EJAM
www.ejamjournal.com

Electronic Journal of Applied Mathematics

Research article

Mersin University Publishing
DOI: ejam.20253188
2025, Vol. 3, No. 1, 1–8

Improved Universal Approximation with Neural Networks Studied via Affine-Invariant Subspaces of $L_2(\mathbb{R}^n)$

Samuel Probst[1], **Cornelia Schneider**[2,*]

1. Mathematics Department, Friedrich-Alexander Universität, 91058 Erlangen, Germany
2. Applied Mathematics III, Friedrich-Alexander Universität, 91058 Erlangen, Germany

*Corresponding author

Abstract

We show that there are no non-trivial closed subspaces of $L_2(\mathbb{R}^n)$ that are invariant under invertible affine transformations[1]. We apply this result to neural networks showing that any nonzero $L_2(\mathbb{R})$ function is an adequate activation function in a one hidden layer neural network in order to approximate every function in $L_2(\mathbb{R})$ with any desired accuracy. This generalizes the universal approximation properties of neural networks in $L_2(\mathbb{R})$ related to Wiener's Tauberian Theorems. Our results extend to the spaces $L_p(\mathbb{R})$ with $p > 1$.

Keywords: Neural networks, activation functions, Wiener's Tauberian Theorems, Lebesgue spaces, affine invariant subspaces, Universal approximation theorem
2020 Mathematics Subject Classification: Primary: 68T07, 41A30; Secondary: 46E30, 28A05
Article history: Received 10 Oct 2024; Accepted 10 Mar 2025; Online 15 Mar 2025

1 Introduction and main results

This paper contributes to the ongoing dialogue between classical approximation theory and neural network theory by providing a more rigorous mathematical foundation for understanding how neural networks approximate L_2 functions. The result aligns with Wiener's classical framework while introducing new implications for modern machine learning architectures.

One of the foundational results regarding approximation theory for neural networks was the *Universal Approximation Theorem* of [5] and [7], which establishes that a feedforward neural network with a single hidden layer and a continuous, non-polynomial activation function

Contact: Samuel Probst ✉ samuel.probst@fau.de; Cornelia Schneider ✉ schneider@math.fau.de

1 See the introduction for a precise definition.

can approximate any continuous function on a compact subset of $\mathbb{R}^n$, given sufficiently many neurons. Subsequent extensions of this result addressed the analogous problem for Lebesgue spaces with a finite exponent [8] and locally integrable spaces [13]. Building upon these approximation theories, Wiener's Tauberian Theorems provide another vital tool in understanding function approximation since they offer criteria for when a set of translates of a function can densely span a function space like $L_1(\mathbb{R})$ or $L_2(\mathbb{R})$. Specifically, Wiener's Tauberian Theorem states that for a function $f \in L_1(\mathbb{R})$, the translates of f span $L_1(\mathbb{R})$ if and only if the Fourier transform of f has no zeros. As was already pointed out in [5], this implies that the set of all one-hidden layer neural networks of the form $\sum_{i=1}^{N} \lambda_i f(x - \theta_i)$ with $N \in \mathbb{N}$, $\lambda_i, \theta_i \in \mathbb{R}$, is dense in $L_1(\mathbb{R})$ if the activation function $f \in L_1(\mathbb{R})$ has Fourier transform $\widehat{f}(x) \neq 0$. This result can be extended to $L_2(\mathbb{R})$, if one assumes that the Fourier transform of f has only zeros on a set of Lebesgue measure zero – making use of Wiener's Tauberian Theorem for $L_2(\mathbb{R})$. As one observes, all previous (universal) approximation results in L_p-spaces have in common that the choice of suitable activations functions is somehow restricted. Let us mention in this context that a very nice overview of the state of the art concerning universal approximation and suitable activation functions can be found in [11, Table 3.1]. The main result of this paper can be considered as a generalization of Wiener's Tauberian Theorem in $L_2(\mathbb{R})$. We show that one can drop the assumption on the Fourier-transform completely and use any $f \in L_2(\mathbb{R}) \setminus \{0\}$ in order to approximate any function in $L_2(\mathbb{R})$:

Theorem 1.1 (Neural network approximation in $L_2(\mathbb{R})$). *The set of all one-hidden layer neural networks with activation function $f \in L_2(\mathbb{R})$ with $f \neq 0$, i.e.,*

$$\sum_{i=1}^{N} \lambda_i f(\alpha_i x - \theta_i) \quad \text{with} \quad N \in \mathbb{N},\ \lambda_i \in \mathbb{R},\ \theta_i \in \mathbb{R},\ \alpha_i \in \mathbb{R} \setminus \{0\},$$

is dense in $L_2(\mathbb{R})$.

We will deduce the above theorem from a more general statement about affine-invariant subspaces of $L_2(\mathbb{R}^n)$. We say a subspace $U \subseteq L_2(\mathbb{R}^n)$ is *affine-invariant* if $f \in U$ implies $f \circ A \in U$ for any invertible affine map $A : \mathbb{R}^n \to \mathbb{R}^n$ (i.e., A is a function of the form $x \mapsto Ax := Mx + v$ for some invertible $n \times n$ matrix M ($\det M \neq 0$) and vector $v \in \mathbb{R}^n$). In particular, we are able to show:

Theorem 1.2 (Affine-invariant subspaces in $L_2(\mathbb{R}^n)$). *Let $U \subseteq L_2(\mathbb{R}^n)$ be a closed affine-invariant subspace. Then either $U = L_2(\mathbb{R}^n)$ or $U = \{0\}$.*

We remark in this context that the similar problem of translation and dilation invariant subspaces has recently been investigated in [1]. In particular, the results from [1, Thm. 6.1] imply that when $n = 1$, Theorem 1.2 (and therefore also Theorem 1.1) remain even true in $L_p(\mathbb{R})$ for arbitrary $p \in (1, \infty)$. However, our method of proof is completely different and more elementary compared to [1] and of interest on its own.

Finally, one observes that Theorem 1.1 is only formulated in dimension $n = 1$, while Theorem 1.2 holds in general dimensions. This is because shallow neural networks as considered in Theorem 1.1 never belong to $L_2(\mathbb{R}^n)$ in dimension $n > 1$; see [20, Thm. 6.1] in this context.

There are several other significant contributions in this field of neural networks and approximation theory which deserve to be mentioned. In the last years, many directions have been explored for both shallow and deep neural networks. For example, in [2] negative results were studied, in [21] the universal approximation theorem for complex-valued neural networks was investigated, whereas [4] deals the problem of approximating any function in a variable Lebesgue space using one-hidden layer neural networks. For neural networks with ReLU activation function, there are various recent papers on similar topics such as approximation for

Besov spaces [19], [18], regression [17], optimization problems [6], and estimates for the errors obtained in the approximation [14], [22]. Furthermore, for more regular activation functions there are also several recent articles such as [3], [9], [12] and many others, which focus on deep neural networks.

2 Affine-invariant subspaces in $L_2(\mathbb{R}^n)$

Our goal in this section is to prove Theorem 1.2, which states that the only closed subspaces of $L_2(\mathbb{R}^n)$ that are affine-invariant are $L_2(\mathbb{R}^n)$ itself and the trivial subspace $\{0\}$. In what follows we reduce Theorem 1.2 to a purely measure-theoretic statement (Theorem 2.5) using the following generalization of Wiener's Tauberian theorem for $L_2(\mathbb{R}^n)$.

Theorem 2.1. *Let $U \subseteq L_2(\mathbb{R}^n)$ be a closed translation-invariant subspace. Then there exists $E \subseteq \mathbb{R}^n$ measurable such that*

$$U = \mathcal{U}(E) := \left\{ f \in L_2(\mathbb{R}^n) : \widehat{f}|_E = 0 \right\}.$$

Proof. See [10]. □

Here U being translation-invariant means that $f(x) \in U$ implies $f(x-\theta) \in U$ for all $\theta \in \mathbb{R}^n$ and $\widehat{f}$ denotes the Fourier transform of f. Note that in particular U is translation-invariant if it is affine-invariant.

Let $\mathcal{M}$ denote the set of (Lebesgue) measurable subsets of $\mathbb{R}^n$ and $\mathcal{L}^n \colon \mathcal{M} \to [0,\infty]$ the n-dimensional Lebesgue measure. We now embark on the proof of Theorem 2.5. The proof relies on the Lebesgue density theorem, which we briefly state for the reader's convenience.

Theorem 2.2 (Lebesgue density theorem). *Let $E \in \mathcal{M}$. Then for almost all $x \in E$*

$$\lim_{r \searrow 0} \frac{\mathcal{L}^n(B_r(x) \cap E)}{\mathcal{L}^n(B_r(x))} = 1,$$

where $B_r(x)$ denotes the open ball in $\mathbb{R}^n$ around x with radius $r > 0$.

Proof. See [15, Theorem 7.7]. □

To deduce Theorem 2.5 from this we first need two lemmas.

Lemma 2.3. *Let $k \geq 3$ and $M_1, ..., M_k \in \mathcal{M}$. If $M_2, ..., M_{k-1}$ have finite measure then*

$$\mathcal{L}^n(M_1 \cap M_k) \geq \sum_{i=1}^{k-1} \mathcal{L}^n(M_i \cap M_{i+1}) - \sum_{i=2}^{k-1} \mathcal{L}^n(M_i).$$

Proof. The proof follows by induction on k: To start, let $k = 3$. Then we have

$$\begin{aligned}
\mathcal{L}^n(M_1 \cap M_2) &= \mathcal{L}^n(M_1 \cap M_2 \cap M_3) + \mathcal{L}^n(M_1 \cap M_2 \cap M_3^{\complement}), \\
\mathcal{L}^n(M_2 \cap M_3) &= \mathcal{L}^n(M_1 \cap M_2 \cap M_3) + \mathcal{L}^n(M_1^{\complement} \cap M_2 \cap M_3), \\
\mathcal{L}^n(M_2) &\geq \mathcal{L}^n(M_1 \cap M_2 \cap M_3) + \mathcal{L}^n(M_1^{\complement} \cap M_2 \cap M_3) + \mathcal{L}^n(M_1 \cap M_2 \cap M_3^{\complement}),
\end{aligned}$$

and because $M_1 \cap M_2 \cap M_3$ has finite measure we see that

$$\mathcal{L}^n(M_2) \geq \mathcal{L}^n(M_1 \cap M_2) + \mathcal{L}^n(M_2 \cap M_3) - \mathcal{L}^n(M_1 \cap M_2 \cap M_3).$$

Therefore, because M_2 has finite measure, we obtain

$$\mathcal{L}^n(M_1 \cap M_3) \geq \mathcal{L}^n(M_1 \cap M_2 \cap M_3) \geq \mathcal{L}^n(M_1 \cap M_2) + \mathcal{L}^n(M_2 \cap M_3) - \mathcal{L}^n(M_2),$$

which shows the claim for $k = 3$.
Now let $k > 3$ and suppose the lemma is true for all $m < k$. Let $M_1, ..., M_k$ be measurable and let $M_2, ..., M_{k-1}$ have finite measure. Using the inductive hypothesis on $M_1, ..., M_{k-1}$ and M_1, M_{k-1}, M_k we obtain

$$\mathcal{L}^n(M_1 \cap M_{k-1}) \geq \sum_{i=1}^{k-2} \mathcal{L}^n(M_i \cap M_{i+1}) - \sum_{i=2}^{k-2} \mathcal{L}^n(M_i),$$
$$\mathcal{L}^n(M_1 \cap M_k) \geq \mathcal{L}^n(M_1 \cap M_{k-1}) + \mathcal{L}^n(M_{k-1} \cap M_k) - \mathcal{L}^n(M_{k-1}).$$

This yields

$$\begin{aligned}\mathcal{L}^n(M_1 \cap M_k) &\geq \mathcal{L}^n(M_1 \cap M_{k-1}) + \mathcal{L}^n(M_{k-1} \cap M_k) - \mathcal{L}^n(M_{k-1}) \\ &\geq \sum_{i=1}^{k-2} \mathcal{L}^n(M_i \cap M_{i+1}) - \sum_{i=2}^{k-2} \mathcal{L}^n(M_i) + \mathcal{L}^n(M_{k-1} \cap M_k) - \mathcal{L}^n(M_{k-1}) \\ &= \sum_{i=1}^{k-1} \mathcal{L}^n(M_i \cap M_{i+1}) - \sum_{i=2}^{k-1} \mathcal{L}^n(M_i),\end{aligned}$$

which completes the proof. □

For a field K let $\mathrm{GL}(K, n)$ denote the set of invertible $n \times n$ matrices with entries in K. Moreover let $\|\cdot\|_{op}$ denote the operator norm induced by the 2-norm $\|\cdot\|_2$.

Lemma 2.4. *Let $\epsilon > 0$, $a, b \in \mathbb{R}^n \setminus \{0\}$ and $r_a, r_b > 0$ such that $\frac{\|b\|_2}{\|a\|_2} = \frac{r_b}{r_a}$. Then there exists $M \in \mathrm{GL}(\mathbb{Q}, n)$ such that*

(i) $|\mathcal{L}^n(M(B_{r_a}(a))) - \mathcal{L}^n(B_{r_b}(b))| < \epsilon \cdot \mathcal{L}^n(B_{r_b}(b))$,
(ii) $|\mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b}(b)) - \mathcal{L}^n(B_{r_b}(b))| < 2\epsilon \cdot \mathcal{L}^n(B_{r_b}(b))$.

Proof. We start by showing there exists a matrix $N \in \mathrm{GL}(\mathbb{R}, n)$ such that $N(B_{r_a}(a)) = B_{r_b}(b)$: Let $a_1 = \frac{a}{\|a\|_2}$ and $b_1 = \frac{b}{\|b\|_2}$. Then using the Gram-Schmidt process we can find orthonormal bases $(a_1, ..., a_n)$ and $(b_1, ..., b_n)$ containing a_1 and b_1 respectively. Now let A be the orthogonal matrix with the a_i as columns and let B be the orthogonal matrix with the b_i as columns. Then $BA^T a_1 = b_1$ and thus $\frac{\|b\|_2}{\|a\|_2} BA^T a = b$. Hence putting $N := \frac{\|b\|_2}{\|a\|_2} BA^T$ we found an invertible matrix that maps a to b. Moreover, the condition $\frac{\|b\|_2}{\|a\|_2} = \frac{r_b}{r_a}$ together with the fact that N only rotates and scales and thus maps spheres to spheres ensures that $N(B_{r_a}(a)) = B_{r_b}(b)$.
Next we can choose a matrix that is sufficiently close to N as our matrix M: Because the determinant is a continuous map there exists $\delta_1 > 0$ such that

$$\|M - N\|_{op} < \delta_1 \implies |\det(M) - \det(N)| < \frac{\epsilon \cdot \mathcal{L}^n(B_{r_b}(b))}{\mathcal{L}^n(B_{r_a}(a))} \tag{1}$$

for all $M \in \mathrm{GL}(\mathbb{R}, n)$. Furthermore there exists $\delta_2 > 0$ such that

$$\frac{\mathcal{L}^n(B_{r_b+\delta_2}(b) \setminus B_{r_b}(b))}{\mathcal{L}^n(B_{r_b}(b))} = \frac{(r_b + \delta_2)^n - r_b^n}{r_b^n} < \epsilon. \tag{2}$$

Finally choose $R > 0$ such that $B_{r_a}(a) \subseteq B_R(0)$ and define $\delta := \min\{\delta_1, \frac{\delta_2}{R}\}$. Then, because $\mathbb{Q}$ is dense in $\mathbb{R}$ and because the set of invertible matrices is open, there exists $M \in \mathrm{GL}(\mathbb{Q}, n)$ such that $\|M - N\|_{op} < \delta$. Now we have

$$|\mathcal{L}^n(M(B_{r_a}(a))) - \mathcal{L}^n(B_{r_b}(b))| = \big||\det(M)| - |\det(N)|\big| \cdot \mathcal{L}^n(B_{r_a}(a)) \leq \epsilon \cdot \mathcal{L}^n(B_{r_b}(b))$$

by (1), which shows that M satisfies (i).
Next let $x \in B_{r_a}(a)$. Then

$$\|Mx - b\|_2 \leq \|Mx - Nx\|_2 + \|Nx - b\|_2 \leq \|M - N\|_{op} \cdot \|x\|_2 + r_b \leq \delta_2 \cdot \frac{\|x\|_2}{R} + r_b \leq \delta_2 + r_b,$$

which shows that $M(B_{r_a}(a)) \subseteq B_{r_b+\delta_2}(b)$.
Thus $M(B_{r_a}(a))$ is the disjoint union of $M(B_{r_a}(a)) \cap B_{r_b}(b)$ and $M(B_{r_a}(a)) \cap B_{r_b+\delta_2}(b) \setminus B_{r_b}(b)$ and therefore using (2) we have

$$\begin{aligned}\mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b}(b)) &= \mathcal{L}^n(M(B_{r_a}(a))) - \mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b+\delta_2}(b) \setminus B_{r_b}(b)) \\ &\geq \mathcal{L}^n(M(B_{r_a}(a))) - \mathcal{L}^n(B_{r_b+\delta_2}(b) \setminus B_{r_b}(b)) \\ &\geq \mathcal{L}^n(M(B_{r_a}(a))) - \epsilon \cdot \mathcal{L}^n(B_{r_b}(b)).\end{aligned}$$

Finally using the fact that M satisfies (i) we obtain

$$\begin{aligned}|\mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b}(b)) - \mathcal{L}^n(B_{r_b}(b))| &= \mathcal{L}^n(B_{r_b}(b)) - \mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b}(b)) \\ &\leq \mathcal{L}^n(B_{r_b}(b)) - \mathcal{L}^n(M(B_{r_a}(a))) + \epsilon \cdot \mathcal{L}^n(B_{r_b}(b)) \\ &\leq |\mathcal{L}^n(B_{r_b}(b)) - \mathcal{L}^n(M(B_{r_a}(a)))| + \epsilon \cdot \mathcal{L}^n(B_{r_b}(b)) \\ &\leq 2\epsilon\mathcal{L}^n(B_{r_b}(b)),\end{aligned}$$

showing that M also satisfies (ii). □

We are now ready to prove Theorem 2.5. For $M, N \in \mathcal{M}$ we write $M \overset{\Delta}{=} N$ if M and N only differ on a set of measure zero, that is if $\mathcal{L}^n((M \setminus N) \cup (N \setminus M)) = 0$. Moreover for $A \subseteq \mathbb{R}^n$ we define $\mathrm{GL}(\mathbb{Q}, n) \cdot A := \{Ma : M \in \mathrm{GL}(\mathbb{Q}, n), a \in A\}$.

Theorem 2.5. *Let $A \in \mathcal{M}$ with positive measure. Then*

$$\mathrm{GL}(\mathbb{Q}, n) \cdot A \overset{\Delta}{=} \mathbb{R}^n.$$

Proof. Let $\epsilon := 0.1$ and $B := (\mathrm{GL}(\mathbb{Q}, n) \cdot A)^{\complement}$. Suppose B had positive measure. Then by the Lebesgue density theorem there would exist (nonzero) points of density $a \in A, b \in B$ for A and B. Hence there would exist $R > 0$ such that

$$\begin{aligned}\mathcal{L}^n(A \cap B_r(a)) &\geq (1 - \epsilon) \cdot \mathcal{L}^n(B_r(a)), \\ \mathcal{L}^n(B \cap B_r(b)) &\geq (1 - \epsilon) \cdot \mathcal{L}^n(B_r(b))\end{aligned}$$

for all $r < R$. Now choose real numbers r_a and r_b satisfying $0 < r_a, r_b < R$ and $r_b \cdot \|a\|_2 = r_a \cdot \|b\|_2$. Then by Lemma 2.4 there exists $M \in \mathrm{GL}(\mathbb{Q}, n)$ such that

$$\begin{aligned}|\mathcal{L}^n(M(B_{r_a}(a))) - \mathcal{L}^n(B_{r_b}(b))| &\leq \epsilon \cdot \mathcal{L}^n(B_{r_b}(b)), \\ |\mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b}(b)) - \mathcal{L}^n(B_{r_b}(b))| &\leq 2\epsilon \cdot \mathcal{L}^n(B_{r_b}(b)).\end{aligned}$$

In particular, we have

$$\begin{aligned}\mathcal{L}^n(M(B_{r_a}(a)) \cap B_{r_b}(b)) &\geq (1 - 2\epsilon)\mathcal{L}^n(B_{r_b}(b)), \\ \mathcal{L}^n(M(B_{r_a}(a))) &\leq (1 + \epsilon)\mathcal{L}^n(B_{r_b}(b)).\end{aligned}$$

Thus using Lemma 2.3 we obtain

$$\begin{aligned}\mathcal{L}^n(B \cap M(A)) &\geq \mathcal{L}^n(B \cap B_{r_b}(b)) + \mathcal{L}^n(B_{r_b}(b) \cap M(B_{r_a}(a))) + \mathcal{L}^n(M(B_{r_a}(a)) \cap M(A)) \\ &\quad - \mathcal{L}^n(B_{r_b}(b)) - \mathcal{L}^n(M(B_{r_a}(a))) \\ &\geq (1-\epsilon)\mathcal{L}^n(B_{r_b}(b)) + (1-2\epsilon)\mathcal{L}^n(B_{r_b}(b)) + \det(M)(1-\epsilon)\mathcal{L}^n(B_{r_a}(a)) \\ &\quad - \mathcal{L}^n(B_{r_b}(b)) - \det(M)\mathcal{L}^n(B_{r_a}(a)) \\ &= (1-3\epsilon)\mathcal{L}^n(B_{r_b}(b)) - \epsilon \cdot \mathcal{L}^n(M(B_{r_a}(a))) \\ &\geq (1-3\epsilon-\epsilon(1+\epsilon))\mathcal{L}^n(B_{r_b}(b)) = (1-4\epsilon-\epsilon^2)\mathcal{L}^n(B_{r_b}(b)) > 0\end{aligned}$$

contradicting the definition of B as the complement of $\mathrm{GL}(\mathbb{Q}, n) \cdot A$. □

Having established Theorem 2.5 we only need one more small lemma in order to be able to proof Theorem 1.2. In what follows we use the following convention: We say that $f \in L_2(\mathbb{R}^n)$ has no zeros in $E \in \mathcal{M}$ if for some (and thus every) representative φ_f of f it holds that $\mathcal{L}^n(\varphi_f^{-1}(\{0\}) \cap E) = 0$.

Lemma 2.6. *Let $E \in \mathcal{M}$, $f \in \mathcal{U}(E) = \{g \in L_2(\mathbb{R}^n) : \widehat{g}|_E = 0\}$, and $M \in \mathcal{M}$ such that $\widehat{f}$ has no zeros in M. Then $\mathcal{L}^n(M \cap E) = 0$.*

Proof. Because $f \in \mathcal{U}(E)$ we have $\widehat{f}|_{M \cap E} = 0$. If $M \cap E$ had nonzero measure this would contradict $\widehat{f}$ having no zeros in M. □

With all this preparation, we can now finally characterize affine-invariant subspaces in $L_2(\mathbb{R}^n)$.

Proof. (of Theorem 1.2): Suppose $U \neq \{0\}$. Then there exists some $f \in U$ with $f \neq 0$ and thus $\widehat{f} \neq 0$ by Plancherel's Theorem (cf. [16, Theorem 1.6.1]) and there exists $A \in \mathcal{M}$ with positive measure such that $\widehat{f}$ has no zeros in A (take for example the complement of the zero set of some representative of $\widehat{f}$). Then by Theorem 2.5 we have $\mathrm{GL}(\mathbb{Q}, n) \cdot A \overset{\Delta}{=} \mathbb{R}^n$. Moreover, for any $M \in \mathrm{GL}(\mathbb{Q}, n)$ the Fourier transform of $f \circ M$ has no zeros in $M^T(A)$. This is because we have

$$\widehat{f \circ M}(x) = \frac{1}{|\det(M)|} \cdot \widehat{f}(M^{-T}x) \quad \text{a.e.},$$

which can be verified directly using the definition of the Fourier transform.

Thus as U is affine-invariant we have $f \circ M \in U$ and Lemma 2.6 yields $\mathcal{L}^n(M^T(A) \cap E) = 0$ for any $M \in \mathrm{GL}(\mathbb{Q}, n)$, where $E \in \mathcal{M}$ is such that $\mathcal{U}(E) = U$ (see Theorem 2.1). Then

$$\mathcal{L}^n(E) = \mathcal{L}^n(E \cap (\mathrm{GL}(\mathbb{Q}, n) \cdot A)) = \mathcal{L}^n\left(\bigcup_{M \in \mathrm{GL}(\mathbb{Q}, n)} (E \cap M^T(A))\right) \leq \sum_{M \in \mathrm{GL}(\mathbb{Q}, n)} \mathcal{L}^n(E \cap M^T(A)) = 0,$$

which implies

$$U = \mathcal{U}(E) = \{g \in L_2(\mathbb{R}^n) : \widehat{g}|_E = 0\} = L_2(\mathbb{R}^n),$$

because all functions are equivalent to the zero function when restricted to a set of measure zero. □

3 Application to neural networks

We now use the result from Theorem 1.2 when $n = 1$ and investigate when one-hidden layer neural networks are dense in $L_2(\mathbb{R})$.

Proof. (of Theorem 1.1): Note that the neural networks considered in Theorem 1.1 form an affine-invariant subspace of $L_2(\mathbb{R})$. Thus, if for $f \in L_2(\mathbb{R}) \setminus \{0\}$ we let

$$N_f := \left\{ g \in L_2(\mathbb{R}) : \ g(x) = \sum_{i=1}^{N} \lambda_i f(\alpha_i x - \theta_i), \ N \in \mathbb{N}, \ \lambda_i, \theta_i \in \mathbb{R}, \ \alpha_i \in \mathbb{R} \setminus \{0\} \right\} \subseteq L_2(\mathbb{R})$$

denote the set of these neural networks, it follows that $\overline{N_f}$ is a closed affine-invariant subspace and Theorem 1.2 implies $\overline{N_f} = L_2(\mathbb{R})$ for nonzero $f \in L_2(\mathbb{R})$. □

Remark 3.1. Most activation functions used in practice are not in $L_2(\mathbb{R})$ and thus Theorem 1.1 does not apply to them directly. However, by Theorem 1.2 it suffices if there is some nonzero L_2 function that can be approximated (using a one-hidden layer neural network with such an activation function) to deduce all L_2 functions can be. Consider for example the ReLU activation function $\mathrm{ReLU}(x) = \max\{0, x\}$. Then the hat function can be represented as

$$(\mathrm{ReLU}(x) - \mathrm{ReLU}(x+1)) - (\mathrm{ReLU}(x+1) - \mathrm{ReLU}(x+2)) = \begin{cases} 0, & x \leq -2 \\ x+2, & -2 \leq x \leq -1 \\ -x, & -1 \leq x \leq 0 \\ 0, & 0 \leq x \end{cases} \in L_2(\mathbb{R}).$$

Hence there exists a nonzero function in N_{ReLU} belonging to $L_2(\mathbb{R})$ and from the structure of N_{ReLU} and Theorem 1.2 we deduce that $\overline{N_{\mathrm{ReLU}}} = L_2(\mathbb{R})$.

4 Declarations

Funding

This research received no external funding.

Competing Interests

The authors declare that they have no competing interests.

Ethical Approval

Not applicable.

Authors's Contributions

All authors contributed equally. All the authors read and approved the final manuscript.

Availability Data and Materials

Not applicable.

Acknowledgements

We thank the two anonymous referees for carefully reading the paper and their comments, which helped to improve the manuscript.

References

[1] A. B. Aleksandrov, *On translation- and dilation-invariant subspaces of the space $L^p(\mathbb{R}^n)$, $0 < p < 1$*, J. Math. Sci. (N.Y.) **148** (2008), no. 6, 785–794.

[2] J. M. Almira, P. E. Lopez de Teruel, D. J. Romero-López, and F. Voigtlaender, *Negative results for approximation using single layer and multilayer feedforward neural networks*, Journal of Mathematical Analysis and Applications **494** (2021), no. 1, 124584.

[3] A. R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Transactions on Information Theory **39** (1993), no. 3, 930–945.

[4] Á. Capel and J. Ocáriz, *Approximation with neural networks in variable lebesgue spaces*, Preprint, arXiv: 2007.04166v1, 2020.

[5] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Systems **2** (1989), 303–314.

[6] S. Giulini and M. Sanguineti, *Approximation schemes for functional optimization problems*, J. Optim. Theory Appl. **140** (2009), 33–54.

[7] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximatiors*, Neural networks **2** (1989), 359–366.

[8] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, Neural networks **4** (1991), no. 3, 251–257.

[9] B. Li, S. Tang, and H. Yu, *Better approximations of high dimensional smooth functions by deep neural networks with rectified power units*, Commun. in Comp. Phys. **27** (2020), 379–411.

[10] J. Ludwig, C. Molitor-Braun, and S. Pusti, *Spectral synthesis in $L^2(G)$*, Colloquium Mathematicum **138** (2015), no. 1, 89–103.

[11] J. Ocáriz, *Universal approximation results on artificial neural networks*, Master thesis, Univ. Rovira i Virgili, 2020.

[12] I. Ohn and Y. Kim, *Smooth function approximation by deep neural networks with general activation functions*, Entropy **21** (2019), no. 7, 627–411.

[13] J. Park and I. W. Sandberg, *Universal approximation using radial-basis-function networks*, Neural computation **3** (1991), no. 2, 246–257.

[14] P. P. Petrushev, *Approximation by ridge function and neural networks*, SIAM Journal on Mathematical Analysis **30** (1999), 155–189.

[15] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill Book Company, 1987.

[16] W. Rudin, *Fourier analysis on groups*, John Wiley & Sons, 1990.

[17] J. Schmidt-Hieber, *Nonparametric regression using deep neural networks with ReLU activation function*, The Annals of Statistics **48** (2020), no. 4, 1875–1897.

[18] J. W. Siegel, *Optimal approximation rates for deep ReLU neural networks on Sobolev and Besov spaces*, Preprint, arXiv:2211.1440v6 (2024).

[19] T. Suzuki, *Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: optimal reate and curse of dimensionality*, 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9 (2019), 785–79.

[20] T. D. H. van Nuland, *Noncompact uniform universal approximation*, Neural Networks **173** (2024), 106181.

[21] F. Voigtlaender, *The universal approximation theorem for complex-valued neural networks*, Appl. and Comput. Harm. Anal. **64** (2023), 33–61.

[22] D. Yarotsky, *Error bounds for approximations with deep ReLU networks*, Neural Networks **94** (2017), 103–114.