Approximation Properties of Complex-valued Neural Networks: An Overview

Paul Geuchen
KU Eichstätt-Ingolstadt
Mathematical Institute for Machine Learning and Data Science (MIDS)
Ingolstadt, Germany
paul.geuchen@ku.de

***Abstract*—While the approximation properties of real-valued neural networks have been a subject of intense mathematical research since the 1990s, the approximation properties of *complex-valued* neural networks have only gained increased attention in recent years. In this paper, we provide an overview of the known results and specifically discuss the similarities and differences between the complex-valued and real-valued case.**

***Index Terms*—complex-valued neural networks, universality, approximation rates**

I. Introduction

Since the foundational work of the 1990s, the approximation properties of real-valued neural networks (RVNNs) have been extensively studied [5, 9, 17, 21, 23, 27, 28, 33]. Both qualitative aspects such as universality and quantitative aspects are subject of investigation. In recent years, however, *complex-valued* neural networks (CVNNs) have emerged as a natural modification of RVNNs, due to their suitability for problems involving inherently complex-valued data such as MRI [8, 16, 30] and radar imaging [4, 25, 34]. Despite their success in practical applications, the theoretical understanding of CVNNs remains comparatively underdeveloped.

This paper provides an overview of the current state of research on the approximation properties of CVNNs. We summarize key results, highlighting their connections to and differences from the well-established theory of RVNNs.

A. Preliminaries

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $\phi : \mathbb{F} \to \mathbb{F}$ be a given function. An $\mathbb{F}$-valued neural network Φ with *activation function* ϕ is a function of the form

$$\Phi = V^{(L)} \circ \phi \circ V^{(L-1)} \circ ... \circ V^{(1)} \circ \phi \circ V^{(0)},$$

where the maps $V^{(\ell)} : \mathbb{F}^{N_\ell} \to \mathbb{F}^{N_{\ell+1}}$ (for proper choices of $N_\ell \in \mathbb{N}$) are $\mathbb{F}$-affine, and the function ϕ is applied *componentwise* to a vector. N_0 is called the number of *input neurons*, N_{L+1} the number of *output neurons*, L is the *depth* and $\max_{\ell=0,...,L+1} N_\ell$ is the *width* of Φ. Φ is called *shallow* if $L = 1$ and *deep* if $L \geq 2$. In the case of a shallow network, we refer to N_1 as the number of *hidden neurons*. Each affine map $V^{(\ell)}$ can be identified by a matrix $W^{(\ell)} \in \mathbb{F}^{N_{\ell+1} \times N_\ell}$ and a *bias* vector $b^{(\ell)} \in \mathbb{F}^{N_{\ell+1}}$. The entries of the $W^{(\ell)}$ and $b^{(\ell)}$ are called the *weights* of Φ.

At first sight, CVNNs could appear to only be a trivial extension of RVNNs by using the identification $\mathbb{C} \cong \mathbb{R}^2$. However, the two network classes differ in the following aspects:

- The activation function in a CVNN is a function $\mathbb{C} \to \mathbb{C}$ (i.e., $\mathbb{R}^2 \to \mathbb{R}^2$), whereas the activation function in an RVNN is a function $\mathbb{R} \to \mathbb{R}$.
- The affine maps in a CVNN are required to be $\mathbb{C}$-affine whereas the affine maps in an RVNN only need to be $\mathbb{R}$-affine.[1]

These two points show that it is *not* possible to view CVNNs as a special case of RVNNs or vice versa. As a consequence, one can (in general) not directly transfer approximation-theoretic properties of RVNNs to the complex-valued case but needs to carefully examine which properties translate, and in which way.

The key advantage of CVNNs in comparison to RVNNs is that it is possible to use activation functions that are able to faithfully handle the complex nature of the inputs of the network. This can for example be achieved by using an activation function that preserves the phase of the input. Two of the most popular complex activation functions are the modReLU as introduced in [3] and the complex cardioid as proposed in [30].

II. Universality

The most fundamental aspect regarding the approximation-theoretic properties of neural networks is the question of *universality*.

Definition II.1. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, $d_1, d_2 \in \mathbb{N}$ and $\mathcal{F} \subseteq C(\mathbb{F}^{d_1}; \mathbb{F}^{d_2})$ be a class of continuous functions. $\mathcal{F}$ is called *universal*, if for every compact set $K \subseteq \mathbb{F}^{d_1}$, every $g \in C(K; \mathbb{C}^{d_2})$ and every $\varepsilon > 0$, there exists a function $f \in \mathcal{F}$ satisfying

$$\|f - g\|_{L^\infty(K)} < \varepsilon.$$

[1]Each $\mathbb{C}$-affine map $V : \mathbb{C}^{N_1} \to \mathbb{C}^{N_2}$ has an associated map $\widetilde{V} : \mathbb{R}^{2N_1} \to \mathbb{R}^{2N_2}$, which is trivially $\mathbb{R}$-affine. On the other hand, given an arbitrary $\mathbb{R}$-affine map $\widetilde{W} : \mathbb{R}^{2N_1} \to \mathbb{R}^{2N_2}$, the associated map $W : \mathbb{C}^{N_1} \to \mathbb{C}^{N_2}$ is *not* necessarily $\mathbb{C}$-affine.

A. The universal approximation theorem for CVNNs

In the most classical setting, one analyzes which properties an activation function has to satisfy in order for the class of associated networks with arbitrary but *fixed* depth to be universal. In the real-valued case, this question was extensively studied in the late 80s and early 90s (see for instance [9, 17]). It could be shown that the set of RVNNs with fixed continuous activation function ϕ is universal if and only if ϕ is non-polynomial [17, Theorem 1], known as the *universal approximation theorem* (UAT). However, it remained open in what way this result generalizes to the case of CVNNs.

Two of the first works to address this issue were [1, 2], by P. Arena, L. Fortuna, R. Re and M. G. Xibilia. These authors in particular observed that a holomorphic activation function *never* gives rise to universal CVNNs. Indeed, if the activation function is holomorphic, then every CVNN that uses this activation function is holomorphic too. Since the locally uniform limit of holomorphic functions is holomorphic, it is easy to see that the set of holomorphic functions is not universal, showing that the set of CVNNs using holomorphic activation functions cannot be universal. Moreover, the authors show universality of shallow CVNNs that use a *discriminatory* activation function, similar to the statement for RVNNs in [9]. However, for an activation function $\phi : \mathbb{C} \to \mathbb{C}$ it is usually not easy to determine whether it is discriminatory or not.

A full characterization of the complex activation functions that yield universal CVNNs was given by F. Voigtlaender in [31]. An activation function $\phi \in C(\mathbb{C}; \mathbb{C})$ is called *polyharmonic*, if[2] $\phi \in C^\infty(\mathbb{C}; \mathbb{C})$ and if there exists $m \in \mathbb{N}_0$ for which $\Delta^m \phi \equiv 0$, where $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ denotes the *Laplace-operator*. Further, ϕ is called *anti-holomorphic*, if $\overline{\phi}$ is holomorphic. The main result from [31] then reads as follows:

Theorem II.2 (UAT for CVNNs, [31, Theorems 1.3, 1.4])**.** *Let $\phi \in C(\mathbb{C}; \mathbb{C})$ and $d, L \in \mathbb{N}$. Then the set of CVNNs with d input neurons, a depth of L and activation function ϕ is universal if and only if:*

1) *ϕ is not polyharmonic ($L = 1$).*
2) *ϕ is neither holomorphic, nor anti-holomorphic, nor a polynomial in z and $\overline{z}$ ($L \geq 2$).*

Remarkably, the characterization of "good" activation functions in the case of shallow CVNNs differs from the characterization when $L \geq 2$. This is a significant difference to the case of RVNNs, where the same characterization is correct for any choice of L.

We further note that both [17] and [31] even show a more general result, considering activation functions that are locally bounded and where the set of discontinuities has measure zero. Trivially, this includes all continuous functions.

The proof of Theorem II.2 works in its main steps similar to the proof of the real-valued statement in [17]: According to the Stone-Weierstraß theorem (see [10, Theorem 4.51]) it suffices to approximate complex polynomials in z and $\overline{z}$. Making heavy use of the *Wirtinger calculus* (see for instance [14, §1]) one can then show that it is possible to approximate complex polynomials in z and $\overline{z}$ locally uniformly by CVNNs.

We further mention a recently published result [29, Theorem 1], in which a universal approximation property for shallow $\mathbb{H}$-valued neural networks is established, where $\mathbb{H}$ is any non-degenerate hypercomplex algebra (see [29, Definitions 1-3]). Note that the complex numbers form such an algebra. While the result is on the one hand remarkable due to its generality since it applies to arbitrary hypercomplex algebras, it is limited on the other hand, since it only considers *split activation functions*. In the complex-valued case, these are functions of the form $\phi(x + \mathrm{i}y) = \psi(x) + \mathrm{i}\psi(y)$ with $\psi : \mathbb{R} \to \mathbb{R}$. The result [29, Theorem 1] hence does not apply to popular complex activation functions such as the modReLU or the complex cardioid.

B. Universality of deep narrow CVNNs

In Section II-A, we studied networks of arbitrary but *fixed* depth and unrestricted width. However, various applications have shown that deep networks are advantageous compared to networks with fewer layers, motivating the investigation of the universality of "deep and narrow" networks, i.e., sets of networks with *unrestricted* depth and restricted width. P. Kidger and T. Lyons [15] showed that every continuous non-affine real activation function yields universal RVNNs in the case of unrestricted depth, which is a larger class of functions than in the case of restricted depth, where the activation function must be non-polynomial in order to yield universality. Further, it is shown that the sufficient width of the RVNNs can be upper bounded by $d_1 + d_2 + 2$, where d_1 and d_2 are the input and output dimension of the considered networks, respectively.

This observation was generalized to the complex-valued setting in [11]. The main result of that paper reads as follows.

Theorem II.3 (UAT for deep narrow CVNNs, [11, Theorem 1.1])**.** *Let $d_1, d_2 \in \mathbb{N}$. Moreover, let $\phi \in C(\mathbb{C}; \mathbb{C})$ be (real) differentiable at some point with non-vanishing derivative at that point. Then the set of CVNNs with activation function ϕ, d_1 input neurons, d_2 output neurons, arbitrary depth and width at most $2d_1 + 2d_2 + 5$ is universal if and only if ϕ is neither holomorphic, nor anti-holomorphic, nor $\mathbb{R}$-affine.*

In particular, Theorem II.3 shows that, similar to the case of RVNNs, the class of "good" activation functions *increases* if one moves from the case of restricted depth to the case of unrestricted depth, in the sense that the condition "not a polynomial in z and $\overline{z}$" is replaced by "not $\mathbb{R}$-affine". Moreover, [11] discusses the optimality of the sufficient width $2d_1 + 2d_2 + 5$. In particular, it is shown that if there exists a point $z_0 \in \mathbb{C}$ at which ϕ is differentiable with

$$\left(\partial_{\mathrm{wirt}}\phi(z_0), \overline{\partial}_{\mathrm{wirt}}\phi(z_0)\right) \in [\mathbb{C}^* \times \{0\}] \cup [\{0\} \times \mathbb{C}^*],$$

then in fact a width of $d_1 + d_2 + 3$ is sufficient. Here, $\mathbb{C}^* := \mathbb{C} \setminus \{0\}$ and ∂_{wirt} and $\overline{\partial}_{\mathrm{wirt}}$ denote the Wirtinger derivatives. However, it is proved that in the given generality of arbitrary

[2] For notations of the form "$C^r(\mathbb{C}; \mathbb{C})$", differentiability is to be understood in the sense of real variables.

continuous activation functions, a width of $\max\{2d_1, 2d_2\}$ is necessary; see [11, Theorem 6.4].

The proof of Theorem II.3 is mainly based on a generalized Taylor expansion for functions from $\mathbb{C}$ to $\mathbb{C}$. Using this expansion and the given assumptions on ϕ, it is possible to approximate elementary building blocks such as the complex identity, complex conjugation and complex multiplication by CVNNs. By carefully combining those building blocks, one can then prove that each complex polynomial in z and $\overline{z}$ can be approximated by CVNNs of a certain width. Again, the Stone-Weierstrass theorem then yields the claim.

C. Holomorphic activation functions

While the set of CVNNs with a holomorphic activation function (of arbitrary depth and width) can surely not be universal (since this set only consists of holomorphic functions), it is then natural to ask *which* functions can be approximated by such CVNNs. In some application areas, it might for example be sufficient to be able to approximate holomorphic functions (and not arbitrary *continuous* functions). This question was analyzed in [22] with the following result.

Theorem II.4 (special case of [22, Theorem 1]). *Let $\phi : \mathbb{C} \to \mathbb{C}$ be an entire non-polynomial function. Moreover, let $K \subseteq \mathbb{C}^d$ be compact and $f \in C(K;\mathbb{C})$. Then the following are equivalent:*

1) *For every $\varepsilon > 0$ there exists a shallow CVNN $\Phi : \mathbb{C}^d \to \mathbb{C}$ with activation function ϕ satisfying $\|f - \Phi\|_{L^\infty(K)} \leq \varepsilon$.*
2) *For every $\varepsilon > 0$ there exists a complex polynomial $P : \mathbb{C}^d \to \mathbb{C}$ satisfying $\|f - P\|_{L^\infty(K)} \leq \varepsilon$.*

Note that the polynomial P appearing in Theorem II.4 is a polynomial solely in the variable z and *not* in the variables z and $\overline{z}$. Unlike in the real case, it is in general quite challenging to determine the set of functions f that satisfy 2). At least in the case $d = 1$ the following characterization is possible: If $\mathbb{C} \setminus K$ is connected, then the set of functions f that satisfy 2) coincides with the set of functions that are holomorphic on the interior K° (see [26, Theorem 20.5]). We also mention [6, Theorem 1] which is a special case of Theorem II.4 by using [26, Theorem 13.11(e)].

III. Quantitative bounds

The previous section focused on *qualitative* results, addressing the existence of networks that approximate a target function up to an arbitrary, given precision. In the present section, we explore *quantitative* bounds, i.e., we examine the network size (in terms of the number of parameters) required to achieve a desired approximation accuracy.

All the results presented in this section are concerned with the approximation of *Sobolev functions* defined on $\Omega_d := [-1,1]^d + \mathrm{i}[-1,1]^d$. To this end, let $\mathcal{B}\left(W^{r,p}_{2d}\right)$ denote the set of L^p-Sobolev functions of regularity r defined on $[-1,1]^{2d}$ satisfying $\|f\|_{W^{r,p}_{2d}} \leq 1$. We then write $\mathcal{B}\left(W^{r,p}_{d}(\mathbb{C})\right)$ for the set of functions $f : \Omega_d \to \mathbb{C}$ with $\mathrm{Re}(f), \mathrm{Im}(f) \in \mathcal{B}\left(W^{r,p}_{2d}\right)$, upon identifiying $\Omega_d \cong [-1,1]^{2d}$. We then seek to establish upper and lower bounds for the error

$$\sup_{f \in \mathcal{B}\left(W^{r,p}_{d}(\mathbb{C})\right)} \inf_{\Phi} \|f - \Phi\|_{L^q(\Omega_d)} \tag{III.1}$$

for various choices of q, where the infimum runs over all CVNNs Φ of a certain size. Typically, the input dimension d and the regularity r of the target functions are considered as fixed and one analyzes, how the error in (III.1) behaves as the size of the considered CVNNs is increased.

The first paper to address this question was [7], in which the expressive power of CVNNs with $\mathrm{modReLU}$ activation function is considered. We here only discuss it briefly, since its main result could be significantly improved in [13]; see also Section III-A. Essentially, [7, Theorem 1] shows that

$$\sup_{f \in \mathcal{B}\left(W^{r,\infty}_{d}(\mathbb{C})\right)} \inf_{\Phi} \|f - \Phi\|_{L^\infty(\Omega_d)} \lesssim_{d,r} \left(n \cdot \ln^{-2}(n)\right)^{-r/(2d)},$$

where the infimum is over all $\mathrm{modReLU}$-CVNNs with at most $C(d,r) \cdot \ln(n)$ layers, n non-zero weights, and all weights bounded in absolute value by $n^{22r} \cdot \ln^{44r}(n)$. The proof of [7, Theorem 1] relies to a large extent on techniques used in [32], where a similar statement for the approximation using deep ReLU-networks in the real-valued case was shown. Moreover, [7] establishes a certain optimality of the upper bound under the restriction that the magnitude of the weights grows at most polynomially with the number of weights; see [7, Theorem 12].

A. The case of continuous weight selection

In the real-valued case it is known that already *shallow* neural networks with *any* activation function that is smooth but non-polynomial on some arbitrary open ball can approximate L^∞-Sobolev functions of regularity r at the rate of $n^{-r/d}$ [21]. The extension to the complex-valued case is contained in [13].

Theorem III.1 ([13, Theorem 3.2]). *Let $d, r \in \mathbb{N}$. Then there exists a constant $C = C(d,r) > 0$ with the following property: For any complex activation function $\phi : \mathbb{C} \to \mathbb{C}$ that is smooth and non-polyharmonic on an open set $\varnothing \neq U \subseteq \mathbb{C}$, any $n \in \mathbb{N}$ and any $f \in \mathcal{B}(W^{r,\infty}_d(\mathbb{C}))$, there exists a shallow CVNN $\Phi : \mathbb{C}^d \to \mathbb{C}$ with n hidden neurons and activation function ϕ satisfying*

$$\|f - \Phi\|_{L^\infty(\Omega_d)} \leq C \cdot n^{-r/(2d)}.$$

The exponent $-r/(2d)$ instead of $-r/d$ in the real setting results from the identification $\mathbb{C} \cong \mathbb{R}^2$. Essentially, Theorem III.1 states that CVNNs *match* the approximation power of RVNNs. Remarkably, Theorem III.1 applies to arbitrary activation functions that are smooth and non-polyharmonic on some non-empty open set. This property can be verified for the popular complex activation functions $\mathrm{modReLU}$ and complex cardioid, see [13, Corollaries A.4&A.6]. Hence, Theorem III.1 *improves* [7, Theorem 1] in the sense that (i) it applies to very general activation functions, (ii) it improves the approximation rate by a logarithmic factor and (iii) it is shown that this rate can already be achieved by shallow CVNNs. On the other hand, [7, Theorem 1] provides a bound on the magnitude of the weights of the approximating networks. Such a bound is not obtained in [13] for the case of arbitrary activation functions.

The proof in [13] works similarly as the proof in the real-valued case in [21]. Essentially, one uses the Wirtinger derivatives combined with a multivariate version of *divided differences* to approximate polynomials in z and $\overline{z}$ using CVNNs of a certain size which only depends on the degree of the approximated polynomial and *not* on the approximation accuracy (see [13, Theorem 3.1]). Finally, one uses the well-known fact that Sobolev functions of regularity r can be approximated at the rate of s^{-r} by polynomials of degree s.

It is even shown in [13] that the weights of the CVNN Φ that connect the input and the hidden layer can be chosen *independent* of the target function f and only the weights connecting hidden layer and output neuron have to be adapted to f. Furthermore, it is shown that the map that assigns to a target function $f \in \mathcal{B}(W_d^{r,\infty}(\mathbb{C}))$ the parameters of the approximating network realizing the upper bound of $C \cdot n^{-r/(2d)}$ can be chosen as a *continuous* map. This setting is called *continous weight selection*.

In the regime of continous weight selection, the following lower bound can be shown.

Theorem III.2 ([13, Theorem 4.1]). *Let $d, r \in \mathbb{N}$. Then there exists a constant $c = c(d, r) > 0$ with the following property: For any $n \in \mathbb{N}$, any map $\overline{a} : \mathcal{B}(W_d^{r,\infty}(\mathbb{C})) \to \mathbb{C}^n$ that is continuous with respect to some norm (on a vector space containing $\mathcal{B}(W_d^{r,\infty}(\mathbb{C}))$) and any map $M : \mathbb{C}^n \to C(\Omega_d; \mathbb{C})$, we have*

$$\sup_{f \in \mathcal{B}(W_d^{r,\infty}(\mathbb{C}))} \|f - M(\overline{a}(f))\|_{L^\infty(\Omega_d)} \geq c \cdot n^{-r/(2d)}.$$

Remarkably, the lower bound holds in the very general setting where one approximates functions from $\mathcal{B}(W_d^{r,\infty}(\mathbb{C}))$ using a class of functions that is parametrizable by n parameters and under the assumption of continuous weight selection. Hence, even if one considers possibly deep CVNNs with $\asymp n$ parameters, the rate of $n^{-r/(2d)}$ cannot be improved (assuming a continuous weight selection). Moreover, it follows that in the setting of continuous weight selection, CVNNs *are not able* to achieve a better approximation rate than RVNNs.

For the sake of completeness, it should be noted that the statements in [13] are formulated for C^r-functions (instead of Sobolev functions or regularity r). However, the proofs in [13] work similarly for Sobolev functions.

B. The case of unrestricted weight selection

It remains to study the case of *unrestricted* weight selection, i.e., the optimal approximation rate when the continuity of the map that assigns to the Sobolev functions the parameters of the approximating network is *not* assumed.

In the real-valued case, this question was solved in [18, 19] by studying sums of n arbitrary *ridge functions*, where a ridge function is a function of the form $\mathbb{R}^d \ni x \mapsto h(a^T x)$ for some $h : \mathbb{R} \to \mathbb{R}$ and a direction $a \in \mathbb{R}^d$. Essentially, it was shown that the error of approximation can be lower bounded by $n^{-r/(d-1)}$ (up to a fixed multiplicative constant) for the large class of *locally integrable* activation functions. Moreover, it could be proved that there exists a smooth activation function that indeed achieves the optimal approximation rate of $n^{-r/(d-1)}$. Note that this is strictly better than the rate of $n^{-r/d}$ obtained in the case of continuous weight selection.

Obtaining a lower bound for CVNNs requires to study sums of n *multivariate* ridge functions, i.e., sums of n functions of the form $\mathbb{R}^d \ni x \to h(Ax)$, where $A \in \mathbb{R}^{\ell \times d}$ and $h : \mathbb{R}^\ell \to \mathbb{R}$ for some fixed $\ell \in \{1, ..., d-1\}$. The special case $\ell = 2$ then readily implies a lower bound for the approximation using CVNNs. This question was recently solved in [12].

Theorem III.3 ([12, Theorem 1.3(1)]). *Let $d, r \in \mathbb{N}$ with $d \geq 2$. Then there exists a constant $c = c(d, r) > 0$ with the following property: For every $n \in \mathbb{N}$ there exists a function $f \in \mathcal{B}(W_d^{\infty,r}(\mathbb{C}))$ such that for every choice of $\phi \in L^1_{\text{loc}}(\mathbb{C}; \mathbb{C})$, and any shallow CVNN $\Phi : \mathbb{C}^d \to \mathbb{C}$ with activation function ϕ and n hidden neurons we have*

$$\|f - \Phi\|_{L^1(\Omega_d)} \geq c \cdot n^{-r/(2d-2)}.$$

Remarkably, this lower bound does not only hold with respect to L^∞ but even with respect to L^1. As already explained above, the proof relies on the study of sums of multivariate ridge functions and essentially combines ideas from [20] and [18]. In fact, [12, Theorem 1.1] contains a general lower bound for such sums (and not just for CVNNs).

Moreover, the following upper bound is derived.

Theorem III.4 ([12, Theorem 1.3(2)]). *There exists $\phi \in C^\infty(\mathbb{C}; \mathbb{C})$ with the following property: For every $d, r \in \mathbb{N}$ with $d \geq 2$ and $p \in [1, \infty]$ there exists a constant $C = C(d, p, r) > 0$ such that for every $n \in \mathbb{N}$ and every function $f \in \mathcal{B}(W_d^{p,r}(\mathbb{C}))$ there exists a shallow CVNN $\Phi : \mathbb{C}^d \to \mathbb{C}$ with n hidden neurons and activation function ϕ satisfying*

$$\|f - \Phi\|_{L^p(\Omega_d)} \leq C \cdot n^{-r/(2d-2)}.$$

In particular, this shows that shallow CVNNs can achieve a *strictly better* approximation rate than RVNNs, for which at most a rate of $n^{-r/(2d-1)}$ is possible. Thus, while CVNNs are in the setting of continuous weight selection *not* able to improve the approximation rate that RVNNs achieve, they are able to do so if the continuity assumption is dropped.

The proof of Theorem III.4 works by transferring various results from [24, Chapter 5] to the complex-valued setting and by then constructing the activation function that realizes the desired approximation rate in a “piecewise” manner.

Lastly, it should be noted that there exist complex activation functions, for which the rate of $n^{-r/(2d)}$ *cannot* be improved (up to a logarithmic factor), even if the continuity assumption on the weight selection is dropped; see [13, Theorem 4.3].

Acknowledgments

The author thanks Felix Voigtlaender for helpful comments. The author acknowledges support by the German Science Foundation (DFG) in the context of the Emmy Noether junior research group VO 2594/1-1.

References

[1] P. Arena, L. Fortuna, R. Re, and M. G. Xibilia. "Multilayer perceptrons to approximate complex valued functions". In: *International Journal of Neural Systems* 6.04 (1995), pp. 435–446.

[2] P. Arena, L. Fortuna, R. Re, and M. G. Xibilia. "On the capability of neural networks with complex neurons in complex valued functions approximation". In: *1993 IEEE international symposium on circuits and systems*. IEEE. 1993, pp. 2168–2171.

[3] M. Arjovsky, A. Shah, and Y. Bengio. "Unitary evolution recurrent neural networks". In: *International conference on machine learning*. PMLR. 2016, pp. 1120–1128.

[4] J. A. Barrachina, C. Ren, C. Morisseau, G. Vieillard, and J.-P. Ovarlez. "Comparison between equivalent architectures of complex-valued and real-valued neural networks-Application on polarimetric SAR image segmentation". In: *Journal of Signal Processing Systems* 95.1 (2023), pp. 57–66. DOI: 10.1007/s11265-022-01793-0.

[5] A. R. Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information theory* 39.3 (1993), pp. 930–945.

[6] M. Calafà, E. Hovad, A. P. Engsig-Karup, and T. Andriollo. "Physics-Informed Holomorphic Neural Networks (PIHNNs): Solving 2D linear elasticity problems". In: *Computer Methods in Applied Mechanics and Engineering* 432 (2024), p. 117406. DOI: https://doi.org/10.1016/j.cma.2024.117406.

[7] A. Caragea, D. G. Lee, J. Maly, G. Pfander, and F. Voigtlaender. "Quantitative approximation results for complex-valued neural networks". In: *SIAM Journal on Mathematics of Data Science* 4.2 (2022), pp. 553–580.

[8] E. Cole, J. Cheng, J. Pauly, and S. Vasanawala. "Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications". In: *Magnetic Resonance in Medicine* 86.2 (2021), pp. 1093–1109. DOI: 10.1002/mrm.28733.

[9] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.

[10] G. B. Folland. *Real analysis: Modern techniques and their applications*. 2nd ed. Pure and applied mathematics. New York: Wiley, 1999. ISBN: 978-0-471-31716-6.

[11] P. Geuchen, T. Jahn, and H. Matt. "Universal approximation with complex-valued deep narrow neural networks". In: *Constructive Approximation* (June 2025). DOI: 10.1007/s00365-025-09713-8.

[12] P. Geuchen, P. Salanevich, O. Schavemaker, and F. Voigtlaender. "On best approximation by multivariate ridge functions with applications to generalized translation networks". In: *Mathematical Foundations of Machine Learning* 1.1 (July 2025), p. 1. DOI: 10.1007/s44439-025-00002-7.

[13] P. Geuchen and F. Voigtlaender. "Optimal approximation using complex-valued neural networks". In: *Advances in Neural Information Processing Systems* 36 (2024).

[14] L. Kaup and B. Kaup. *Holomorphic functions of several variables: an introduction to the fundamental theory*. Vol. 3. Walter de Gruyter, 2011.

[15] P. Kidger and T. Lyons. "Universal approximation with deep narrow networks". In: *Conference on learning theory*. PMLR. 2020, pp. 2306–2327.

[16] T. Küstner, N. Fuin, K. Hammernik, A. Bustin, H. Qi, R. Hajhosseiny, P. G. Masci, R. Neji, D. Rueckert, R. M. Botnar, and C. Prieto. "CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions". In: *Scientific Reports* 10 (2020), article no. 13710, 13 pp. DOI: 10.1038/s41598-020-70551-8.

[17] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural networks* 6.6 (1993), pp. 861–867.

[18] V. Maiorov. "Best approximation by ridge functions in L p-spaces". In: *Ukrainian Mathematical Journal* 62 (2010), pp. 452–466.

[19] V. Maiorov and A. Pinkus. "Lower bounds for approximation by MLP neural networks". In: *Neurocomputing* 25.1-3 (1999), pp. 81–91.

[20] V. E. Maiorov. "On best approximation by ridge functions". In: *Journal of Approximation Theory* 99.1 (1999), pp. 68–94.

[21] H. N. Mhaskar. "Neural networks for optimal approximation of smooth and analytic functions". In: *Neural computation* 8.1 (1996), pp. 164–177.

[22] J. Park and S. Wojtowytsch. "Qualitative Neural Network Approximation over $\mathbb{R}$ and $\mathbb{C}$: Elementary Proofs for Analytic and Polynomial Activation". In: *Explorations in the Mathematics of Data Science: The Inaugural Volume of the Center for Approximation and Mathematical Data Analytics*. Ed. by S. Foucart and S. Wojtowytsch. Cham: Springer Nature Switzerland, 2024, pp. 31–64. ISBN: 978-3-031-66497-7. DOI: 10.1007/978-3-031-66497-7_3.

[23] P. Petersen and F. Voigtlaender. "Optimal approximation of piecewise smooth functions using deep ReLU neural networks". In: *Neural Networks* 108 (2018), pp. 296–330.

[24] A. Pinkus. *Ridge functions*. Vol. 205. Cambridge University Press, 2015.

[25] Y. Ren, W. Jiang, and Y. Liu. "A new architecture of a complex-valued convolutional neural network for PolSAR image classification". In: *Remote Sensing* 15.19 (2023), article no. 4801, 27 pp. DOI: 10.3390/rs15194801.

[26] W. Rudin. *Real and Complex Analysis*. Higher Mathematics Series. McGraw-Hill Education, 1987. ISBN: 9780070542341.

[27] Z. Shen. "Deep Network Approximation Characterized by Number of Neurons". In: *Communications in Computational Physics* 28.5 (2020), pp. 1768–1811.

[28] J. W. Siegel and J. Xu. "Approximation rates for neural networks with general activation functions". In: *Neural Networks* 128 (2020), pp. 313–321. DOI: 10.1016/j.neunet.2020.05.019.

[29] M. E. Valle, W. L. Vital, and G. Vieira. "Universal approximation theorem for vector- and hypercomplex-valued neural networks". In: *Neural Networks* 180 (2024), p. 106632. DOI: 10.1016/j.neunet.2024.106632.

[30] P. Virtue, S. X. Yu, and M. Lustig. "Better than real: Complex-valued neural nets for MRI fingerprinting". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 3953–3957.

[31] F. Voigtlaender. "The universal approximation theorem for complex-valued neural networks". In: *Applied and computational harmonic analysis* 64 (2023), pp. 33–61.

[32] D. Yarotsky. "Error bounds for approximations with deep ReLU networks". In: *Neural Networks* 94 (2017), pp. 103–114. DOI: 10.1016/j.neunet.2017.07.002.

[33] D. Yarotsky and A. Zhevnerchuk. "The phase diagram of approximation rates for deep neural networks". In: *Advances in neural information processing systems* 33 (2020), pp. 13005–13015.

[34] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin. "Complex-valued convolutional neural network and its application in polarimetric SAR image classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.12 (2017), pp. 7177–7188. DOI: 10.1109/TGRS.2017.2743222.