Hindawi
Journal of Function Spaces
Volume 2022, Article ID 1713912, 10 pages
https://doi.org/10.1155/2022/1713912

Research Article

The Construction and Approximation of ReLU Neural Network Operators

Hengjie Chen,[1] **Dansheng Yu**,[2] **and Zhong Li**[1,3]

[1]Department of Mathematical Sciences, School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2]School of Mathematics, Hangzhou Normal University, Hangzhou 310036, China
[3]School of Information Engineering, Huzhou University, Huzhou 313000, China

Correspondence should be addressed to Zhong Li; lizhong@zstu.edu.cn

Received 10 August 2022; Accepted 14 September 2022; Published 28 September 2022

Academic Editor: Yoshihiro Sawano

In the present paper, we construct a new type of two-hidden-layer feedforward neural network operators with ReLU activation function. We estimate the rate of approximation by the new operators by using the modulus of continuity of the target function. Furthermore, we analyze features such as parameter sharing and local connectivity in this kind of network structure.

1. Introduction

Artificial neural network is a fundamental method in machine learning, which has been applied in many fields such as pattern recognition, automatic control, signal processing, auxiliary decision-making, and artificial intelligence. In particular, the successful applications of deep (multihidden layer) neural networks in image recognition, natural language processing, computer vision, etc. developed in recent years have made neural networks attract great attention. In fact, ever the function of XOR gate was implemented by adding one layer from the simplest perceptron, which leaded to the single-hidden-layer feedforward neural network.

A single-hidden-layer feedforward neural network has the expression form

$$\mathcal{N}(x) = \sum_{i=1}^{n_1} c_i \phi(\omega_i \cdot x + \theta_i), \tag{1}$$

where $c_i, \theta_i (i = 1, 2, \cdots, n)$ are called as output weights and thresholds, the dimension of input weights $\omega_i (i = 1, 2, \cdots, n)$ corresponds to that of the input x, ϕ, is called the activation function of this network, and n_1 is the number of neurons in the hidden layer. If $x \in \mathbb{R}^d$, $A_1 = [\omega_1 \omega_2 \cdots \omega_{n_1}]^T$ (T denotes the transpose) is the input weight matrix of size $n_1 \times d$, $\Theta_1 = [\theta_1, \theta_2, \cdots, \theta_n]^T$ and $C_1 = [c_1, c_2, \cdots, c_n]^T$ are vectors of thresholds and output weights, respectively, then (1) can be written as

$$\mathcal{N}(x) = C_1^T \phi(A_1 x + \Theta_1), \tag{2}$$

where $\phi(A_1 x + \Theta_1)$ means that ϕ acts on each component of $A_1 x + \Theta_1$. Now, the architecture of the neural network with two hidden layers is really not difficult to understand. If the second hidden layer contains n_2 neurons, the input weight matrix A_2 is the size of $n_2 \times n_1$, the vector of thresholds is Θ_2, and the output weight vector is O; then, the two-hidden-layer feedforward neural network can be mathematically expressed as

$$\bar{\mathcal{N}}(x) = O^T \phi(A_2 \phi(A_1 x + \Theta_1) + \Theta_2). \tag{3}$$

We call $w = \max\{n_1, n_2\}$ as the width of the network $\bar{\mathcal{N}}(x)$, and its depth is naturally 2.

The theoretical research and applications of the single-hidden-layer neural network model had been greatly developed in the 80's and 90's of last century; particularly, there were also some research results on the neural networks with multihidden layers at that time. So, in [1], Pinkus pointed out that "Nonetheless there seems to be reasonable to conjecture that the two-hidden-layer model may be significantly more promising than the single layer model, at least from a purely approximation-theoretical point of view. This problem certainly warrants further study." However, whether it is a single-hidden-layer or multihidden-layer neural network, three fundamental issues are always involved: density, complexity, and algorithms.

The so-called density or universal approximation of a neural network structure means that for any given error accuracy and the target function in a function space with some metrics, there is a specific neural network model (except for the input x, other parameters are determined) such that the error between the output and target is less than the preaccuracy. In the 1980s and 1990s, the research on the density of feedforward neural network has achieved many satisfactory results [2–9]. Since the single-hidden-layer neural network is an extreme case of the multilayer neural networks, the current focus of neural network research is still on complexity and algorithms. So-called the complexity of a neural network means that to guarantee a prescribed degree of approximation, a neural network model requires the numbers of structural parameters, including the number of layers (or depth), the number of neuron in each layer (sometimes use width), and the number of link weights and the number of thresholds. In particular, it is of interest to have more equal weights and thresholds, which is called as the parameter sharing, as this reduces computational complexity. The representation ability that has attracted much attention in deep neural networks is actually the study of complexity problem, which needs to be investigated extensively and urgently.

The constructive method is an important approach to the study of complexity, which is applicable to single- and multiple-hidden-layer neural network. In fact, there are two cases here: one is that the depth, width, and approximation degree are given, while the weights and thresholds are uncertain; the other is that all these are given; that is, the neural network model is completely determined. In order to determine the weights and thresholds in the first kind of neural network, we simply use samples to learn or train. Theoretically, the second kind of neural network can be applied directly, but the parameters are often fine-tuned with a small number of samples before use. There have been many results about the constructions of network operators [10–26]. It can be seen that these research results have an important guiding role in the construction and design of neural networks. Therefore, the purpose of this paper is to construct a kind of two-hidden-layer feedforward neural network operators with the ReLU activation function, and the upper bound estimate of approximation (or regression) ability of this neural network for two variable continuous function defined on $[-1,1]^2$ is given.

The rest of the paper is organized as follows: in Section 2, we introduce new two-hidden-layer neural network operators with ReLU activation function and give the approximation rate of approximation by the new operators. In Section 3, we give the proof of the main result. Finally, in Section 4, we give some numerical experiments and discussions.

2. Construction of ReLU Neural Network Operators and Its Approximation Properties

Let $r:\mathbb{R}\longrightarrow\mathbb{R}$ denote the rectified linear unit (ReLU), i.e., $r(x)=\max\{0,x\}$. For any $(x_1,x_2)\in\mathbb{R}\times\mathbb{R}=:\mathbb{R}^2$, we define

$$\sigma(x_1,x_2):=\frac{3}{4}r\left(1-\frac{1}{2}|x_1+x_2|-\frac{1}{2}|x_1-x_2|\right). \tag{4}$$

Obviously, σ is a continuous function of two variables supported on $[-1,1]^2$. By using the fact that $|x|=r(x)+r(-x)$, σ can be rewritten as follows:

$$\begin{aligned}\sigma(x_1,x_2)=\frac{3}{4}r\bigg(1-\frac{1}{2}[&r(x_1+x_2)+r(x_1-x_2)\\&+r(-x_1-x_2)+r(-x_1+x_2)]\bigg).\end{aligned} \tag{5}$$

From the above representation, we see that $\sigma(x_1,x_2)$ can be explained as the output of a two-hidden-layer feedforward neural network. It is obvious that σ possesses the following some important properties:

(A1) $\sigma(-x_1,x_2)=\sigma(x_1,x_2),\sigma(x_1,-x_2)=\sigma(x_1,x_2)$

(A2) For any given $x_1\in\mathbb{R},\sigma(x_1,x_2)$, it is nondecreasing for $x_2\le 0$ and nonincreasing for $x_2\ge 0$. Simultaneously, for any given $x_2\in\mathbb{R},\sigma(x_1,x_2)$,it is nondecreasing for $x_1\le 0$ and nonincreasing for $x_1\ge 0$

(A3) $0\le\sigma(x_1,x_2)\le 3/4$

(A4) $\int_{-1}^{1}\int_{-1}^{1}\sigma(x_1,x_2)dx_1dx_2=1$

For any continuous function $f(x_1,x_2)$ on $[-1,1]^2$, we define the following neural network operator:

$$\begin{aligned}\mathcal{N}_n(f;x_1,x_2)=\frac{1}{n}&\sum_{k_1=\lceil -n-\sqrt{n}\rceil}^{\lfloor n+\sqrt{n}\rfloor}\sum_{k_2=\lceil -n-\sqrt{n}\rceil}^{\lfloor n+\sqrt{n}\rfloor} f\left(\frac{k_1}{n+\sqrt{n}},\frac{k_2}{n+\sqrt{n}}\right)\\&\cdot\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right),\end{aligned} \tag{6}$$

where $\lfloor x\rfloor$ is the largest integer not greater than x, and $\lceil x\rceil$ denotes the smallest integer not less than x.

We prove that the rate of approximation by $\mathcal{N}_n(f)$ can be estimated by using the modulus of smoothness of the target function. In fact, we have

Theorem 1. *Let $f(x_1,x_2)$ is a continuous function defined on $[-1,1]^2$. Then,*

$$\begin{aligned}&|\mathcal{N}_n(f;x_1,x_2)-f(x_1,x_2)|\\&\quad\le\frac{3}{2}\left(2+\frac{1}{\sqrt{n}}\right)^2\omega\left(f;\frac{1}{\sqrt{n}}\right)\\&\quad+\frac{2M_f}{\sqrt{n}}\left(11+\frac{3}{2\sqrt{n}}\right),(x_1,x_2)\in[-1,1]^2,\end{aligned} \tag{7}$$

where $M_f = \max_{(x_1,x_2)\in[-1,1]^2} |f(x_1, x_2)|$ *and* $\omega(f\,;(1/\sqrt{n}))$ *are the modulus of continuity of* f *defined by*

$$\omega(f\,;\delta) = \sup_{|x_1-x_1'|\le\delta,|x_2-x_2'|\le\delta} \left|f(x_1,x_2) - f\left(x_1',x_2'\right)\right|. \tag{8}$$

Remark 2. For $0 < \alpha < 1$, we define the following neural network operators:

$$\begin{aligned}\tilde{\mathcal{N}}_n(f\,;x_1,x_2) = \frac{1}{n^{2\alpha}} \sum_{\lceil -n-n^\alpha\rceil}^{[n+n^\alpha]} \sum_{\lceil -n-n^\alpha\rceil}^{[n+n^\alpha]} f\left(\frac{k_1}{n+n^\alpha}, \frac{k_2}{n+n^\alpha}\right) \\ \cdot\,\sigma\left(\frac{nx_1-k_1}{n^\alpha}, \frac{nx_2-k_2}{n^\alpha}\right), (x_1,x_2)\in[-1,1]^2.\end{aligned} \tag{9}$$

Using a similar process of the proof in Theorem 1, we can get

$$\begin{aligned}&\left|\tilde{\mathcal{N}}_n(f\,;x_1,x_2) - f(x_1,x_2)\right| \\ &\le \frac{3}{2}\left(2+\frac{1}{n^\alpha}\right)^2 \omega\left(f\,;\frac{1}{n^{1-\alpha}}\right) \\ &+ \frac{2M_f}{n^\alpha}\left(11+\frac{3}{2n^\alpha}\right), \quad (x_1,x_2)\in[-1,1]^2.\end{aligned} \tag{10}$$

Remark 3. Let β $(0<\beta\le 1)$ be a fixed number, if there is a constant $L>0$ such that

$$\left|f(x_1,x_2) - f\left(x_1',x_2'\right)\right| \le L\left(\sqrt{\left(x_1-x_1'\right)^2 + \left(x_2-x_2'\right)^2}\right)^\beta, \tag{11}$$

for any $(x_1,x_2),(x_{1'},x_{2'})\in[-1,1]^2$, we say that f is a Lipschitz function of order β and write $f\in \mathrm{Lip}_L^\beta$. Obviously, when $f\in \mathrm{Lip}_L^\beta$, we have $\omega(f\,;\delta)\le\sqrt{2}L\delta^\beta$. Consequently, it follows from (7) that

$$\begin{aligned}&\left|\tilde{\mathcal{N}}_n(f\,;x_1,x_2) - f(x_1,x_2)\right| \\ &\le \frac{3}{2}\left(2+\frac{1}{n^\alpha}\right)^2 \frac{\sqrt{2}L}{n^{(1-\alpha)\beta}} \\ &+ \frac{2M_f}{n^\alpha}\left(11+\frac{3}{2n^\alpha}\right), \quad (x_1,x_2)\in[-1,1]^2.\end{aligned} \tag{12}$$

Remark 4. Now, we describe the structure of $\mathcal{N}_n(f)$ by using the form (3).

The input matrix of the first hidden layer is

$$A_1 = \begin{bmatrix} \sqrt{n} & -\sqrt{n} & -\sqrt{n} & \sqrt{n}\cdots\sqrt{n} & -\sqrt{n} & -\sqrt{n} & \sqrt{n} \\ -\sqrt{n} & -\sqrt{n} & \sqrt{n} & \sqrt{n}\cdots-\sqrt{n} & -\sqrt{n} & \sqrt{n} & \sqrt{n}\end{bmatrix}^T, \tag{13}$$

and its size is $4([n+\sqrt{n}] - \lceil -n-\sqrt{n}\rceil + 1)^2 \times 2$. The bias vector of the first hidden layer is

$$\begin{aligned}\Theta_1 = \Bigg[&-\frac{2[n+\sqrt{n}]}{\sqrt{n}}, \frac{2[n+\sqrt{n}]}{\sqrt{n}}, 0, 0, \cdots, \\ &-\frac{k_1+k_2}{\sqrt{n}}, \frac{k_1+k_2}{\sqrt{n}}, \frac{k_1-k_2}{\sqrt{n}}, \frac{k_2-k_1}{\sqrt{n}}, \cdots\Bigg]^T,\end{aligned} \tag{14}$$

and the dimension is $4([n+\sqrt{n}] - \lceil -n-\sqrt{n}\rceil + 1)^2$. The input matrix of the second hidden layer is

$$A_2 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}, \tag{15}$$

and its size is $([n+\sqrt{n}] - \lceil -n-\sqrt{n}\rceil + 1)^2 \times 4([n+\sqrt{n}] - \lceil -n-\sqrt{n}\rceil + 1)^2$. Θ_2 is a constant 1 vector with dimension $([n+\sqrt{n}] - \lceil -n-\sqrt{n}\rceil + 1)^2$. The output weight vector is

$$O = \left[f\left(\frac{[n+\sqrt{n}]}{n+\sqrt{n}}, \frac{[n+\sqrt{n}]}{n+\sqrt{n}}\right), \cdots, f\left(\frac{\lceil -n-\sqrt{n}\rceil}{n+\sqrt{n}}, \frac{\lceil -n-\sqrt{n}\rceil}{n+\sqrt{n}}\right)\right]^T. \tag{16}$$

Its general term and dimension are $f((k_1/n+\sqrt{n}), (k_2/n+\sqrt{n}))$ and $([n+\sqrt{n}] - \lceil -n-\sqrt{n}\rceil + 1)^2$, respectively.

We can see that there are two different numbers in weight matrices A_1 and A_2, respectively. That is, neural network operators $\mathcal{N}_n(f)$ have a strong weight sharing feature. There are some results about the constructions of this kind of neural networks [14, 27–29]. Moreover, A_2 shows that this neural network is locally connected. Finally, the simplicity of bias vector Θ_2 also greatly reduces the complexity of the neural network.

3. Proof of the Main Result

To prove Theorem 1, we need the following auxiliary lemma.

Lemma 5. *For function $\sigma(x_1, x_2)$, we have*

$$\begin{aligned}
&\int_{\lceil -n-\sqrt{n}\rceil -1}^{[nx_1]-1}\int_{\lceil -n-\sqrt{n}\rceil -1}^{[nx_2]-1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2\\
&\quad\le \sum_{k_1=\lceil -n-\sqrt{n}\rceil}^{[nx_1]-1}\sum_{k_2=\lceil -n-\sqrt{n}\rceil}^{[nx_2]-1}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\\
&\quad\le \int_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]}\int_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2,\\
&\int_{\lceil -n-\sqrt{n}\rceil -1}^{[nx_1]-1}\int_{\lceil nx_2\rceil +1}^{[n+\sqrt{n}]+1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2\\
&\quad\le \sum_{k_1=\lceil -n-\sqrt{n}\rceil}^{[nx_1]-1}\sum_{k_2=\lceil nx_2\rceil +1}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\\
&\quad\le \int_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]}\int_{\lceil nx_2\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2,\\
&\int_{\lceil nx_1\rceil +1}^{[n+\sqrt{n}]+1}\int_{\lceil -n-\sqrt{n}\rceil -1}^{[nx_2]-1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2\\
&\quad\le \sum_{k_1=\lceil nx_1\rceil +1}^{[n+\sqrt{n}]}\sum_{k_2=\lceil -n-\sqrt{n}\rceil}^{[nx_2]-1}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\\
&\quad\le \int_{\lceil nx_1\rceil}^{[n+\sqrt{n}]}\int_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2,\\
&\int_{\lceil nx_1\rceil +1}^{[n+\sqrt{n}]+1}\int_{\lceil nx_2\rceil +1}^{[n+\sqrt{n}]+1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2\\
&\quad\le \sum_{k_1=\lceil nx_1\rceil +1}^{[n+\sqrt{n}]}\sum_{k_2=\lceil nx_2\rceil +1}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\\
&\quad\le \int_{\lceil nx_1\rceil}^{[n+\sqrt{n}]}\int_{\lceil nx_2\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2.
\end{aligned}\tag{17}$$

Proof. We only prove (1), and (2), (3), and (4) can be proved similarly.

(1) When $k_i-1<k_i<k_i+1\le[nx_i]-1 (i=1,2)$, we have

$$nx_i-(k_i-1)>nx_i-k_i>nx_i-(k_i+1)\ge 1. \tag{18}$$

Considering the monotonicity of $\sigma(x_1,x_2)$, we have

$$\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\le\int_{k_1}^{k_1+1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)dt_1,\tag{19}$$

$$\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\le\int_{k_2}^{k_2+1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_2.\tag{20}$$

Combining (19) and (20) leads to

$$\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\le\int_{k_1}^{k_1+1}\int_{k_2}^{k_2+1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2.\tag{21}$$

Similarly, we have

$$\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\ge\int_{k_1-1}^{k_1}\int_{k_2-1}^{k_2}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2.\tag{22}$$

By (21), (22), and summation from $\lceil -n-\sqrt{n}\rceil$ to $[nx_i]-1(i=1,2)$, we obtain (1) of Lemma 5.

(2) When $k_2+1>k_2>k_2-1\ge\lceil nx_2\rceil+1$, we have

$$nx_2-(k_2+1)<nx_2-k_2<nx_2-(k_2-1)\le -1.\tag{23}$$

From (18), (23), and in a similar way to the proof in proving (1), we get

$$\begin{aligned}
&\int_{k_1-1}^{k_1}\int_{k_2}^{k_2+1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2\le\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right),\\
&\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\le\int_{k_1}^{k_1+1}\int_{k_2-1}^{k_2}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2.
\end{aligned}\tag{24}$$

By summation for $\lceil -n-\sqrt{n}\rceil\le k_1\le[nx_1]-1$ and $\lceil nx_2\rceil+1\le k_2\le[n+\sqrt{n}]$, we obtain (2) of Lemma 5. □

Proof of Theorem 6. Let

$$M_f=\max_{(x_1,x_2)\in[-1,1]^2}|f(x_1,x_2)|.\tag{25}$$

Then,

$$\begin{aligned}
&|\mathcal{N}_n(f;x_1,x_2)-f(x_1,x_2)|\\
&\quad=\left|\frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}f\left(\frac{k_1}{n+\sqrt{n}},\frac{k_2}{n+\sqrt{n}}\right)\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)-f(x_1,x_2)\right|\\
&\quad\le\frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\left|f\left(\frac{k_1}{n+\sqrt{n}},\frac{k_2}{n+\sqrt{n}}\right)-f(x_1,x_2)\right|\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\\
&\qquad+M_f\left|\frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)-1\right|=:I_1+M_fI_2.
\end{aligned}\tag{26}$$

We further estimate $|\mathcal{N}_n(f;x_1,x_2)-f(x_1,x_2)|$ by estimating I_1 and I_2, respectively.

Set

$$\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]} = \sum_{|x_1-k_1/n|<1/\sqrt{n}}\ \sum_{|x_2-k_2/n|<1/\sqrt{n}} + \sum_{|x_1-k_1/n|\ge 1/\sqrt{n}}\ \sum_{|x_2-k_2/n|<1/\sqrt{n}} + \sum_{|x_1-k_1/n|<1/\sqrt{n}}\ \sum_{|x_2-k_2/n|\ge 1/\sqrt{n}} + \sum_{|x_1-k_1/n|\ge 1/\sqrt{n}}\ \sum_{|x_2-k_2/n|\ge 1/\sqrt{n}} =: \sum^1 + \sum^2 + \sum^3 + \sum^4. \tag{27}$$

Since in $\sum^2 + \sum^3 + \sum^4$, at least one of inequalities $|x_1 - k_1/n| \ge 1/\sqrt{n}$ and $|x_2 - k_2/n| \ge 1/\sqrt{n}$ holds; so, either

$$\left|\left(x_1 - \frac{k_1}{n}\right) + \left(x_2 - \frac{k_2}{n}\right)\right| + \left|\left(x_1 - \frac{k_1}{n}\right) + \left(x_2 - \frac{k_2}{n}\right)\right| \ge 2\left|x_1 - \frac{k_1}{n}\right| \tag{28}$$

or

$$\left|\left(x_1 - \frac{k_1}{n}\right) + \left(x_2 - \frac{k_2}{n}\right)\right| + \left|\left(x_1 - \frac{k_1}{n}\right) - \left(x_2 - \frac{k_2}{n}\right)\right| \ge 2\left|x_2 - \frac{k_2}{n}\right| \tag{29}$$

is valid. Therefore,

$$1 - \frac{1}{2}\left[\left|\frac{nx_1 - k_1}{\sqrt{n}} + \frac{nx_2 - k_2}{\sqrt{n}}\right| + \left|\frac{nx_1 - k_1}{\sqrt{n}} - \frac{nx_2 - k_2}{\sqrt{n}}\right|\right] \le 0, \tag{30}$$

which implies that

$$\left(\sum^2 + \sum^3 + \sum^4\right)\left|f\left(\frac{k_1}{n+\sqrt{n}}, \frac{k_2}{n+\sqrt{n}}\right) - f(x_1, x_2)\right|\sigma\left(\frac{nx_1 - k_1}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) = 0. \tag{31}$$

For $\sum^1$, by the facts that $|x_1 - k_1/n| < 1/\sqrt{n}$, $|x_2 - k_2/n| < 1/\sqrt{n}$, for $(x_1, x_2) \in [-1, 1]^2$, we obtain that

$$\begin{gathered} -\sqrt{n} < k_i - nx_i < \sqrt{n}, i = 1, 2\,; \\ \left|\frac{k_i}{n+\sqrt{n}} - x_i\right| \le \frac{|nx_i - k_i| + \sqrt{n}}{n+\sqrt{n}} \le \frac{2}{\sqrt{n}}, i = 1, 2. \end{gathered} \tag{32}$$

Hence,

$$\begin{aligned} &\frac{1}{n}\left(\sum^1\right)\left|f\left(\frac{k_1}{n+\sqrt{n}}, \frac{k_2}{n+\sqrt{n}}\right) - f(x_1, x_2)\right|\sigma\left(\frac{nx_1 - k_1}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &\quad \le \omega\left(f\,; \frac{2}{\sqrt{n}}\right)\frac{1}{n}\sum^1 \sigma\left(\frac{nx_1 - k_1}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &\quad \le \frac{3}{2}\left(2 + \frac{1}{\sqrt{n}}\right)^2 \omega\left(f\,; \frac{1}{\sqrt{n}}\right), \end{aligned} \tag{33}$$

where we have used the inequality $0 \le \sigma \le 3/4$, and the fact that the number of the terms in $\sum^1$ is no more than $(2\sqrt{n} + 1)^2$. From (27)-(33), it follows that

$$I_1 \le \frac{3}{2}\left(2 + \frac{1}{\sqrt{n}}\right)^2 \omega\left(f\,; \frac{1}{\sqrt{n}}\right). \tag{34}$$

Set

$$D := \sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]-1}\sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]-1} + \sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]-1}\sum_{\lceil nx_2\rceil+1}^{[n+\sqrt{n}]} + \sum_{\lceil nx_1\rceil+1}^{[n+\sqrt{n}]}\sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]-1} + \sum_{\lceil nx_1\rceil+1}^{[n+\sqrt{n}]}\sum_{\lceil nx_2\rceil+1}^{[n+\sqrt{n}]}. \tag{35}$$

Then,

$$\begin{aligned} I_2 &\le \left|D\left(\frac{1}{n}\sigma\left(\frac{nx_1 - k_1}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right)\right) - 1\right| \\ &\quad + \frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1 - [nx_1]}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &\quad + \frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1 - \lceil nx_1\rceil}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &\quad + \frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1 - k_1}{\sqrt{n}}, \frac{nx_2 - [nx_2]}{\sqrt{n}}\right) \\ &\quad + \frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1 - k_1}{\sqrt{n}}, \frac{nx_2 - \lceil nx_2\rceil}{\sqrt{n}}\right) \\ &=: I_{21} + I_{22} + I_{23} + I_{24} + I_{25}. \end{aligned} \tag{36}$$

Firstly, we have

$$\begin{aligned} I_{22} &= \frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1 - [nx_1]}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_{|x_2-k_2/n|\ge 1/\sqrt{n}}\sigma\left(\frac{nx_1 - [nx_1]}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &\quad + \frac{1}{n}\sum_{|x_2-k_2/n|<1/\sqrt{n}}\sigma\left(\frac{nx_1 - [nx_1]}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right). =: \Delta_1 + \Delta_2. \end{aligned} \tag{37}$$

Noting that $|x_2 - k_2/n| \ge 1/\sqrt{n}$, we get $\Delta_1 = 0$ by the similar arguments for estimating $\sum^2 + \sum^3 + \sum^4$ in (27). Therefore,

$$\begin{aligned} \Delta_2 &= \frac{1}{n}\sum_{|x_2-k_2/n|<1/\sqrt{n}}\sigma\left(\frac{nx_1 - [nx_1]}{\sqrt{n}}, \frac{nx_2 - k_2}{\sqrt{n}}\right) \\ &\le \frac{3}{4n}\cdot\left(2\sqrt{n} + 1\right) \le \frac{3}{2\sqrt{n}}\left(1 + \frac{1}{2\sqrt{n}}\right). \end{aligned} \tag{38}$$

Consequently,

$$I_{22} \le \frac{3}{2\sqrt{n}}\left(1+\frac{1}{2\sqrt{n}}\right). \tag{39}$$

Similarly, we have

$$I_{2i} \le \frac{3}{2\sqrt{n}}\left(1+\frac{1}{2\sqrt{n}}\right), i = 3, 4, 5. \tag{40}$$

Thus, we already have

$$I_{22}+I_{23}+I_{24}+I_{25} \le \frac{6}{\sqrt{n}}\left(1+\frac{1}{2\sqrt{n}}\right). \tag{41}$$

Now, let us estimate I_{21}. By

$$\begin{gathered}\frac{1}{n}\int_{\lceil -n-\sqrt{n}\rceil -1}^{[nx_1]-1}\int_{\lceil -n-\sqrt{n}\rceil -1}^{[nx_2]-1}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2 \\ = \int_{nx_1-[nx_1]+1/\sqrt{n}}^{nx_1-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\int_{nx_2-[nx_2]+1/\sqrt{n}}^{nx_2-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\sigma(t_1,t_2)dt_1dt_2, \\ \frac{1}{\sqrt{n}} \le \frac{nx_i-[nx_i]+1}{\sqrt{n}} \le \frac{2}{\sqrt{n}};\ \frac{nx_i-\lceil -n-\sqrt{n}\rceil+1}{\sqrt{n}} \ge 1, i = 1, 2,\end{gathered} \tag{42}$$

we deduce that

$$\begin{gathered}\left|\int_{nx_1-[nx_1]+1/\sqrt{n}}^{nx_1-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\int_{nx_2-[nx_2]+1/\sqrt{n}}^{nx_2-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\sigma(t_1,t_2)dt_1dt_2\right. \\ \left.-\int_0^1\int_0^1\sigma(t_1,t_2)dt_1dt_2\right| \le \frac{4}{\sqrt{n}},\end{gathered} \tag{43}$$

where we used the fact that the support of $\sigma(t_1, t_2)$ is $[-1, 1]^2$.

Similarly, by

$$\begin{gathered}\frac{1}{n}\int_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]}\int_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]}\sigma\left(\frac{nx_1-t_1}{\sqrt{n}},\frac{nx_2-t_2}{\sqrt{n}}\right)dt_1dt_2 \\ = \int_{nx_1-[nx_1]+1/\sqrt{n}}^{nx_1-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\int_{nx_2-[nx_2]+1/\sqrt{n}}^{nx_2-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\sigma(t_1,t_2)dt_1dt_2, \\ 0 \le \frac{nx_i-[nx_i]}{\sqrt{n}} \le \frac{1}{\sqrt{n}};\ \frac{nx_i-\lceil -n-\sqrt{n}\rceil}{\sqrt{n}} > 1-\frac{1}{\sqrt{n}}, i = 1, 2,\end{gathered} \tag{44}$$

we have

$$\begin{gathered}\left|\int_{nx_1-[nx_1]+1/\sqrt{n}}^{nx_1-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\int_{nx_2-[nx_2]+1/\sqrt{n}}^{nx_2-\lceil -n-\sqrt{n}\rceil+1/\sqrt{n}}\sigma(t_1,t_2)dt_1dt_2\right. \\ \left.-\int_0^1\int_0^1\sigma(t_1,t_2)dt_1dt_2\right| \le \frac{4}{\sqrt{n}}.\end{gathered} \tag{45}$$

By (1) of Lemma 5, (43), and (45), we have

$$\begin{gathered}\left|\frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]-1}\sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]-1}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\right. \\ \left.-\int_0^1\int_0^1\sigma(t_1,t_2)dt_1dt_2\right| \le \frac{4}{\sqrt{n}}.\end{gathered} \tag{46}$$

By (2)-(4) of Lemma 5, and the arguments similar to (43) and (45), we obtain that

$$\begin{gathered}\left|\frac{1}{n}\sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_1]-1}\sum_{\lceil nx_2\rceil+1}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\right. \\ \left.-\int_0^1\int_{-1}^0\sigma(t_1,t_2)dt_1dt_2\right| \le \frac{4}{\sqrt{n}},\end{gathered} \tag{47}$$

$$\begin{gathered}\left|\frac{1}{n}\sum_{\lceil nx_1\rceil+1}^{[n+\sqrt{n}]}\sum_{\lceil -n-\sqrt{n}\rceil}^{[nx_2]-1}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\right. \\ \left.-\int_{-1}^0\int_0^1\sigma(t_1,t_2)dt_1dt_2\right| \le \frac{4}{\sqrt{n}},\end{gathered} \tag{48}$$

$$\begin{gathered}\left|\frac{1}{n}\sum_{\lceil nx_1\rceil+1}^{[n+\sqrt{n}]}\sum_{\lceil nx_2\rceil+1}^{[n+\sqrt{n}]}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\right. \\ \left.-\int_{-1}^0\int_{-1}^0\sigma(t_1,t_2)dt_1dt_2\right| \le \frac{4}{\sqrt{n}}.\end{gathered} \tag{49}$$

By (46)-(49) and the identity $\int_{-1}^1\int_{-1}^1\sigma(t_1,t_2)dt_1dt_2 = 1$, we have

$$I_{21} = \left|D\left(\frac{1}{n}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\right)-1\right| \le \frac{16}{\sqrt{n}}. \tag{50}$$

It follows from (26), (34)-(41), and (50) that

$$\begin{gathered}\left|\mathcal{N}_n(f;x_1,x_2)-f(x_1,x_2)\right| \\ \le \frac{3}{2}\left(2+\frac{1}{\sqrt{n}}\right)^2\omega\left(f;\frac{1}{\sqrt{n}}\right) \\ +\frac{2M_f}{\sqrt{n}}\left(11+\frac{3}{2\sqrt{n}}\right), \quad (x_1,x_2)\in[-1,1]^2,\end{gathered} \tag{51}$$

which completes the proof of Theorem 6. □

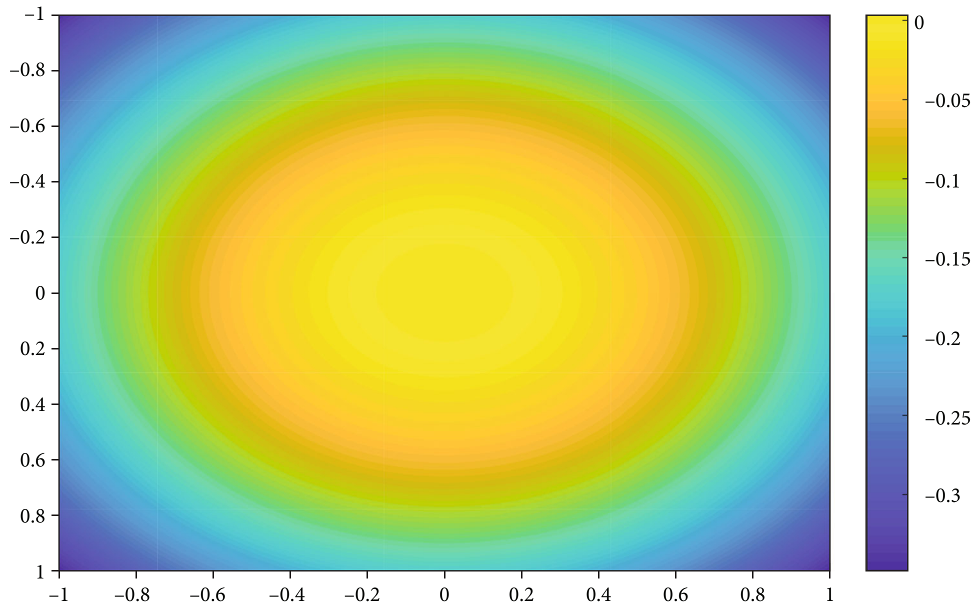

Figure 1: Errors of approximation of network operators (6) with $n = 100$.

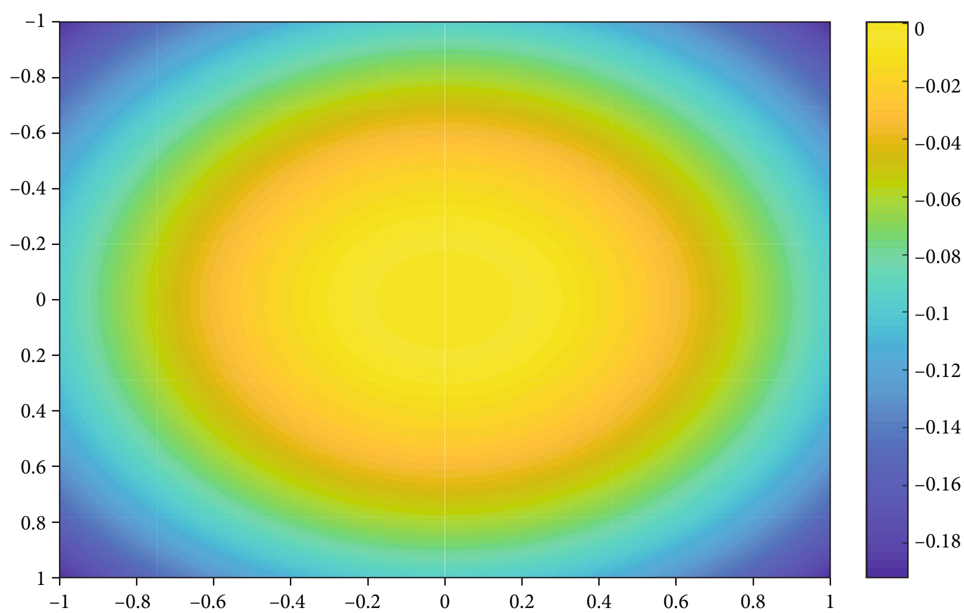

Figure 2: Errors of approximation of network operators (6) with $n = 1000$.

4. Numerical Experiments and Some Discussions

In this section, we give some numerical experiments to illustrate the theoretical results. We take $f(x_1, x_2) = x_1^2 + x_2^2$, $(x_1, x_2) \in [-1, 1]^2$ as the target function.

Set

$$e_n(x_1, x_2) \coloneqq \mathcal{N}_n(f\,; x_1, x_2) - f(x_1, x_2), (x_1, x_2) \in [-1, 1]^2. \tag{52}$$

Figures 1–3 show the results of $e_{100}(x_1, x_2)$, $e_{1000}(x_1, x_2)$ and $e_{10000}(x_1, x_2)$, respectively. When n equals to 10^6, the amount of calculation of $\mathcal{N}_n(f\,; x_1, x_2)$ is large. Therefore, we choose 6 specific points and the corresponding values of $e_n(x_1, x_2)$, which are shown in Table 1.

From the results of experiments we see that as the parameter n of neural network operators increases, the approximation effect increases; we only need to notice $M_f = 2$, $\omega(f\,; (1/\sqrt{n})) = 4/\sqrt{n}$, and after the simple calculation, we can demonstrate the validity of the obtained result.

If we investigate network operators (6) carefully, we cannot help but ask why we use $f((k_1/n + \sqrt{n}), (k_2/n + \sqrt{n}))$

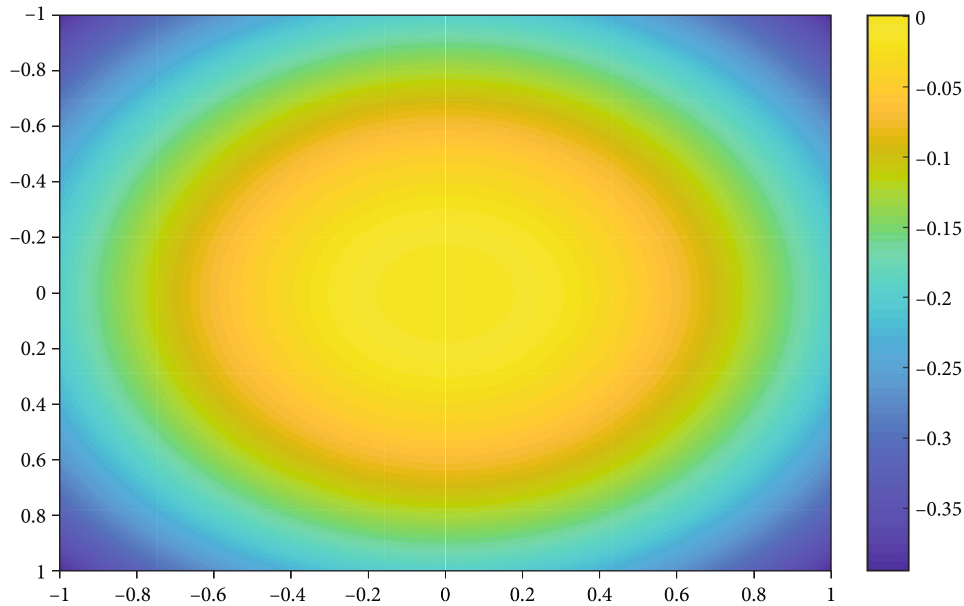

FIGURE 3: Errors of approximation of network operators (6) with $n = 10000$.

TABLE 1: The error values of $e_n(x_1, x_2)$ for 6 specific points with $n = 1000000$.

(x_1, x_2)	$(0, -1)$	$(-1, 1)$	$(0.5, 0.5)$	$(0, 0)$	$(0.5, -0.6)$	$(0.25, 0.8)$
$e_n(x_1, x_2)$	0.0020	0.0040	-9.982e-04	3.992e-07	0.0012	-0.0014

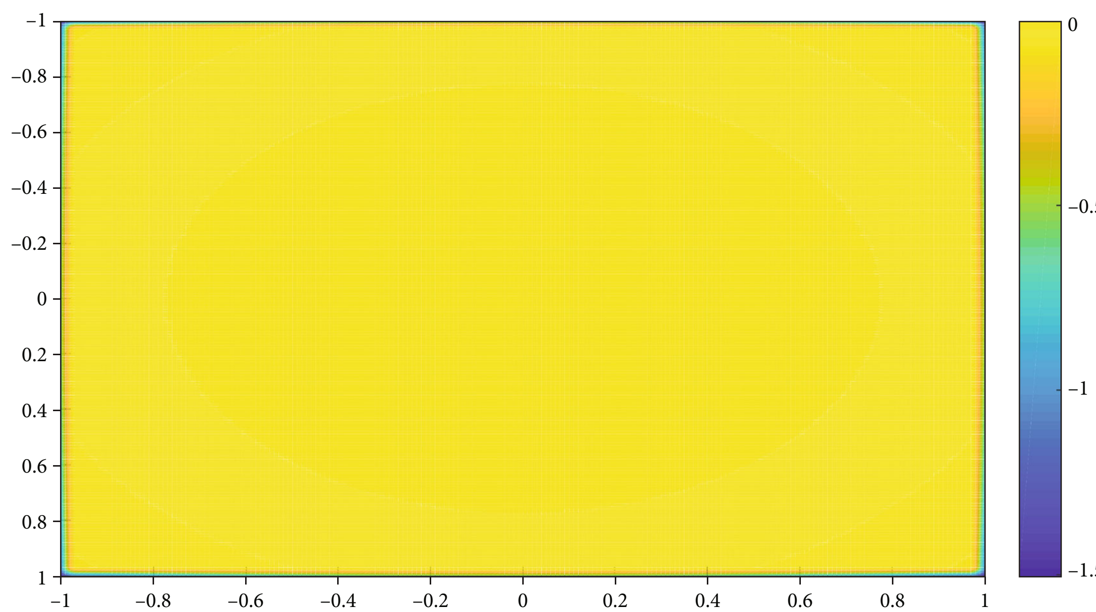

FIGURE 4: Errors of approximation of network operators (53) with $n = 1000$.

TABLE 2: The error values of $\tilde{e}_n(x_1, x_2)$ and $e_n(x_1, x_2)$ for 6 specific points with $n = 10000$.

(x_1, x_2)	$(0, -1)$	$(-1, 1)$	$(0.5, 0.5)$	$(0, 0)$	$(0.5, -0.6)$	$(0.25, 0.8)$
$\tilde{e}_n(x_1, x_2)$	-0.5000	-1.4862	2.749e-05	3.999e-05	2.47e-05	2.24e-05
$e_n(x_1, x\ 2)$	-0.0197	-0.0394	-0.0098	3.921e-05	-0.0120	-0.0138

instead of $f((k_1/n),(k_2/n))$ in (6). Because $((k_1/n),(k_2/n))$ are the conventional grid points on $[-1,1]^2$, this will reduce the amount of calculation. Now, we might as well introduce the following network operators:

$$\bar{\mathcal{N}}_n(f;x_1,x_2)=\frac{1}{n}\sum_{-n}^{n}\sum_{-n}^{n}f\left(\frac{k_1}{n},\frac{k_2}{n}\right)\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right). \tag{53}$$

Then, from the proof of Theorem 6, we have

$$\begin{aligned}
&\left|\bar{\mathcal{N}}_n(f;x_1,x_2)-f(x_1,x_2)\right|\\
&\quad=\left|\frac{1}{n}\sum_{-n}^{n}\sum_{-n}^{n}f\left(\frac{k_1}{n},\frac{k_2}{n}\right)\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)-f(x_1,x_2)\right|\\
&\quad\le\frac{1}{n}\sum_{-n}^{n}\sum_{-n}^{n}\left|f\left(\frac{k_1}{n},\frac{k_2}{n}\right)-f(x_1,x_2)\right|\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)\\
&\qquad+M_f\left|\frac{1}{n}\sum_{-n}^{n}\sum_{-n}^{n}\sigma\left(\frac{nx_1-k_1}{\sqrt{n}},\frac{nx_2-k_2}{\sqrt{n}}\right)-1\right|\\
&\quad=:I_1'+M_fI_2'.
\end{aligned} \tag{54}$$

It is not difficult to obtain the same estimate of I' as I_1, but it is not inconvenient to estimate I_2'. In fact, if we set

$$\tilde{e}_n(x_1,x_2):=\bar{\mathcal{N}}_n(f;x_1,x_2)-f(x_1,x_2),(x_1,x_2)\in[-1,1]^2. \tag{55}$$

Figure 4 shows the $\tilde{e}_n(x_1,x_2)$ with $n=1000$. We can see that next to the border of $[-1,1]^2$ the effect of $\bar{\mathcal{N}}_n(f;x_1,x_2)$ approximating f is not satisfactory. Particularly, we can see this phenomenon from Table 2 below. So, we modified $f((k_1/n),(k_2/n))$ to construct operators (6). Then, we obtain the error estimation of approximation of operators (6) and give the numerical experiments.

Data Availability

Data are available on request from the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant No. 12171434 and Zhejiang Provincial Natural Science Foundation of China under Grant No. LZ19A010002.

References

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

[2] T. P. Chen, H. Chen, and R. W. Liu, "Approximation capability in $C(R^n)$ by multilayer feedforward networks and related problems," *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 25–30, 1995.

[3] T. P. Chen and H. Chen, "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," *IEEE Transactions on Neural Networks*, vol. 6, no. 4, pp. 911–917, 1995.

[4] C. K. Chui and X. Li, "Approximation by ridge functions and neural networks with one hidden layer," *Journal of Approximation Theory*, vol. 70, no. 2, pp. 131–141, 1992.

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

[6] K. I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, no. 3, pp. 183–192, 1989.

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

[8] V. Kuková, "Kolmogorov's theorem and multilayer neural networks," *Neural Networks*, vol. 5, no. 3, pp. 501–506, 1992.

[9] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993.

[10] G. A. Anastassiou, "Rate of convergence of some neural network operators to the unit-univariate case," *Journal of Mathematical Analysis and Applications*, vol. 212, no. 1, pp. 237–262, 1997.

[11] G. A. Anastassiou, "Univariate hyperbolic tangent neural network approximation," *Mathematical and Computer Modelling*, vol. 53, no. 5-6, pp. 1111–1132, 2011.

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

[13] F. Cao, T. Xie, and Z. B. Xu, "The estimate for approximation error of neural networks: a constructive approach," *Neurocomputing*, vol. 71, no. 4-6, pp. 626–630, 2008.

[14] F. Cao and T. Xie, "The construction and approximation for feedforword neural networks with fixed weights," in *2010 International Conference on Machine Learning and Cybernetics*, pp. 3164–3168, Qingdao, China, 2010.

[15] D. B. Chen, "Degree of approximation by superpositions of a sigmoidal function," *Approximation Theory and its Applications*, vol. 9, pp. 17–28, 1993.

[16] Z. Chen and F. Cao, "The approximation operators with sigmoidal functions," *Computers & Mathematics with Applications*, vol. 58, no. 4, pp. 758–765, 2009.

[17] L. Chen and C. W. Wu, "A note on the expressive power of deep rectified linear unit networks in high- dimensional spaces," *Mathematicsl Methods in the Applied Sciences*, vol. 42, no. 9, pp. 3400–3404, 2019.

[18] D. Costarelli, "Interpolation by neural network operators activated by ramp functions," *Journal of Mathematical Analysis and Applications*, vol. 419, no. 1, pp. 574–582, 2014.

[19] D. Costarelli, "Neural network operators: constructive interpolation of multivariate functions," *Neural Networks*, vol. 67, pp. 28–36, 2015.

[20] D. Costarelli and G. Vinti, "Voronovskaja type theorems and high-order convergence neural network operators with sigmoidal functions," *Mediterranean Journal of Mathematics*, vol. 17, no. 3, 2020.

[21] I. Daubechies, R. Devore, S. Foucart, B. Hanin, and G. Petrova, "Nonlinear approximation and (deep) ReLU networks," *Constructive Approximation*, vol. 55, no. 1, pp. 127–172, 2022.

[22] H. N. Mhaskar and C. A. Micchelli, "Degree of approximation by neural and translation networks with a single hidden layer," *Advances in Applied Mathematics*, vol. 16, no. 2, pp. 151–183, 1995.

[23] Y. Y. Qian and D. S. Yu, "Rates of approximation by neural network interpolation operators," *Applied Mathematics and Computation*, vol. 418, article 126781, 2022.

[24] U. Shaham, A. Cloninger, and R. Coifman, "Provable approximation properties for deep neural networks," *Applied and Computational Harmonic Analysis*, vol. 44, no. 3, pp. 537–557, 2018.

[25] D. Yarotsky, "Error bounds for approximations with deep ReLU networks," *Neural Networks*, vol. 94, pp. 103–114, 2017.

[26] D. X. Zhou, "Universality of deep convolutional neural networks," *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 787–794, 2020.

[27] N. J. Guliyev and V. E. Ismailov, "Approximation capability of two hidden layer feedforward neural networks with fixed weights," *Neurocomputing*, vol. 316, pp. 262–269, 2018.

[28] N. J. Guliyev and V. E. Ismailov, "On the approximation by single hidden layer feedforward neural networks with fixed weights," *Neural Networks*, vol. 98, pp. 296–304, 2018.

[29] N. Hahm and B. I. Hong, "An approximation by neural networkswith a fixed weight," *Computers & Mathematcs with Applications*, vol. 47, no. 12, pp. 1897–1903, 2004.