Analysis in Theory and Applications
Anal. Theory Appl., Vol. **39**, No. 1 (2023), pp. 93-104

DOI: 10.4208/ata.OA-2021-0006
March 2023

On Approximation by Neural Networks with Optimized Activation Functions and Fixed Weights

Dansheng Yu[1,*], Yunyou Qian[1] and Fengjun Li[2]

[1] *Department of Mathematics, Hangzhou Normal University, Hangzhou, Zhejiang 310036, China*
[2] *School of Mathematics and Statistics, Ningxia University, Yinchuan, Ningxia 750021, China*

Received 25 March 2021; Accepted (in revised version) 23 February 2022

Abstract. Recently, Li [16] introduced three kinds of single-hidden layer feed-forward neural networks with optimized piecewise linear activation functions and fixed weights, and obtained the upper and lower bound estimations on the approximation accuracy of the FNNs, for continuous function defined on bounded intervals. In the present paper, we point out that there are some errors both in the definitions of the FNNs and in the proof of the upper estimations in [16]. By using new methods, we also give right approximation rate estimations of the approximation by Li's neural networks.

Key Words: Approximation rate, modulus of continuity, modulus of smoothness, neural network operators.

AMS Subject Classifications: 41A35, 41A25, 41A20

---

1 Introduction

Feed-forward neural networks (FNNS) have been investigated entensively and deeply because of their universal approximation capbilities on compact input sets and approximation in a finite set. In the present paper, we deal with the FNNS with one hidden layer, which can be mathematically expressed as

$$N_n(x) = \sum_{j=0}^{n} c_j \sigma\left(\langle a_j \cdot x \rangle + b_j\right), \quad x \in \mathbb{R}^s, \quad s \in \mathbb{N},$$

where for $0 \leq j \leq n, b_j \in \mathbb{R}$ are the thresholds, $a_j \in \mathbb{R}^s$ are the connection weights, $c_j \in \mathbb{R}$ are the coefficients, $\langle a_j \cdot x \rangle$ is the inner product of a_j and x, and σ is the activation

*Corresponding author. *Email addresses:* `dsyu@hznu.edu.cn` (D. Yu), `yunyouqian@yeah.net` (Y. Qian), `fjli@nxu.edu.cn` (F. Li)

function. In many fundamental network models, the activation function σ is usually taken to be a sigmoidal function.

As we know, FNNS are universal approximators. Theoretically, any continuous function defined on a compact set can be approximated to any desired degree of accuracy by increasing the number of hidden neurons. A lot of results concerning the existence of an approximation and determining the number of neurons required to guarantee that all functions (belong to a certain class) can be approximated to the prescribed degree of accuracy, have been achieved by many authors (see [1-21] and [23-27]).

Let $\sigma : \mathbf{R} \to [0,c]$ be the ramp function defined by

$$\sigma(x) := \begin{cases} 0, & x \le -\mu_0, \\ c, & x \ge \mu_0, \\ \dfrac{x+\mu_0}{2\mu_0}c, & -\mu_0 < x < \mu_0, \end{cases} \qquad c \in \mathbf{R}^+, \quad 0 < \mu_0 \le \frac{1}{2}.$$

Define

$$\begin{aligned} \varphi_1(x) &:= \sigma(x+\mu_0) - \sigma(x-\mu_0) \\ &= \begin{cases} 0, & |x| \ge 2\mu_0, \\ \left(1 - \dfrac{1}{2\mu_0}|x|\right)c, & |x| < 2\mu_0, \end{cases} \\ \varphi_2(x) &:= \sigma(x+2\mu_0) - \sigma(x-2\mu_0) \\ &= \begin{cases} 0, & |x| \ge 3\mu_0, \\ \left(\dfrac{3}{2} - \dfrac{1}{2\mu_0}|x|\right)c, & \mu_0 < |x| < 3\mu_0, \\ c, & |x| \le \mu_0. \end{cases} \end{aligned}$$

Obviously, $\varphi_1(x)$ and $\varphi_2(x)$ are triangle function and trapezoidal function, respectively. Furthermore, $\varphi_1(x)$ and $\varphi_2(x)$ are nonnegative even functions, and are non-increasing for $x > 0$. By using $\varphi_j(x)$ as the activation functions, Li [16] introduced the following single-hidden layer feed-forward neural network operators:

$$N_{n,j}(f,x) := \frac{\sum\limits_{k=0}^{n} f(x_k)\varphi_j\left(\frac{1}{h}(x-x_k)\right)}{\sum\limits_{k=0}^{n} \varphi_j\left(\frac{1}{h}(x-x_k)\right)}, \tag{1.1}$$

where $x_k = a + kh, k = 0,1,\cdots,n$, are the uniform space nodes on the interval $[a,d]$, with $h = \frac{d-a}{n}$.

In [16], Li obtained the following approximation rate of $N_{n,j}(f,x)$ for functions $f(x) \in C[a,d]$:

$$\|N_{n,j}(f) - f\| \le 4\omega_2(f,h), \quad n \in \mathbf{Z}^+, \tag{1.2}$$

where $\omega_2(f,x)$ is the second order modulus of smoothness of $f(x)$.

The first purpose of the present paper is to clarify the following some facts:

(1) The operators $N_{n,1}(f,x)$ and $N_{n,2}(f,x)$ are not well defined for $0 < \mu_0 \leq \frac{1}{4}$ and $0 < \mu_0 \leq \frac{1}{6}$, respectively;

(2) Under the conditions either $j = 1$, $\frac{1}{4} < \mu_0 < \frac{1}{2}$ or $j = 2$, $\frac{1}{6} < \mu_0 \leq \frac{1}{2}$, (1.2) is not valid;

(3) When $j = 1$, $\mu_0 = \frac{1}{2}$, (1.2) is true. However, the proof of it in [16] is incorrect. In fact, in [16], the following inequality is used:

$$|f(x) - f(x_i)| \leq \omega_2(f,h) \tag{1.3}$$

for $|x - x_i| \leq h$. By taking $f(x) = x$, we see that the left hand side of (1.3) equals to $|x - x_i| > 0$ when $x \neq x_i$, but the right hand side always is zero, which means the invalidaty of (1.3). The second purpose of us is to give right approximation rate estimations of approximation by $N_{n,j}(f,x)$ under the conditions either $j = 1$, $\frac{1}{4} < \mu_0 < \frac{1}{2}$ or $j = 2$, $\frac{1}{6} < \mu_0 \leq \frac{1}{2}$ (see Theorem 3.1 in Section 3). Our final purpose is to give a correct proof of (1.2) when $j = 1$, $\mu_0 = \frac{1}{2}$. In fact, we also improve (1.2) by using a sharper constant $\frac{9}{8}$ to replace the constant 2 (see Theorem 3.2 in Section 3).

2 Some counterexamples

Proposition 2.1. *The operators $N_{n,1}(f,x)$ and $N_{n,2}(f,x)$ are not well defined for $0 < \mu_0 \leq \frac{1}{4}$ and $0 < \mu_0 \leq \frac{1}{6}$, respectively.*

Proof. When $j = 1$ and $0 < \mu_0 \leq \frac{1}{4}$, we have

$$[x_i + 2\mu_0 h, x_{i+1} - 2\mu_0 h] \subset (x_i, x_{i+1}), \quad i = 1, 2, \cdots, n-1.$$

Therefore, for any given $x \in [x_i + 2\mu_0 h, x_{i+1} - 2\mu_0 h]$, it holds that

$$\frac{1}{h}|x - x_k| \geq \frac{1}{h}\min\left(|x - x_i|, |x - x_{i+1}|\right) \geq 2\mu_0$$

for $k = 0, 1, \cdots, n$. By the definition of $\varphi_1(x)$, we see that

$$\varphi_1\left(\frac{1}{h}|x - x_k|\right) \equiv 0, \quad k = 0, 1, \cdots, n,$$

which means that the denominator of $N_{n,1}(f,x)$

$$\sum_{k=0}^{n} \varphi_1\left(\frac{1}{h}|x - x_k|\right) \equiv 0$$

for $x \in [x_i + 2\mu_0 h, x_{i+1} - 2\mu_0 h]$. Therefore, $N_{n,1}(f,x)$ are not well defined for $0 < \mu_0 \le \frac{1}{4}$.

Similarly, when $0 < \mu_0 \le \frac{1}{6}$, we have

$$\sum_{k=0}^{n} \varphi_2\left(\frac{1}{h}|x - x_k|\right) \equiv 0$$

for $x \in [x_i + 3\mu_0 h, x_{i+1} - 3\mu_0 h]$, $i = 1, 2, \cdots, n-1$, which means that $N_{n,2}(f,x)$ are not well defined for $0 < \mu_0 \le \frac{1}{6}$. $\square$

Proposition 2.2. *Inequality* (1.2) *does not hold under the conditions either*

(i) $j = 1$, $\frac{1}{4} < \mu_0 < \frac{1}{2}$, *or*

(ii) $j = 2$, $\frac{1}{6} < \mu_0 \le \frac{1}{2}$.

Proof. We first consider the case when $j = 1$ and $\frac{1}{4} < \mu_0 < \frac{1}{2}$. Taking $f(x) = x$ and $x_i^* = x_i - \frac{\mu_0}{8}h$, $1 \le i \le n$.

If $\frac{1}{4} < \mu_0 \le \frac{8}{17}$. Then

$$|x_i^* - x_i| = \frac{\mu_0}{8}h, \tag{2.1a}$$

$$|x_i^* - x_{i-1}| = h - \frac{\mu_0}{8}h \ge 2\mu_0 h, \tag{2.1b}$$

$$|x_i^* - x_k| > h \ge 2\mu_0 h, \quad k \ne i, i-1. \tag{2.1c}$$

By the definition of $\varphi_1(x)$, we have

$$\sum_{k=0}^{n} \varphi_1\left(\frac{1}{h}|x_i^* - x_k|\right) = \varphi_1\left(\frac{1}{h}|x_i^* - x_i|\right) = \varphi_1\left(\frac{\mu_0}{8}\right) = \frac{15}{16}c,$$

and

$$\begin{aligned} |N_{n,1}(f, x_i^*) - f(x_i^*)| = &\left|\frac{(f(x_i) - f(x_i^*))\varphi_1\left(\frac{\mu_0}{8}\right)}{\varphi_1\left(\frac{\mu_0}{8}\right)}\right| \\ =&|x^* - x_i| = \frac{\mu_0}{8}h. \end{aligned} \tag{2.2}$$

If $\frac{8}{17} < \mu_0 < \frac{1}{2}$, we have (2.1a), (2.1c) and

$$|x_i^* - x_{i-1}| = h - \frac{\mu_0}{8}h < 2\mu_0 h. \tag{2.3}$$

By (2.1a), (2.1c), (2.3) and the definition of $\varphi_1(x)$, we have

$$
\begin{aligned}
|N_{n,1}(f,x_i^*)-f(x_i^*)| &= \left|\frac{(f(x_i)-f(x_i^*))\varphi_1\left(\frac{\mu_0}{8}\right)+(f(x_{i-1})-f(x_i^*))\varphi_1\left(1-\frac{\mu_0}{8}\right)}{\varphi_1\left(\frac{\mu_0}{8}\right)+\varphi_1\left(1-\frac{\mu_0}{8}\right)}\right| \\
&= \left|\frac{\frac{15}{16}c\times\frac{\mu_0}{8}h+\left(\frac{17}{16}-\frac{1}{2\mu_0}\right)c\times\left(-h+\frac{\mu_0}{8}h\right)}{\frac{15}{16}c+\frac{17}{16}c-\frac{1}{2\mu_0}c}\right| \\
&= \frac{h|2\mu_0^2-9\mu_0+4|}{16\mu_0-4}>0.
\end{aligned}
\tag{2.4}
$$

By (2.2) and (2.4), we observe that

$$
\|N_{n,1}(f)-f\|_\infty>0
$$

for $\frac{1}{4}<\mu_0<\frac{1}{2}$. On the other hand, it is obvious that $\omega_2(f,t)=\omega_2(x,t)=0$ for any $t>0$. Therefore, (1.2) does not hold.

Now, we consider the case when $j=2$ and $\frac{1}{6}<\mu_0\le\frac{1}{2}$. Taking $f(x)=x$ and $\bar{x}_i=x_i-\mu_0 h,\ i=1,2,\cdots,n-1$. Direct calculations yield that

$$
\begin{aligned}
&|\bar{x}_i-x_i|=\mu_0 h,\quad |\bar{x}_i-x_{i-1}|=(1-\mu_0)h,\\
&|\bar{x}_i-x_k|\ge(1+\mu_0)h\ge 3\mu_0 h,\quad k\ne i-1,i.
\end{aligned}
$$

When $\frac{1}{6}<\mu_0\le\frac{1}{4}$, we have $1-\mu_0\ge 3\mu_0$, which implies that $|\bar{x}_i-x_{i-1}|\ge 3\mu_0 h$. Thus,

$$
|N_{n,2}(f,\bar{x}_i)-f(\bar{x}_i)|=\left|\frac{(f(x_i)-f(\bar{x}_i))\varphi_2(\mu_0)}{\varphi_2(\mu_0)}\right|=\mu_0 h. \tag{2.5}
$$

Similarly, when $\frac{1}{4}<\mu_0<\frac{1}{2}$, we have $\mu_0<|\bar{x}_i-x_{i-1}|<3\mu_0 h$. Thus,

$$
\begin{aligned}
|N_{n,2}(f,\bar{x}_i)-f(\bar{x}_i)| &= \left|\frac{(f(x_i)-f(\bar{x}_i))\varphi_2(\mu_0)+(f(x_{i-1})-f(\bar{x}_i))\varphi_2(1-\mu_0)}{\varphi_2(\mu_0)+\varphi_2(1-\mu_0)}\right| \\
&= \left|\frac{\mu_0 h+(-1+\mu_0)h\left(2-\frac{1}{2\mu_0}\right)}{3-\frac{1}{2\mu_0}}\right| \\
&= \frac{h|6\mu_0^2-5\mu_0+1|}{6\mu_0-1}.
\end{aligned}
\tag{2.6}
$$

From (2.5) and (2.6), we get

$$
|N_{n,2}(f,\bar{x}_i)-f(\bar{x}_i)|>0
$$

for $\frac{1}{6}<\mu_0<\frac{1}{2}$, $\mu_0\ne\frac{1}{3}$.

Now, we consider the case when $\mu_0 = \frac{1}{3}$. Taking $x_i' = x_i - \frac{1}{6}h$. Then

$$\begin{aligned}|N_{n,2}(f,x_i') - f(x_i')| =& \frac{\left|(f(x_i) - f(x_i'))\varphi_2\left(\frac{1}{6}\right) + (f(x_{i-1}) - f(x_i'))\varphi_2\left(\frac{5}{6}\right)\right|}{\varphi_2\left(\frac{1}{6}\right) + \varphi_2\left(\frac{5}{6}\right)}\\ =&\frac{1}{30}h > 0.\end{aligned}$$

Finally, if $\mu_0 = \frac{1}{2}$, by taking $x_0'' = x_0 + \frac{1}{16}h$, we have

$$\begin{aligned}&|x_0 - x_0''| = \frac{1}{16}h, \quad \frac{1}{2}h < |x_1 - x_0''| = \frac{15}{16}h < \frac{3}{2}h,\\ &|x_0 - x_k| \geq |x_2 - x_0''| = \frac{31}{16}h > \frac{3}{2}h, \quad k \geq 2.\end{aligned}$$

Then

$$\begin{aligned}&|N_{n,2}(f,x_0'') - f(x_0'')|\\ =&\frac{\left|(f(x_0) - f(x_0''))\varphi_2\left(\frac{1}{16}\right) + (f(x_1) - f(x_0''))\varphi_2\left(\frac{15}{16}\right)\right|}{\varphi_2\left(\frac{1}{16}\right) + \varphi_2\left(\frac{15}{16}\right)}\\ =&\frac{134}{25}h > 0.\end{aligned}$$

In conclusion, we have that

$$\|N_{n,2}(f) - f\|_\infty > 0,$$

and so (1.2) does not hold for $j = 2$, $\frac{1}{6} < \mu_0 \leq \frac{1}{2}$. □

3 Approximation rate of $N_{n,j}(f,x)$

Firstly, we have

Theorem 3.1. *Assume that $\frac{1}{4} < \mu_0 < \frac{1}{2}$ for $j = 1$, and $\frac{1}{6} < \mu_0 \leq \frac{1}{2}$ for $j = 2$. Then, for any $f \in C[a,d]$, we have*

$$\|N_{n,j}(f) - f\| \leq \omega\left(f,(j+1)\mu_0 h\right), \quad n \in \mathbf{Z}^+. \tag{3.1}$$

Proof. By the interpolation of $N_{n,j}(f,x)$ at the nodes x_i, $i = 0,1,\cdots,n$, we may assume that $x \neq x_i$, $i = 0,1,\cdots,n$.

We first prove (3.1) in the case when $j = 1$ and $\frac{1}{4} < \mu_0 < \frac{1}{2}$.

For any $x \in (x_i, x_{i+1})$, $i = 0,1,\cdots,n-1$, denote by

$$A_k = \{k \in \mathbf{Z} : 0 \leq k \leq n, \text{ and } |x - x_k| < 2\mu_0 h\},$$

$\sharp A_k$ the cardinal number of A_k. We have

$$\begin{aligned} &|x-x_{i-1}|>x_i-x_{i-1}=h>2\mu_0 h, &&\text{if } i\geq 1,\\ &|x-x_{i+2}|>x_{i+2}-x_{i+1}=h>2\mu_0 h, &&\text{if } i\leq n-2. \end{aligned}$$

Thus,

$$|x-x_k|>2\mu_0 h,\quad k\neq i,i+1.$$

On the other hand, if $|x-x_i|\geq 2\mu_0 h$, i.e., $x_i+2\mu_0 h\leq x<x_{i+1}$, then

$$|x-x_{i+1}|=x_{i+1}-x<(1-2\mu_0)h<2\mu_0 h.$$

In conclusion, we have $1\leq \sharp A_k\leq 2$. Therefore,

$$\begin{aligned} |N_{n,1}(f,x)-f(x)|&=\left|\frac{\sum\limits_{k\in A_k}\left(f(x_k)-f(x)\varphi_2\left(\frac{1}{h}|x-x_k|\right)\right.}{\sum\limits_{k\in A_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}\right|\\ \leq&\frac{\sum\limits_{k\in A_k}|f(x_k)-f(x)|\varphi_2\left(\frac{1}{h}|x-x_k|\right)}{\sum\limits_{k\in A_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}\leq\frac{\sum\limits_{k\in A_k}\omega(f,|x-x_k|)\varphi_2\left(\frac{1}{h}|x-x_k|\right)}{\sum\limits_{k\in A_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}\\ \leq&\omega(f,2\mu_0 h)\frac{\sum\limits_{k\in A_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}{\sum\limits_{k\in A_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}=\omega(f,2\mu_0 h). \end{aligned}$$

Now, we prove (3.1) in the case when $j=2$ and $\frac{1}{6}<\mu_0\leq\frac{1}{2}$. For any $x\in(x_i,x_{i+1})$, $i=0,1,\cdots,n-1$, denote by

$$B_k=\{k\in\mathbf{Z}:0\leq k\leq n,\ \text{and}\ |x-x_k|<3\mu_0 h\},$$

$\sharp B_k$ the cardinal number of B_k. We have

$$\begin{aligned} &|x-x_{i-2}|>x_i-x_{i-2}=2h>3\mu_0 h, &&\text{if } i\geq 2,\\ &|x-x_{i+3}|>x_{i+3}-x_{i+1}=2h>3\mu_0 h, &&\text{if } i\leq n-3. \end{aligned}$$

Then

$$|x-x_k|>3\mu_0 h,\quad k\neq i-1,i,i+1,i+2.$$

On the other hand, if $|x-x_i|\geq 3\mu_0 h$, i.e., $x_i+3\mu_0 h\leq x<x_{i+1}$, then

$$|x-x_{i+1}|=x_{i+1}-x<(1-3\mu_0)h<3\mu_0 h.$$

Therefore, $1 \leq \sharp B_k \leq 4$, and

$$\begin{aligned}|N_{n,2}(f,x)-f(x)| =&\left|\frac{\sum\limits_{k\in B_k}(f(x_k)-f(x))\varphi_2\left(\frac{1}{h}|x-x_k|\right)}{\sum\limits_{k\in B_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}\right|\\ \leq&\frac{\sum\limits_{k\in B_k}\omega(f,|x-x_k|)\varphi_2\left(\frac{1}{h}|x-x_k|\right)}{\sum\limits_{k\in B_k}\varphi_2\left(\frac{1}{h}|x-x_k|\right)}\\ \leq&\omega(f,3\mu_0 h).\end{aligned}$$

Thus, we complete the proof. □

To obtain the approximation rate of $N_{n,1}(f,x)$ when $\mu_0 = \frac{1}{2}$, we need the following two lemmas.

Lemma 3.1. *It holds that*

$$N_{n,1}(t,x) = x, \quad x \in [a,d].$$

Proof. Since the interpolation of $N_{n,1}(f,x)$ at the nodes, that is, $N_{n,1}(f,x_i) = f(x_i)$, $i = 0,1,\cdots,n$, we may assume that $x \neq x_i$, $i = 0,1,\cdots,n$.

For any $x \in [a,d]$, assume that x_i is the closest node to x. If $i = 0$, then,

$$\begin{aligned}&|x-x_0| \leq \frac{1}{2}(x_1-x_0) = \frac{1}{2}h,\\ &\frac{1}{2}h \leq |x-x_1| < h = 2\mu_0 h,\\ &|x-x_k| > \frac{3}{2}h > 2\mu_0 h, \quad k \neq 0,1.\end{aligned}$$

Hence,

$$\begin{aligned}N_{n,1}(t,x) =& x_0\varphi_1\left(\frac{1}{h}|x-x_0|\right) + x_1\varphi_1\left(\frac{1}{h}|x-x_1|\right)\\ =& x_0\left(1-\frac{1}{h}(x-x_0)\right) + x_1\left(1-\frac{1}{h}(x_1-x)\right)\\ =& x_0 + x_1 - \frac{x_1^2-x_0^2}{h} + \frac{x(x_1-x_0)}{h} = x.\end{aligned}$$

Similarly, if $i = n$, we have

$$N_{n,1}(t,x) = x_n\varphi_1\left(\frac{1}{h}|x-x_n|\right) + x_{n-1}\varphi_1\left(\frac{1}{h}|x-x_{n-1}|\right) = x.$$

Now, we consider the case $1 \le i \le n-1$. Without loss of generality, we may assume that $x \in \left[\frac{x_{i-1}+x_i}{2}, x_i\right)$. Then

$$|x-x_i| \le \frac{h}{2}, \quad \frac{h}{2} \le |x-x_{i-1}| < h,$$
$$|x-x_k| > h = 2\mu_0 h, \quad k \ne i-1, i.$$

Therefore,

$$N_{n,1}(t,x) = x_i\varphi_1\left(\frac{1}{h}|x-x_i|\right) + x_{i-1}\varphi_1\left(\frac{1}{h}|x-x_{i-1}|\right) = x.$$

Combining all the above discussions, we finish the proof of Lemma 3.1. □

Lemma 3.2. *It holds that*

$$|N_{n,1}((t-x)^2,x)| \le \frac{1}{4}h^2, \quad x \in [a,d].$$

Proof. Assume that x_i is the closest node to x. We only consider the case when $1 \le i \le n-1$ and $x \in \left[\frac{x_{i-1}+x_i}{2}, x_i\right)$, the other cases can be treated similarly. By Lemma 3.1 and the definition of $\varphi_1(x)$, we get

$$\begin{aligned} N_{n,1}((t-x)^2,x) =& x^2 - 2xN_{n,1}(t,x) + N_{n,1}(t^2,x) \\ =& x^2 - 2x^2 + x_i^2\varphi_1\left(\frac{1}{h}|x-x_i|\right) + x_{i-1}\varphi_1\left(\frac{1}{h}|x-x_{i-1}|\right) \\ =& -x^2 + x(x_{i-1}+x_i) - x_{i-1}x_i \\ =& (x-x_{i-1})(x_i-x). \end{aligned}$$

It is easy to see that $N_{n,1}((t-x)^2,x)$ attains its maximum value $\frac{1}{4}h^2$ at the point $x = \frac{x_i+x_{i-1}}{2}$, i.e.,

$$|N_{n,1}((t-x)^2,x)| \le \frac{1}{4}h^2.$$

We complete the proof. □

For $f(x) \in C[a,d]$, define the second order Steklov function as follows:

$$f_{hh}(x) := \frac{1}{h^2}\int_{-h/2}^{h/2} ds \int_{-h/2}^{h/2} f(x+s+t)dt.$$

Then [22]

$$\|f - f_{hh}\|_\infty \le \frac{1}{2}\omega_2(f,h), \tag{3.2a}$$

$$\|f''_{hh}\|_\infty \le \frac{1}{h^2}\omega_2(f,h). \tag{3.2b}$$

Theorem 3.2. *For any $f \in C[a,d]$, we have*

$$\|N_{n,1}(f)-f\|_\infty \le \frac{9}{8}\omega_2(f,h), \quad n\in \mathbf{Z}^+. \tag{3.3}$$

Proof. It is obvious that

$$|N_{n,1}(f,x)| \le \frac{\sum\limits_{k=0}^{n}|f(x_k)|\varphi_1\left(\frac{1}{h}(x-x_k)\right)}{\sum\limits_{k=0}^{n}\varphi_1\left(\frac{1}{h}(x-x_k)\right)} \le \|f\|_\infty.$$

Then, by (3.2a), we have

$$\begin{aligned}|N_{n,1}(f,x)-f(x)| =&|N_{n,1}(f-f_{hh},x)+N_{n,1}(f_{hh},x)-f_{hh}(x)+f_{hh}(x)-f(x)|\\ \le&2\|f-f_{hh}\|_\infty+|N_{n,1}(f_{hh},x)-f_{hh}(x)|\\ \le&\omega_2(f,h)+|N_{n,1}(f_{hh},x)-f_{hh}(x)|. \end{aligned}\tag{3.4}$$

By using Lemmas 3.1 and 3.2, and Taylor's expanssion formular:

$$f_{hh}(x_k)=f_{hh}(x)+f'_{hh}(x)(x_k-x)+\frac{1}{2}f''_{hh}(\xi_k)(x-x_k)^2, \quad \xi_k\in(x,x_k) \quad \text{or} \quad (x_k,x),$$

we deduce that

$$\begin{aligned}|N_{n,1}(f_{hh},x)-f_{hh}(x)| =&\left|\frac{\sum\limits_{k=0}^{n}(f_{hh}(x_k)-f_{hh}(x))\varphi_1\left(\frac{1}{h}(x-x_k)\right)}{\sum\limits_{k=0}^{n}\varphi_1\left(\frac{1}{h}(x-x_k)\right)}\right|\\ =&\left|\frac{\sum\limits_{k=0}^{n}\frac{1}{2}f''_{hh}(\xi_k)(x-x_k)^2\varphi_1\left(\frac{1}{h}(x-x_k)\right)}{\sum\limits_{k=0}^{n}\varphi_1\left(\frac{1}{h}(x-x_k)\right)}\right|\\ \le&\frac{1}{2}\|f''_{hh}\|_\infty N_{n,1}((t-x)^2,x)\\ \le&\frac{1}{8}\omega_2(f,h). \end{aligned}\tag{3.5}$$

We prove (3.3) by combining (3.4) and (3.5). □

Remark 3.1. In Theorems 3.1 and 3.2, we obtain the direct results of approximation by the operators $N_{n,1}(f,x)$ and $N_{n,2}(f,x)$. It will be of interesting to investigate the inverse results of approximation by these operators. Also, we only consider the neural network operators based on the equally spaced nodes, it is of interesting to extend our results to some other nodes. It is valuable to extend the main results to the multivariate case.

Acknowledgements

The third author is supported by NSFC (No. 12061055).

References

[1] G. A. Anastassiou, Rate of convergence of some neural network operators to the unit-univariate case, J. Math. Anal. Appl., 212 (1997), 237–262.

[2] A. R. Barron, Universal approximation bounds for superpositions of a sigmodal function, IEEE Trans. Inform. Theory, 39 (1993), 930–945.

[3] F. L. Cao, T. F. Xie, and Z. B. Xu, The estimate for approximation error of neural networks: A constructive approach, Neurocomputing, 71 (2008), 626–636.

[4] T. P. Chen, and H. Chen, Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to a dynamic system, IEEE Trans. Neural Netw., 6 (1995), 911–917.

[5] Z. X. Chen, and F. L. Cao, The approximation operators with sigmoidal functions, Comput. Math. Appl., 58 (2009), 758–765.

[6] C. K. Chui, and X. Li, Approximation by ridge functions and neural networks with one hidden layer, J. Approx. Theory, 70 (1992), 131–141.

[7] G. Cybenko, Approximation by superpositions of sigmoidal function, Math. Control Signals Sys., 2 (1989), 303–314.

[8] D. Costarelli, Interpolation by neural network operators activated by ramp functions, J. Math. Anal. Appl., 419 (2014), 574–582.

[9] D. Costarelli, Neural network operators: Constructive interpolation of multivate function, Neural Netw., 67 (2015), 210–226.

[10] S. Ferrari, and R. F. Stengel, Smooth function approximation using neural networks, IEEE Trans. Neural Netw., 16 (2005), 24–38.

[11] K. I. Funahashi, On the approximate realization of continuous mapping by neural networks, Neural Netw., 2 (1989), 183–192.

[12] N. Hahm, and B. I. Hong, An approximation by neural networks with a fixed weight, Comput. Math. Appl., 47 (2004), 1897–1903.

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

[14] K. Hornik, M. Stinchcombe, and H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Netw., 3 (1990), 551–560.

[15] M. Leshno, V. Y. Lin, A. Pinks, and S. Schocken, Multilayer feedward networks with a non-polynomial activation function can approximate any function, Neural Netw., 6 (1993), 861–867.

[16] F. J. Li, Constructive function approximation by neural networks with optimized activation functions and fixed weights, Neural Comput. Appl., 31 (2019), 4613–4628.

[17] G. C. Ma, D. S. Yu, and P. Zhou, On approximation by univariate sigmoidal neural networks, Acta Math. Sinica (Chinese Edition), 57 (2014), 89–100.

[18] V. Maiorov, and R. S. Meir, Approximation bounds for smooth functions in $C(R^d)$ by neural and moxture networks, IEEE Trans. Neural Netw., 9 (1998), 969–978.

[19] Y. Makovoz, Uinform approximation by neural networks, J. Approx. Theory, 95 (1998), 215–228.

[20] H. N. Mhaskar, and C. A. Micchelli, Approximation by superposition of a sigmodal function, Adv. Appl. Math., 13 (1992), 350–373.

[21] H. N. Mhaskar, and C. A. Micchelli, Degree of approximation by neural networks with a single hidden layer, Adv. Appl. Math., 16 (1995), 151–183.

[22] T. F. Xie, and S. P. Zhou, Approximation Theory of Real Functions, Press of Hangzhou Uinversity, 1998.

[23] Z. B. Xu, and F. L. Cao, The essential order of approximation for neural networks, Sci. China Ser. F, 47 (2004), 97–112.

[24] D. S. Yu, Approximation by neural networks with sigmoidal functions, Acta Math. Sinica, 29 (2013), 2013–2026.

[25] D. S. Yu, Y. Zhao, and P. Zhou, Error estimates for modified approximate approximation with Gaussian kernel, Calcolo, 50 (2013), 195–208

[26] D. S. Yu, and P. Zhou, Rates of approximation by neural networks with four layers, J. Comput. Anal. Appl., 18 (2015), 551–558.

[27] D. S. Yu, and P. Zhou, Approximation by neural network operators activated by smooth ramp functions, 59 (2016), 623–638.