. Sci. in Press.