Hindawi
Journal of Function Spaces
Volume 2024, Article ID 3262798, 6 pages
https://doi.org/10.1155/2024/3262798

*Research Article*

# Global Universality of the Two-Layer Neural Network with the $k$-Rectified Linear Unit

**Naoya Hatano**[1], **Masahiro Ikeda**[2,3], **Isao Ishikawa**[3,4] and **Yoshihiro Sawano**[1,2,3]

[1]*Department of Mathematics, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan*
[2]*Department of Mathematics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*
[3]*Center for Advanced Intelligence Project, RIKEN, Japan*
[4]*Center for Data Science, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*

Correspondence should be addressed to Yoshihiro Sawano; yoshihiro-sawano@celery.ocn.ne.jp

Received 7 January 2022; Revised 7 February 2023; Accepted 27 November 2023; Published 18 January 2024

Academic Editor: Andrea Scapellato



This paper concerns the universality of the two-layer neural network with the $k$-rectified linear unit activation function with $k = 1, 2, \ldots$ with a suitable norm without any restriction on the shape of the domain in the real line. This type of result is called global universality, which extends the previous result for $k = 1$ by the present authors. This paper covers $k$-sigmoidal functions as an application of the fundamental result on $k$-rectified linear unit functions.

## 1. Introduction

The goal of this note is to specify the closure of linear subspaces generated by the $k$-rectified linear unit functions under various norms. As in [1], for $k \in \mathbb{N}$, we set

$$\mathrm{ReLU}(x) := \begin{cases} 0, & x \le 0, \\ x, & x > 0, \end{cases} \qquad (1)$$
$$\mathrm{ReLU}^k(x) := \mathrm{ReLU}(x)^k.$$

The function $\mathrm{ReLU}^k$ is called the $k$-rectified linear unit ($k$-ReLU for short), which is introduced to compensate for the properties that ReLU does not have. Our approach will be a completely mathematical one. Recently, increasing attention has been paid to the $k$-ReLU function as well as the original ReLU function. For example, if $k \ge 2$, the function $k$-ReLU is in the class $C^{k-1}$, so that it is smoother than the ReLU function. When we study neural networks, the function $k$-ReLU is called an activation function. As in [2], $k$-ReLU functions are used to reduce the amount of computation. Using this smoothness property, Siegel and Xu investigated the error estimates of the approximation [1]. Mhaskar and Micchelli worked in compact sets in $\mathbb{R}^n$, while in the present work, we consider the approximation on the whole real line.

A problem arises when we deal with $k$-ReLU as a function over the whole line. The function $k$-ReLU is not bounded on $\mathbb{R}$. Our goal in this paper is to propose a Banach space that allows us to handle such unbounded functions. Actually, for $k = 1, 2, \ldots$, we let

$$\mathscr{Y}_k(\mathbb{R}) := \left\{ f \in C(\mathbb{R}) : \lim_{x \to \pm\infty} \frac{f(x)}{1 + |x|^k} \text{ exists} \right\}, \qquad (2)$$

equipped with the norm

$$\|f\|_{\mathscr{Y}_k} := \sup_{x \in \mathbb{R}} \frac{|f(x)|}{1 + |x|^k}, \qquad (3)$$

and define

$$H_{\mathrm{ReLU}^k}(\mathbb{R}) := \mathrm{Span}\left( \left\{ \mathrm{ReLU}^k(a \cdot + b) : a \ne 0, b \in \mathbb{R} \right\} \right). \qquad (4)$$

Note that any element in $\mathcal{Y}_k(\mathbb{R})$, divided by $1+|\cdot|^k$, is a continuous function over $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{\pm\infty\}$. Our main result in this paper is as follows:

**Theorem 1.** *The linear subspace $H_{ReLU^k}(\mathbb{R})$ is dense in $\mathcal{Y}_k(\mathbb{R})$.*

Understanding the structure of $H_{\mathrm{ReLU}^k}(\mathbb{R}^n)$ is important in the field machine learning in the last decade. We refer to [4, 5] for example. Furthermore, dealing with unbounded activation functions is important from the viewpoint of application (see [6]). Remark that the approximation over bounded domains has a long history (see [7]).

As is seen from the definition of the norm $\|\cdot\|_{\mathcal{Y}_k}$, when we have a function $f \in \mathcal{Y}_k(\mathbb{R})$, with ease, we can find a function $g \in H_{\mathrm{ReLU}^k}(\mathbb{R})$ such that $\lim_{x \longrightarrow \pm\infty} (|f(x)-g(x)|/1+|x|^k) = 0$. However, after choosing such a function $g$, we have to look for a way to control $f-g$ inside any compact interval by a function $h \in \mathcal{Y}_k(\mathbb{R}) \cap C_c(\mathbb{R})$. Although $\mathcal{Y}_k(\mathbb{R})$ consists of unbounded functions, we can manage to do so by induction on $k$. Actually, we will find $h \in \mathcal{Y}_k(\mathbb{R}) \cap C_c(\mathbb{R})$ such that $f-g-h$ is sufficiently small once we are given a compact interval.

Theorem 1 says that the space $\mathcal{Y}_k(\mathbb{R})$ is mathematically suitable when we consider the activation function $k$-ReLU. We compare Theorem 1 with the following fundamental result by Cybenko. For a function space $X(\mathbb{R})$ over the real line $\mathbb{R}$ and an open set $\Omega$, $X(\Omega)$ stands for the restriction of each element $f$ to $\Omega$, that is,

$$X(\Omega) = \{f|\Omega : f \in X(\mathbb{R})\}, \tag{5}$$

and the norm is given by

$$\|f\|_{X(\Omega)} = \inf \{\|g\|_X : g \in X(\mathbb{R}), f = g|\Omega\}. \tag{6}$$

**Theorem 2** (see Cybenko [8]). *Let $K \subset \Omega$ be a compact set and $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous sigmoidal function. Then, for all $f \in C(K)$ and $\varepsilon > 0$, there exists $g \in H_\sigma(\Omega)$ such that*

$$\sup_{x \in K} |g(x) - f(x)| < \varepsilon. \tag{7}$$

*We remark that Theorem 1 is not a direct consequence of Theorem 2. Theorem 2 concerns the uniform approximation over compact intervals, while Theorem 2 deals with the uniform approximation over the whole real line. We will prove Theorem 1 without using Theorem 2.*

*Let $k = 0, 1, \ldots$. Our results readily can be carried over to the case of k-sigmoidal functions. As in Definition 4.1 in [7], a continuous function $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ is k-sigmoidal if*

$$\lim_{x \longrightarrow -\infty} \frac{\sigma(x)}{x^k} = 0, \quad \lim_{x \longrightarrow \infty} \frac{\sigma(x)}{x^k} = 1. \tag{8}$$

*Needless to say, ReLU$^k$ is k-sigmoidal. If $k = 0$, then we say that $\sigma$ is a continuous sigmoidal. As a corollary of Theorem 1, we extend this theorem to the case of k-sigmoidal.*

**Theorem 3.** *If $\sigma$ is $k$ -sigmoidal, then the linear subspace $H_\sigma(\mathbb{R})$ is dense in $\mathcal{Y}_k(\mathbb{R})$.*

We can transplant Theorem 3 to various Banach lattices over any open set $\Omega$ on the real line $\mathbb{R}$. Here and below, $L^0(\Omega)$ denotes the set of all Lebesgue measurable functions from $\Omega$ to $\mathbb{C}$. Let $X(\Omega)$ be a Banach space contained in $L^0(\Omega)$ endowed with the norm $\|\cdot\|_{X(\Omega)}$. We say that $X(\Omega)$ is a Banach lattice if for any $f \in L^0(\Omega)$ and $g \in X(\Omega)$ satisfying the estimate $|f(x)| \le |g(x)|$, i.e., $x \in \Omega$, $f \in X(\Omega)$, and the estimate $\|f\|_{X(\Omega)} \le \|g\|_{X(\Omega)}$ holds. We refer to [3] for the case where X is the variable exponent Lebesgue spaces. See [9] for the function spaces to which Theorem 1 is applicable.

We write

$$\mathrm{ReLU}_+^k(x) = \mathrm{ReLU}^k(x) = \max(0, x)^k, \quad \mathrm{ReLU}_-^k(x) = \mathrm{ReLU}_+^k(-x). \tag{9}$$

**Theorem 4** (Universality on Banach lattices). *Let $k \in \mathbb{N}$, $\Omega \subset \mathbb{R}$ be an open set. Assume that $\mathcal{Y}_k(\Omega)$ is continuously embedded into $X(\Omega)$. Assume that $\chi_\Omega \in X(\Omega)$. Then,*

$$\overline{H_{ReLU^k}(\Omega)}^{\|\cdot\|_{X(\Omega)}} = \overline{\mathbb{C}(ReLU_+^k|\Omega) + \mathbb{C}(ReLU^k|\Omega) + C_c(\mathbb{R}^n)|\Omega}^{\|\cdot\|_{X(\Omega)}}. \tag{10}$$

It is noteworthy that we can deal with the case of $\Omega = \mathbb{R}$.

*Remark 5.*

(1) The condition that $\chi_\Omega \in X(\Omega)$ is a natural condition, since $\sigma \in X(\Omega)$

(2) If $k = 0$, then we saw in [9] that our result recaptures the result by Funahashi [10]. So, our result includes a further extension of his result

*Remark 6.* Let $X(\Omega)$ be a Banach lattice, and let $\sigma$ be a 1 -sigmoidal. We put

$$\sigma_0(x) \equiv \mathrm{ReLU}(x) - \mathrm{ReLU}(x-1), x \in \mathbb{R}. \tag{11}$$

Then, by the result for the case of $k = 1$,

$$\overline{H_{\sigma_0}(\Omega)}^{\|\cdot\|_{X(\Omega)}} = \overline{H_\sigma(\Omega)}^{\|\cdot\|_{X(\Omega)}}. \tag{12}$$

## 2. Proof of Theorem 1

We need the following lemmas: we embed $\mathcal{Y}_k(\mathbb{R})$ into a function space over $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.

**Lemma 7.** *The operator $\mathcal{Y}_k(\mathbb{R}) \longrightarrow BC(\overline{\mathbb{R}})$, $f \mapsto f/1+|\cdot|^k$, is an isomorphism.*

If $k = 1$, then this can be found in Lemma 3 in [9].

*Proof.* Observe that the inverse is given for $F \in BC(\overline{\mathbb{R}})$ as follows:

$$f(x) = (1 + |x|)^k F(x) \quad (x \in \mathbb{R}). \tag{13}$$

Since the operator $\mathscr{Y}_k(\mathbb{R}) \longrightarrow BC(\overline{\mathbb{R}})$, $f \mapsto f/1 + |\cdot|^k$, preserves the norms, we see that this operator is an isomorphism.

We set

$$H^+_{\mathrm{ReLU}^k}(\mathbb{R}) = \mathrm{Span}\left\{\mathrm{ReLU}^k(\cdot - t) \colon t \in \mathbb{R}\right\}. \tag{14}$$

We will use the following algebraic relation for $H^+_{\mathrm{ReLU}^k}(\mathbb{R})$. □

**Lemma 8.** *Let $k \in \mathbb{N}$. Then, for all $x \in \mathbb{R}$,*

$$\sum_{j=0}^{k+1} \binom{k+1}{j} (-1)^j (x-j)^k = 0, \sum_{j=0}^{k} \binom{k}{j} (-1)^j (x-j)^k = (-1)^k. \tag{15}$$

*Proof of Lemma 8.* By comparing the coefficients, we may reduce the matter to the proof of the following two equalities:

$$\sum_{j=0}^{k} \binom{k+1}{j} (-1)^j j^\ell = 0, \sum_{j=0}^{k} \binom{k+1}{j} (-1)^j j^k (-1)^k, \tag{16}$$

for each $\ell = 0, 1 \ldots, k-1$. We compute

$$\left(e^t - 1\right)^k = \sum_{j=1}^{k} \binom{k}{j} (-1)^{k-j} e^{jt}, \tag{17}$$

and then,

$$t^k \left( \sum_{\ell=0}^{\infty} \frac{t^\ell}{(\ell+1)!} \right)^k = (-1)^k \sum_{\ell=0}^{\infty} \left[ \sum_{j=1}^{k} \binom{k}{j} (-1)^j j^\ell \right] \frac{t^\ell}{\ell!}. \tag{18}$$

Hence,

$$\begin{aligned} &\sum_{j=0}^{k} \binom{k+1}{j} (-1)^j j^\ell = 0, \\ &\sum_{j=0}^{k} \binom{k+1}{j} (-1)^j j^k = (-1)^k \end{aligned} \tag{19}$$

for each $\ell = 0, 1 \ldots, k-1$. □

Although $\mathrm{ReLU}^k$ is unbounded, if we consider suitable linear combinations, we can approximate any function in $_c(\mathbb{R})$.

**Lemma 9.** *Any function in $C_c(\mathbb{R})$ can be approximated uniformly over $\mathbb{R}$ by the functions in $H_{ReLU^k}(\mathbb{R}) \cap C_c(\mathbb{R})$. More precisely, if a function $f \in C_c(\mathbb{R})$ is contained in an interval $[-a', a]$ and $\varepsilon > 0$, then there exists $\tau \in H_{ReLU^k}(\mathbb{R})$ such that $\mathrm{supp}\,(\tau) \subset [-a', a]$ and that $\|\tau - f\|_{L^\infty} \le C\varepsilon$.*

For the proof, we will use the following observation: if

$$f = \sum_{j=1}^{N} a_j \mathrm{ReLU}^{k-1}\left(\cdot - t_j\right), \tag{20}$$

then, by the definition of $\mathrm{ReLU}^k$,

$$\int_{-\infty}^{t} f(s) ds = \sum_{j=1}^{N} \frac{a_j}{k} \mathrm{ReLU}^k\left(\cdot - t_j\right). \tag{21}$$

*Proof.* We induct on $k$. The base case $k = 1$ was proved already [9]. Suppose that we have $f \in C_c(\mathbb{R})$ with $\mathrm{supp}\,(f) \subset [-a', a]$ for $a, a' > 0$. In fact, we can approximate $f$ with the functions in $H_{\mathrm{ReLU}^k}(\mathbb{R})$ supported in $[-a', a]$. Let $\varepsilon > 0$ be given. By mollification and dilation, we may assume $f \in C^1(\mathbb{R})$. By the induction assumption, there exists $\psi \in H_{\mathrm{ReLU}^{k-1}}(\mathbb{R})$ such that

$$\left\|f' - \psi\right\|_{L^\infty} < (1 + \ell)^{-1} \varepsilon, \mathrm{supp}\,(\psi) \subset \left[-a', a\right], \tag{22}$$

where $\ell = a + a' = \mathrm{diam}(\mathrm{supp}\,(f))$. Note that

$$\varphi(t) = \int_{-\infty}^{t} \psi(s) ds \quad (t \in \mathbb{R}) \tag{23}$$

is a function in $H_{\mathrm{ReLU}^k}(\mathbb{R})$. Note that

$$\varphi(t) = 0 \quad \text{if} \quad t \le -a', \varphi(t) = \int_{\mathbb{R}} \psi(s) ds = \varphi(a) \quad \text{if} \quad t \ge a. \tag{24}$$

Integrating estimate (22), we obtain

$$\begin{aligned} |f(t) - \varphi(t)| &\le \int_{-a'}^{t} \left|f'(s) - \varphi'(s)\right| ds \\ &\le \int_{-a'}^{a} \left|f'(s) - \varphi'(s)\right| ds \\ &\le \frac{\left(a + a'\right)\varepsilon}{1 + \ell} < \varepsilon, \end{aligned} \tag{25}$$

for $t \ge -a'$. In particular,

$$\varphi(a) = \varphi(a) - f(a) \in (-\varepsilon, \varepsilon). \tag{26}$$

Thus, $\|f - \varphi\|_{L^\infty} < \varepsilon$. Using Lemma 8, the dilation and translation, we choose $\varphi^* \in H_{\mathrm{ReLU}^k}(\mathbb{R})$, which depends on $k$, $a$, and $a'$, such that $\mathrm{supp}\,(\varphi^*) \subset (-a', \infty)$ and that $\varphi^*$

agrees with 1 over $(a, \infty)$. If $t < -a'$, then for $\tau = \varphi - \varphi(a)\varphi^*$,

$$f(t) - \tau(t) = f(t) - \varphi(t) \in (-\varepsilon, \varepsilon). \tag{27}$$

If $-a' < t < a$, then

$$\begin{aligned} f(t) - \tau(t) &= f(t) - \varphi(t) + \varphi(a)\varphi^*(t) \\ &\cdot \in \left(-(1 + \|\varphi^*\|_{L^\infty})\varepsilon, (1 + \|\varphi^*\|_{L^\infty})\varepsilon\right). \end{aligned} \tag{28}$$

Finally, if $t > a$, then

$$f(t) - \tau(t) = f(t) - \varphi(t) + \varphi(a)\varphi^*(t) = f(t) - \varphi(t) + \varphi(a) = 0. \tag{29}$$

Therefore, the function $\tau$ is a function in $H_{\mathrm{ReLU}^k}(\mathbb{R})$ satisfying supp $(\tau) \subset [-a', a]$ and $\|f - \tau\|_{L^\infty} < C\varepsilon$, where $C$ depends on $k$, $a$, and $a'$, that is, $k$ and $f$. □

We will prove Theorems 1 and 3.

*Proof of Theorem 1.* We identify $\mathscr{Y}_k$ with $\mathrm{BC}(\overline{\mathbb{R}})$ as in Lemma 7. We have to show that any finite Borel measure $\mu$ in $\overline{\mathbb{R}}$ which annihilates $H_{\mathrm{ReLU}^k}(\mathbb{R})$ is zero. Since $C_c(\mathbb{R})$ is contained in the closure of the space $H_{\mathrm{ReLU}^k}(\mathbb{R})$ as we have seen in Lemma 9, $\mu$ is not supported on $\mathbb{R}$. Therefore, we have only to show that $\mu(\{\infty\}) = 0$ and that $\mu(\{-\infty\}) = 0$. However, since we have shown that $\mu$ is not supported on $\mathbb{R}$, this is a direct consequence of the following observations:

$$\begin{aligned} \mu(\{\infty\}) &= \int_{\overline{\mathbb{R}}} \frac{\mathrm{ReLU}^k(t)}{1 + |t|^k} d\mu(t) = 0, \\ \mu(\{-\infty\}) &= (-1)^k \int_{\overline{\mathbb{R}}} \frac{\mathrm{ReLU}^k(-t)}{1 + |t|^k} d\mu(t) = 0. \end{aligned} \tag{30}$$

Thus, $\mu = 0$. □

*Proof of Theorem 3.* We identify $\mathscr{Y}_k$ with $\mathrm{BC}(\overline{\mathbb{R}})$ as in Lemma 7 once again. Then to show that

$$H_\sigma(\mathbb{R}) = \mathrm{Span}\left\{\frac{\sigma(ax - b)}{1 + |x|^k} : a, b \in \mathbb{R}\right\} \tag{31}$$

is dense in $\mathrm{BC}(\overline{\mathbb{R}})$ under this identification, it suffices to show that any finite measure $\mu$ over $\overline{\mathbb{R}}$ is zero if it annihilates $H_\sigma(\mathbb{R})$.

Assuming that $\mu$ annihilates $H_\sigma(\mathbb{R})$, we see that

$$\int_{\overline{\mathbb{R}}} \frac{\sigma(ax - ab)}{1 + |x|^k} d\mu(x) = 0, \tag{32}$$

for any $a \neq 0$ and $b \in \mathbb{R}$. Since $\sigma$ is $k$-sigmoidal,

$$\sup_{x \in \mathbb{R}^n} \sup_{a \in \mathbb{R} \setminus [-1,1]} \left| \frac{\sigma(ax - ab)}{a^k \left(1 + |x|^k\right)} \right| < \infty, \tag{33}$$

for any fixed $b \in \mathbb{R}$. Furthermore,

$$\begin{aligned} \lim_{a \longrightarrow \infty} \frac{\sigma(ax - ab)}{|a|^k \left(1 + |x|^k\right)} &= \frac{(x - b)_+^k}{1 + |x|^k}, \\ \lim_{a \longrightarrow -\infty} \frac{\sigma(ax - ab)}{|a|^k \left(1 + |x|^k\right)} &= \frac{(b - x)_+^k}{1 + |x|^k}. \end{aligned} \tag{34}$$

Therefore, by the Lebesgue convergence theorem, letting $a \longrightarrow \pm\infty$ in (32), we have

$$\int_{\overline{\mathbb{R}}} \frac{(x - b)_+^k}{1 + |x|^k} d\mu(x) = \int_{\overline{\mathbb{R}}} \frac{(b - x)_+^k}{1 + |x|^k} d\mu(x) = 0. \tag{35}$$

This means that $\mu$ annihilates $H_{\mathrm{ReLU}^k}(\mathbb{R})$. Thus, by Theorem 1, $\mu = 0$. □

## 3. Proof of Theorem 4—Application of Theorem 1

We show

$$H_{\mathrm{ReLU}^k}(\mathbb{R}) = \mathbb{C}\mathrm{ReLU}_+^k + \mathbb{C}\mathrm{ReLU}_-^k + \overline{C_c(\mathbb{R})}^{\mathscr{Y}_k}. \tag{36}$$

We have

$$H_{\mathrm{ReLU}^k}(\mathbb{R}) \supset \overline{C_c(\mathbb{R})}^{\mathscr{Y}_k}, \tag{37}$$

by Lemma 9. Hence,

$$H_{\mathrm{ReLU}^k}(\mathbb{R}) \supset \mathbb{C}\mathrm{ReLU}_+^k + \mathbb{C}\mathrm{ReLU}_-^k + \overline{C_c(\mathbb{R})}^{\mathscr{Y}_k}. \tag{38}$$

Thus, we prove the opposite inclusion.

For any $f \in H_{\mathrm{ReLU}^k}(\mathbb{R})$, there exist $\beta_\pm \in \mathbb{C}$ such that $g_0(x) = f(x) - \beta_+ \mathrm{ReLU}_+^k(x) - \beta_- \mathrm{ReLU}_-^k(x)$ is a polynomial of degree $(k - 1)$ both on $[K, \infty)$ and on $(-\infty, -K]$ for $K \gg 1$. Fix $R \gg K$ for the time being. Then, we have

$$g(x) = \beta_+ \mathrm{ReLU}_+^k(x) + \beta_- \mathrm{ReLU}_-^k(x), \tag{39}$$

satisfying

$$\sup_{x \in \mathbb{R} \setminus [-R,R]} \frac{|g_0(x)|}{1 + |x|^k} = \sup_{x \in \mathbb{R} \setminus [-R,R]} \frac{|f(x) - g(x)|}{1 + |x|^k} = \mathrm{O}\left(R^{-1}\right). \tag{40}$$

We define

$$F(x)=\begin{cases}-g_0(R)(x-R)+g_0(R), & R\le x\le R+1,\\ g(x), & |x|\le R,\\ g_0(-R)(x+R)+g_0(-R), & -R-1\le x\le -R,\\ 0, & \text{otherwise.}\end{cases} \tag{41}$$

By the use of Lemma 9, we choose a compactly supported function

$$h\in H_{\mathrm{ReLU}^k}(\mathbb{R})\cap C_{\mathrm{c}}(\mathbb{R}), \tag{42}$$

supported on $[-R-2, R+2]$ so that

$$\sup_{x\in\mathbb{R}}|F(x)-h(x)|=\sup_{x\in[-R-2,R+2]}|F(x)-h(x)|\le R^{-1}. \tag{43}$$

Then, we have

$$\begin{aligned}&\sup_{x\in[-R,R]}\frac{|f(x)-g(x)-h(x)|}{1+|x|^k}\le CR^{-1},\\ &\sup_{x\in[-R-2,R+2]}|f(x)-g(x)-h(x)|\le C\,R^{k-1}.\end{aligned} \tag{44}$$

Then, we have

$$\begin{aligned}\|g_0-h\|_{\mathcal{Y}_k}&=\sup_{x\in\mathbb{R}}\frac{|f(x)-g(x)-h(x)|}{1+|x|^k}\\ &\le\sup_{x\in[-R,R]}\frac{|f(x)-g(x)-h(x)|}{1+|x|^k}+\sup_{x\in\mathbb{R}\setminus[-R-2,R+2]}\\ &\quad\cdot\frac{|f(x)-g(x)|}{1+|x|^k}+\sup_{x\in[-R-2,R+2]\setminus[-R,R]}\\ &\quad\cdot\frac{|f(x)-g(x)-h(x)|}{1+|x|^k}\le\sup_{x\in[-R,R]}\\ &\quad\cdot\frac{|f(x)-g(x)-h(x)|}{1+|x|^k}+\sup_{x\in\mathbb{R}\setminus[-R-2,R+2]}\\ &\quad\cdot\frac{|f(x)-g(x)|}{1+|x|^k}+C\sup_{x\in[-R-2,R+2]\setminus[-R,R]}\\ &\quad\cdot\frac{R^{k-1}}{1+|x|^k}=\mathrm{O}\left(R^{-1}\right).\end{aligned} \tag{45}$$

Since $g\in\mathbb{C}\mathrm{ReLU}_+^k+\mathbb{C}\mathrm{ReLU}_-^k$, $h\in C_c(\mathbb{R})$, and $\|f-h-g\|_{\mathcal{Y}_k}=\|g_0-h\|_{\mathcal{Y}_k}<\varepsilon$ as long as $R$ is large enough, $f-g\in\overline{C_{\mathrm{c}}(\mathbb{R})}^{\mathcal{Y}_k}$. Thus, we obtain (36).

From (36), we deduce

$$\begin{aligned}\overline{H_{\mathrm{ReLU}^k}(\Omega)}^{\mathcal{Y}_k}&\subset\overline{\mathbb{C}\left(\mathrm{ReLU}_+^k|\Omega\right)+\mathbb{C}\left(\mathrm{ReLU}^k|\Omega\right)+\overline{C_{\mathrm{c}}(\mathbb{R})|\Omega}^{\mathcal{Y}_k}}^{\mathcal{Y}_k}\\ &=\overline{\mathbb{C}\left(\mathrm{ReLU}_+^k|\Omega\right)+\mathbb{C}\left(\mathrm{ReLU}^k|\Omega\right)+C_{\mathrm{c}}(\mathbb{R})|\Omega}^{\mathcal{Y}_k}.\end{aligned} \tag{46}$$

Thus, the proof is complete if $X=\mathcal{Y}_k$. For general Banach lattices $X$, we use a routine approximation procedure. We prove

$$\overline{H_{\mathrm{ReLU}^k}(\Omega)}^{\|\cdot\|_{X(\Omega)}}=\overline{\mathbb{C}\left(\mathrm{ReLU}_+^k|\Omega\right)+\mathbb{C}\left(\mathrm{ReLU}^k|\Omega\right)+C_{\mathrm{c}}(\mathbb{R})|\Omega}^{\|\cdot\|_{X(\Omega)}}. \tag{47}$$

Let $f\in\overline{H_{\mathrm{ReLU}^k}(\Omega)}^{\|\cdot\|_{X(\Omega)}}$ and $\varepsilon>0$. Then since $f\in\overline{H_{\mathrm{ReLU}^k}(\Omega)}^{\|\cdot\|_{X(\Omega)}}$, there exists $f_0\in H_{\mathrm{ReLU}^k}$ such that

$$\|f-f_0|\Omega\|_{X(\Omega)}<\varepsilon. \tag{48}$$

Since we know that

$$f_0\in H_{\mathrm{ReLU}^k}\subset\mathbb{C}\mathrm{ReLU}_+^k+\mathbb{C}\mathrm{ReLU}_-^k+\overline{C_{\mathrm{c}}}^{\mathcal{Y}_k}, \tag{49}$$

there exist constants $\beta_\pm$ and $h\in C_{\mathrm{c}}(\mathbb{R})$ such that $\|f_0-\beta_+\mathrm{ReLU}_+^k-\beta_-\mathrm{ReLU}_-^k-h\|_{\mathcal{Y}_k}<\varepsilon$. Hence for such $\beta_\pm$ and $h\in C_{\mathrm{c}}(\mathbb{R})$, we have $\|f_0|\Omega-\beta_+\mathrm{ReLU}_+^k|\Omega-\beta_-\mathrm{ReLU}_-^k|\Omega-h|\Omega\|_{\mathcal{Y}_k(\Omega)}<\varepsilon$. Since we assume that $\mathcal{Y}_k(\Omega)$ is continuously embedded into $X(\Omega)$, we have $\|f-\beta_+\mathrm{ReLU}_+^k|\Omega-\beta_-\mathrm{ReLU}_-^k|\Omega-h|\Omega\|_{X(\Omega)}<C\varepsilon$. Therefore, we have

$$\overline{H_{\mathrm{ReLU}^k}(\Omega)}^{\|\cdot\|_{X(\Omega)}}\subset\overline{\mathbb{C}\left(\mathrm{ReLU}_+^k|\Omega|\right)+\mathbb{C}\left(\mathrm{ReLU}^k|\Omega|\right)+C_{\mathrm{c}}(\mathbb{R})|\Omega}^{\|\cdot\|_{X(\Omega)}}. \tag{50}$$

## 4. Conclusion

We specified the closure of $H_{\mathrm{ReLU}^k}(\mathbb{R})$ under the norm $\|\cdot\|_{\mathcal{Y}_k}$. This is useful when we consider the approximation by functions in the function space $H_{\mathrm{ReLU}^k}(\mathbb{R})$. We illustrated this situation using Banach lattices. Our result contains the existing result on the approximation by means of a variable exponent Lebesgue space. It is also remarkable that our attempt can be located as an attempt of understanding the neural network. For example, Carroll and Dikinson used the Radon transform [11], and other research employed some other topologies (see [12, 13]).

Remark that this note is submitted as a preprint coded: https://arxiv.org/abs/2212.13713.

## 5. Discussion

So far, we can manage to handle the case where $k$ is a non-negative integer. Our discussion heavily depended on the algebraic relation such as Lemma 8. So, we do not know how to handle the case where $k$ is not an integer. Even for the case where $k = 1/2$, the problem is difficult.

## Data Availability

No data and material were used to support this study.

## Disclosure

This paper is posted as https://export.arxiv.org/pdf/2212.13713 (see [14]).

## Conflicts of Interest

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

## Authors' Contributions

The four authors contributed equally to this paper. All of them read the whole manuscript and approved the content of the paper.

## Acknowledgments

This work was supported by a JST CREST Grant (Number JPMJCR1913, Japan). This work was also supported by the RIKEN Junior Research Associate Program. The second author was supported by a Grant-in-Aid for Young Scientists Research (No. 19K14581), Japan Society for the Promotion of Science. The fourth author was supported by a Grant-in-Aid for Scientific Research (C) (19K03546), Japan Society for the Promotion of Science.

## References

[1] J. W. Siegel and J. Xu, "High-order approximation rates for neural network with ReLU$^k$ activation functions," https://arxiv.org/abs/2012.07205.

[2] G. Singh, R. Ganvir, M. Püschel, and M. Vechev, "Beyond the single neuron convex barrier for neural network certification," https://files.sri.inf.ethz.ch/website/papers/neurips19_krelu.pdf.

[3] Á. Capel and J. Ocáriz, "Approximation with neural networks in variable Lebesgue spaces," https://arxiv.org/abs/2007.04166v1.

[4] B. Hanin, "Universal function approximation by deep neural nets with bounded width and ReLU activations," *Mathematics*, vol. 7, no. 10, p. 992, 2019.

[5] A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numerica*, vol. 8, pp. 143–195, 1999.

[6] S. Sonoda and N. Murata, "Neural network with unbounded activation functions is universal approximator," *Applied and Computational Harmonic Analysis*, vol. 43, no. 2, pp. 233–268, 2017.

[7] H. N. Mhaskar and C. A. Micchelli, "Approximation by superposition of sigmoidal and radial basis functions," *Advances in Applied Mathematics*, vol. 13, no. 3, pp. 350–373, 1992.

[8] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.

[9] N. Hatano, M. Ikeda, I. Ishikawa, and Y. Sawano, "A global universality of two-layer neural networks with ReLU activations," *Journal of Function Spaces*, vol. 2021, Article ID 6637220, 3 pages, 2021.

[10] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, pp. 183–192, 1989.

[11] M. Carroll and B. W. Dikinson, "Construction of neural nets using the Radon transform," in *Proceedings of the IEEE 1989 International Joint Conference on Neural Networks*, pp. 607–611, New York, 1989.

[12] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[13] Y. T. Sun, A. Gilbert, and A. Tewari, "On the approximation properties of random ReLU features," https://arxiv.org/pdf/1810.04374.pdf.

[14] N. Hatano, M. Ikeda, I. Ishikawa, and Y. Sawano, "Global universality of the two-layer neural network with the k-rectified linear unit," https://export.arxiv.org/pdf/2212.13713.