Computers and Mathematics with Applications 58 (2009) 758–765

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

The approximation operators with sigmoidal functions[☆]

Zhixiang Chen[a], Feilong Cao[b,*]

[a] Department of Mathematics, Shaoxing University, Shaoxing, 312000, Zhejiang Province, PR China

[b] Department of Information and Mathematics Sciences, China Jiliang University, Hangzhou, 310018, Zhejiang Province, PR China

ARTICLE INFO

Article history:
Received 28 February 2009
Accepted 6 May 2009

Keywords:
Approximation
Sigmoidal function
Neural networks
Quasi-interpolation
Error estimates

ABSTRACT

The aim of this paper is to investigate the error which results from the method of approximation operators with logarithmic sigmoidal function. By means of the method of extending functions, a class of feed-forward neural network operators is introduced. Using these operators as approximation tools, the upper bounds of errors, in uniform norm, approximating continuous functions, are estimated. Also, a class of quasi-interpolation operators with logarithmic sigmoidal function is constructed for approximating continuous functions defined on the total real axis.

1. Introduction

A function σ defined on $\mathbb{R}$ is called a sigmoidal function if the following conditions are satisfied:

$$\lim_{x\to+\infty} \sigma(x) = 1, \qquad \lim_{x\to-\infty} \sigma(x) = 0.$$

The sigmoidal function is a class of important functions, which takes an important role in the research into neural networks. It is usually used to take play the role of an activation function in the hidden layer of neural networks.

One of the most familiar sigmoidal functions is the logarithmic type function defined by

$$s(x) = \frac{1}{1+e^{-x}}, \quad x \in \mathbb{R}.$$

In fact, the sigmoidal function $s(x)$ is a logistic model. This model is an important one and has been widely used in biology, demography and so on (see [1,2]). On the other hand, the function has higher *smoothness*.

For the logarithmic type function $s(x)$, we define

$$\phi(x) := \frac{1}{2}\left(s(x+1) - s(x-1)\right), \quad x \in \mathbb{R}.$$

Then some "better" properties, such as $\int_{-\infty}^{+\infty} \phi(x)\mathrm{d}x = 1$, the Fourier transforms of $\phi(x)$ being equal to 0, and $\sum_{k=-\infty}^{+\infty} \phi(x-k) = 1$, will be implied (see Section 2). Our aim, in this paper, is to introduce and study approximation operators with the function $\phi(x)$, i.e., the neural network operators and quasi-interpolation operators. Using these operators as approximation tools, we will consider the estimates of the rate of approximating continuous functions.

The first main result, Theorem 3 in Section 3, is on the estimations of rates of approximation of the constructed neural network operators. By using the modulus of continuity as metric, a Jackson type inequality is established, which reveals, to some extent, the relation between the convergence rates and the topological structure of the networks. The second one, which is discussed in Section 4, is on the construction and approximation of quasi-interpolation operators with $\phi(x)$.

☆ This research was supported by the National Natural Science Foundation of China (No. 90818020).

* Corresponding author.
E-mail address: flcao@263.net (F. Cao).

0898-1221/$ – see front matter Crown Copyright © 2009 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2009.05.001

We will construct a class of quasi-interpolation operators to approximate the continuous function defined on the total real axis, and the rate of approximation is also estimated.

2. The partition of unity of a logarithmic sigmoidal function

For the logarithmic type sigmoidal function defined by

$$s(x) = \frac{1}{1+e^{-x}}, \quad x \in \mathbb{R}$$

we let

$$g(x) := s(x+1) - s(x-1), \quad x \in \mathbb{R}.$$

It is clear that $g(x)$ is even, and

$$\begin{aligned} g(x) &= (e - e^{-1})\frac{e^{-x}}{(1+e^{-x-1})(1+e^{-x+1})} \\ &= \frac{e^2-1}{e^2}\frac{1}{(1+e^{x-1})(1+e^{-x-1})}. \end{aligned}$$

A simple computation gives

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{e^{-x}}{(1+e^{-x-1})(1+e^{-x+1})}dx &= 2\int_{0}^{+\infty} \frac{e^{-x}}{(1+e^{-x-1})(1+e^{-x+1})}dx \\ &= \frac{2e}{e^2-1}, \end{aligned}$$

which implies

$$\frac{1}{2}\int_{-\infty}^{+\infty} g(x) = \frac{e^2-1}{e} \cdot \frac{e}{e^2-1} = 1.$$

Set

$$\phi(x) := \frac{1}{2}g(x) = \frac{1}{2}(s(x+1) - s(x-1)) = \frac{e^2-1}{2e^2}\frac{1}{(1+e^{x-1})(1+e^{-x-1})}.$$

Then there exists a constant C, such that

$$|\phi(x)| \le C(1+|x|)^2, \quad x \in \mathbb{R}.$$

For function $f \in L(\mathbb{R})$, its Fourier transforms $\hat{f}$ are defined by

$$\hat{f}(x) := \int_{-\infty}^{+\infty} f(t)e^{-2\pi ixt}dt.$$

Noting that $\phi(x)$ is even, it follows that for any $k \in \mathbb{Z}$

$$\hat{\phi}(k) = \hat{\phi}(-k),$$

and for $k \in \mathbb{Z}$ and $k \neq 0$,

$$\hat{\phi}(k)e^{2\pi ikx} + \hat{\phi}(-k)e^{-2\pi ikx} = 2\hat{\phi}(k)\cos 2\pi kx.$$

We now prove the following result on Fourier transforms.

Theorem 1. *For any $k \in \mathbb{Z}$ and $k \neq 0$, we have*

$$\hat{\phi}(k) = 0.$$

Proof. Since $\phi(x)$ is an even function, we have

$$\begin{aligned} \hat{\phi}(k) &= \int_{-\infty}^{+\infty} \phi(x)e^{-2\pi ikx}dx \\ &= \frac{e^2-1}{e^2}\int_{0}^{+\infty} \frac{\cos 2\pi kx}{(1+e^{x-1})(1+e^{-x-1})}dx \\ &= \frac{e^2-1}{e^2}I. \end{aligned}$$

Using the method of complex analysis we see that

$$I = \frac{1}{2}\int_{-\infty}^{+\infty} \frac{\cos 2\pi kx}{(1+e^{x-1})(1+e^{-x-1})}dx = \frac{1}{2}\mathrm{Re}\int_{-\infty}^{+\infty} \frac{e^{2\pi ikx}}{(1+e^{x-1})(1+e^{-x-1})}dx.$$

We now introduce an analytic function defined on the plane $w = u + iv$:

$$f(\omega) = \frac{e^{2\pi ik\omega}}{(1+e^{w-1})(1+e^{-w-1})}.$$

Clearly, the function has two poles of order 1, $w_1 = 1 + \pi i$ and $w_2 = -1 + \pi i$, in $0 < I_{mw} < 2\pi$. Consider the rectangle with the vertices $-u_2, u_2, -u_2+2\pi i$ and $u_2+2\pi i$ $(1 < u_2 < +\infty)$. Then the integrate of the function $f(w)$ on the boundary of the rectangle is

$$\begin{aligned}
&\int_{-u_2}^{u_2} \frac{e^{2\pi iku}}{(1+e^{u-1})(1+e^{-u-1})}du + \int_0^{2\pi} \frac{e^{2\pi ik(u_2+vi)}}{(1+e^{u_2+vi-1})(1+e^{-u_2-vi-1})}idv \\
&\quad + \int_{u_2}^{-u_2} \frac{e^{2\pi ik(u+2\pi i)}}{(1+e^{u+2\pi i-1})(1+e^{-u-2\pi i-1})}du + \int_{2\pi}^{0} \frac{e^{2\pi ik(-u_2+vi)}}{(1+e^{-u_2+vi-1})(1+e^{u_2-vi-1})}idv \\
&= (1-e^{-4k\pi^2})\int_{-u_2}^{u_2} \frac{e^{2\pi iku}}{(1+e^{u-1})(1+e^{-u-1})}du + I_1 + I_2 \\
&= 2\pi i\left(\mathrm{Res}(f(w), 1+\pi i) + \mathrm{Res}\left(f(w), -1+\pi i\right)\right).
\end{aligned}$$

Also, as $u_2 \to +\infty$, it holds that $I_1 \to 0$, $I_2 \to 0$. So,

$$\int_{-\infty}^{+\infty} \frac{e^{2\pi iku}}{(1+e^{u-1})(1+e^{-u-1})}du = \frac{1}{1-e^{-4k\pi^2}}\frac{4\pi ie^{-2k\pi^2}}{e^{-2}-1}.$$

Therefore,

$$I = \frac{1}{2}\mathrm{Re}\left(\frac{1}{1-e^{-4k\pi^2}} \cdot \frac{4\pi ie^{-2k\pi^2}}{e^{-2}-1}\right) = 0.$$

From this it follows that $\hat{\phi}(k) = 0$. □

From Theorem 1 and the fact that $\int_{-\infty}^{+\infty}\phi(x)dx = 1$, it follows that

$$\sum_{k=-\infty}^{+\infty} \hat{\phi}(k)e^{2\pi ikx} = 1.$$

Using the above result and the fact that $|\phi(x)| \le C(1+|x|)^2$ we see that the Poisson summation formula is valid. From this, the following result follows.

Theorem 2. *For the function $\phi(x)$, it holds that*

$$\sum_{k=-\infty}^{+\infty} \phi(x-k) = 1, \quad x \in \mathbb{R}.$$

3. The approximation operators of neural networks

Feed-forward neural networks (FNNs) with one hidden layer, the only type that we are concerned with in this paper, are mathematically expressed as

$$N_n(x) = \sum_{j=0}^{n} c_j\sigma\left(\langle a_j \cdot x\rangle + b_j\right), \quad x \in \mathbb{R}^s,\ s \in \mathbb{N}, \tag{3.1}$$

where for $0 \le j \le n$, $b_j \in \mathbb{R}$ are the thresholds, $a_j \in \mathbb{R}^s$ are the connection weights, $c_j \in \mathbb{R}$ are the coefficients, $< a_j \cdot x >$ is the inner product of a_j and x, and σ is the activation function of the network. In many fundamental network models, the activation function must be a sigmoidal function.

It is well known that 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. It was proved by Cybenko [3] and Funahashi [4] that any continuous function can be approximated on a compact set with uniform topology by a network of the form given in Eq. (3.1), using any continuous, sigmoidal activation function. Hornik et al. in [5] have shown that any measurable function can be approached with such a network. Furthermore, these authors proved in [6] that any function of the Sobolev spaces can be approached with all derivatives. Various density results on FNN approximations of multivariate functions were later established by many authors using various methods, for more or less general situations: [7] by Leshno et al., [8] by Mhaskar and Micchelli, [9] by Chui and Li, [10] by Chen and Chen, [11] by Hahm and Hong, etc.

Yet these results only give theorems concerning the existence of an approximation. A related and important problem is that of complexity: determining the number of neurons required to guarantee that all functions (belonging to a certain class) can be approximated to the prescribed degree of accuracy ϵ. For example, a classical result of Barron [12] shows that if the function is assumed to satisfy certain conditions expressed in terms of its Fourier transform, and if each of the neurons evaluates a sigmoidal activation function, then at most $\mathcal{O}(\epsilon^{-2})$ neurons are needed to achieve the order of approximation ϵ. Previously, some authors have published similar results on the complexity of FNN approximations: Mhaskar and Micchelli [13], Suzuki [14], Maiorov and Meir [15], Makovoz [16], Ferrari and Stengel [17], Xu and Cao [18], Anastassiou [19], and Cao et al. [20], etc.

To aid our description, we introduce some notation. For a function $f \in C_{[a,b]}$, the *modulus of continuity* is defined by

$$\omega(f,t) = \sup_{0<h\le t} \max_{x,x+h\in[a,b]} |f(x) - f(x+h)|.$$

This modulus is usually used as a tool for measuring approximation error. It is also used to measure the smoothness of a function and its accuracy in approximation theory and Fourier analysis (see [21–23]). The function f is called (M,α)-Lipschitz continuous $(0 < \alpha \le 1)$. It can be written as $f \in \mathrm{Lip}(M,\alpha)$ if and only if there exists a constant $M > 0$ such that

$$\omega(f,\delta) \le M\delta^{\alpha}.$$

For given $n \in \mathbb{N}$, the fact that

$$\sum_{k=-\infty}^{+\infty} \phi(x-k) = 1,$$

gives

$$\sum_{k=-\infty}^{+\infty} \phi(nx-k) = 1.$$

So, for $f \in C_{[-1,1]}$, $d \in \mathbb{N}$ and $d \le n$, we construct the feed-forward neural networks

$$G_{n,d}(f,x) = \sum_{j=0}^{2(n+d)} c_j \phi(nx - j + (n+d)), \tag{3.2}$$

where

$$c_j = \begin{cases} f(-1), & 0 \le j \le d-1; \\ f\left(\dfrac{j-(n+d)}{n}\right), & d \le j \le 2n+d; \\ f(1), & 2n+d+1 \le j \le 2(n+d). \end{cases}$$

We now establish the following approximation results.

Theorem 3. *Let $f \in C_{[-1,1]}$. For the neural networks approximation operators $G_{n,d}(f,x)$ defined by (3.2), the following estimate of approximation degree holds:*

$$\|f - G_{n,d}(f)\|_\infty \le \omega\left(f, \frac{1}{\sqrt{n}}\right) + 16\|f\|_\infty \left(\mathrm{e}^{-d} + \mathrm{e}^{-\sqrt{n}}\right),$$

where $\|f\|_\infty = \max_{-1\le t\le 1} |f(t)|$.

Clearly, the density results are implied by the above result. In fact, for any $\epsilon > 0$, using the monotonically decreasing property of the modulus of smoothness, we can choose n and d large enough that

$$\|f - G_{n,d}(f)\|_\infty < \epsilon,$$

which shows that for any $f \in C_{[-1,1]}$ the constructed networks $G_{n,d}(f,x)$ can approximate functions $f(x)$ to arbitrary precision. On the other hand, the results reveal the relationships between approximation speed, the number of hidden neurons $2(n+d)$, and the smoothness of the target functions to be approximated. We can conclude that the approximation speed of the FNNs depends not only on the number of the hidden neurons, but also on the smoothness of the target function. So in a certain sense, the results obtained gave a solution for the complexity of approximation by FNNs. In particular, for the (M,α)-Lipschitz continuous function class $\mathrm{Lip}(M,\alpha)$, the approximation speed is directly proportional both to the number of hidden neurons and to the smoothness exponents of the target function α.

Our approach in this paper is constructive, mainly based on an extending function method. This approach is easily realizable in computations. Naturally, the primary targets approximated by FNNs should be multivariate functions. In the development of this paper, we shall consider the difficult problem of how to extend our method to the case of multiple inputs. We will discuss this question further in other papers.

We now turn to beginning the proof of Theorem 3. For $f \in C_{[-1,1]}$, we can extend $f(x)$ from $[-1, 1]$ to $[-2, 2]$:

$$F(x) = \begin{cases} f(-1), & x \in [-2, -1); \\ f(x), & x \in [-1, 1]; \\ f(1), & x \in (1, 2]. \end{cases}$$

Clearly,

$$\|F\|_\infty \le \|f\|_\infty, \qquad \omega(F, \delta) = \omega(f, \delta),\ \delta > 0.$$

From the definition of the neural network operators, we can rewrite $G_{n,d}(f, x)$: as

$$G_{n,d}(f, x) = \sum_{k=-n-d}^{-n-1} f(-1)\phi(nx - k) + \sum_{k=-n}^{n} f\left(\frac{k}{n}\right)\phi(nx - k) + \sum_{k=n+1}^{n+d} f(1)\phi(nx - k).$$

Therefore, for $x \in [-1, 1]$, we have

$$\begin{aligned} |f(x) - G_{n,d}(f, x)| &= \left|\sum_{k=-\infty}^{+\infty} f(x)\phi(nx - k) - \sum_{k=-n-d}^{n+d} F\left(\frac{k}{n}\right)\phi(nx - k)\right| \\ &= \left|\sum_{k=-\infty}^{+\infty} F(x)\phi(nx - k) - \sum_{k=-n-d}^{n+d} F\left(\frac{k}{n}\right)\phi(nx - k)\right| \\ &\le \sum_{k=-n-d}^{n+d} \left|F(x) - F\left(\frac{k}{n}\right)\right|\phi(nx - k) + \sum_{|k|\ge n+d+1} |F(x)|\phi(nx - k) \\ &:= \Delta_1 + \Delta_2. \end{aligned}$$

Next, we estimate Δ_1 and Δ_2, respectively. From $|x| \le 1$ and $|k| \ge n + d + 1$, it follows that $|nx - k| \ge d + 1$. Let

$$\psi(x) = \frac{1}{(1 + e^{x-1})(1 + e^{-x-1})}.$$

Then

$$\psi'(x) = -\frac{e^x - e^{-x}}{e(1 + e^{x-1})^2(1 + e^{-x-1})^2} \le 0, \quad x \ge 0.$$

Therefore,

$$\begin{aligned} \Delta_2 &\le \|f\|_\infty \sum_{|k|\ge n+d+1} \phi(nx - k) \\ &\le \frac{e^2 - 1}{e^2}\|f\|_\infty \int_d^{+\infty} \frac{1}{(1 + e^{x-1})(1 + e^{-x-1})} dx \\ &\le \frac{e^2 - 1}{e}\|f\|_\infty e^{-d} \\ &\le 3\|f\|_\infty e^{-d}. \end{aligned}$$

For Δ_1, we have

$$\begin{aligned} \Delta_1 &= \sum_{|x-\frac{k}{n}|\le\frac{1}{\sqrt{n}}} \left|F(x) - F\left(\frac{k}{n}\right)\right|\phi(nx - k) + \sum_{|x-\frac{k}{n}|>\frac{1}{\sqrt{n}}} \left|F(x) - F\left(\frac{k}{n}\right)\right|\phi(nx - k) \\ &\le \omega\left(F, \frac{1}{\sqrt{n}}\right)\sum_{k=-\infty}^{+\infty} \phi(nx - k) + 2\|F\|_\infty \sum_{|nx-k|>\sqrt{n}} \phi(nx - k) \\ &\le \omega\left(f, \frac{1}{\sqrt{n}}\right) + 16\|f\|_\infty e^{-\sqrt{n}}. \end{aligned}$$

Collecting the estimates of Δ_1 and Δ_2 gives

$$\|f - G_{n,d}(f)\|_\infty \le \omega\left(f, \frac{1}{\sqrt{n}}\right) + 16\|f\|_\infty \left(e^{-d} + e^{-\sqrt{n}}\right).$$

This finishes the proof of Theorem 2.

Remark 1. From

$$\phi(nx - k) = \frac{1}{2}[s(nx - k + 1) - s(nx - k - 1)]$$

we can represent the neural network operators $G_{n,d}(f,x)$ as

$$G_{n,d}(f,x) = \sum_{k=-n-d}^{-n-1} \frac{f(-1)}{2} s(nx-k+1) + \sum_{k=-n-d}^{-n-1} \left(-\frac{f(-1)}{2}\right) s(nx-k-1) + \sum_{k=-n}^{n} \frac{f\left(\frac{k}{n}\right)}{2} s(nx-k+1)$$
$$+ \sum_{k=-n}^{n} \left(-\frac{f\left(\frac{k}{n}\right)}{2}\right) s(nx-k-1) + \sum_{k=n+1}^{n+d} \frac{f(1)}{2} s(nx-k+1) + \sum_{k=n+1}^{n+d} \left(-\frac{f(1)}{2}\right) s(nx-k-1).$$

Remark 2. From the result

$$\sum_{k=-\infty}^{+\infty} \phi(nx-k) = 1,$$

it can be seen that for any $n \in \mathbb{N}$

$$\left|1 - \sum_{k=-n}^{n} \phi(nx-k)\right| = \sum_{k=-\infty}^{-n-1} \phi(nx-k) + \sum_{k=n+1}^{+\infty} \phi(nx-k).$$

Taking $x = 1$ gives

$$\left|1 - \sum_{k=-n}^{n} \phi(nx-k)\right| > \phi(1),$$

which shows that the error of the networks

$$G_{n,0}(f,x) = \sum_{k=-n}^{n} f\left(\frac{k}{n}\right) \phi(nx-k)$$

approximating the function $f \equiv 1$ is $\|1 - G_{n,0}(1)\|_\infty > \phi(1)$. This forms our main motivation for introducing parameter d in the network operators.

4. The quasi-interpolation operators

In this section, we introduce and study a class of quasi-interpolation operators with function $\phi(x)$.

Let $f \in C_0(\mathbb{R})$ (the set of continuous and bounded functions defined on $\mathbb{R}$). We define quasi-interpolation operators:

$$G_n(f,x) := \sum_{k=-\infty}^{+\infty} f\left(\frac{k}{n}\right) \phi(nx-k). \tag{4.1}$$

The main result of this section is as follows.

Theorem 4. *If $f \in C_0(\mathbb{R})$, then for $G_n(f)$ defined by (4.1), it holds that*

$$|f(x) - G_n(f,x)| \le \omega\left(f, \frac{1}{\sqrt{n}}\right) + 16\|f\|_\infty e^{-\sqrt{n}}.$$

Furthermore, if $\{f(\frac{k}{n})\}_{k=-\infty}^{+\infty}$ is monotone increasing, then so is $G_n(f,x)$.

Proof. By definition (4.1) we have

$$\begin{aligned} f(x) - \sum_{k=-\infty}^{+\infty} f\left(\frac{k}{n}\right)\phi(nx-k) &= \sum_{k=-\infty}^{+\infty} \left(f(x) - f\left(\frac{k}{n}\right)\right)\phi(nx-k) \\ &= \sum_{\left|x-\frac{k}{n}\right| \le \frac{1}{\sqrt{n}}} \left(f(x) - f\left(\frac{k}{n}\right)\right)\phi(nx-k) + \sum_{\left|x-\frac{k}{n}\right| > \frac{1}{\sqrt{n}}} \left(f(x) - f\left(\frac{k}{n}\right)\right)\phi(nx-k) \\ &:= \Delta_4 + \Delta_5. \end{aligned}$$

Noting that

$$|\Delta_4| \le \omega\left(f, \frac{1}{\sqrt{n}}\right) \sum_{\left|x-\frac{k}{n}\right| \le \frac{1}{\sqrt{n}}} \phi(nx-k) \le \omega\left(f, \frac{1}{\sqrt{n}}\right),$$

and

$$|\Delta_5| \leq 2\|f\|_\infty \sum_{|nx-k|>\sqrt{n}} \phi(nx-k) \leq 16\|f\|_\infty \mathrm{e}^{-\sqrt{n}}.$$

we get

$$\left| f(x) - \sum_{k=-\infty}^{+\infty} f\left(\frac{k}{n}\right) \phi(nx-k) \right| \leq \omega\left(f, \frac{1}{\sqrt{n}}\right) + 16\|f\|_\infty \mathrm{e}^{-\sqrt{n}}.$$

This competes the proof of the first conclusion of Theorem 4. Next we prove the second one. In fact, for given $x \in \mathbb{R}$, and $0 < n\Delta x < 1$, there exists $k_0 \in \mathbb{Z}$ such that

$$nx - k_0 \leq 0 < nx - k_0 + 1.$$

Then for $k \geq k_0 + 1$,

$$\phi(n(x + \Delta x) - k) - \phi(nx - k) \geq 0.$$

And for $k \leq k_0 - 1$,

$$\phi(n(x + \Delta x) - k) - \phi(nx - k) \leq 0.$$

Hence,

$$\begin{aligned}
G_n(f, x + \Delta x) - G_n(f, x) &= \sum_{k=-\infty}^{+\infty} f\left(\frac{k}{n}\right) (\phi(nx + n\Delta x - k) - \phi(nx - k)) \\
&= \sum_{k=-\infty}^{k_0-1} f\left(\frac{k}{n}\right) (\phi(nx + n\Delta x - k) - \phi(nx - k)) \\
&\quad + f\left(\frac{k_0}{n}\right) (\phi(nx + n\Delta x - k_0) - \phi(nx - k_0)) \\
&\quad + \sum_{k=k_0+1}^{+\infty} f\left(\frac{k_0}{n}\right) (\phi(nx + n\Delta x - k) - \phi(nx - k_0)) \\
&\geq \sum_{k=-\infty}^{k_0-1} f\left(\frac{k_0}{n}\right) (\phi(nx + n\Delta x - k) - \phi(nx - k_0)) \\
&\quad + f\left(\frac{k_0}{n}\right) (\phi(nx + n\Delta x - k_0) - \phi(nx - k_0)) \\
&\quad + \sum_{k=k_0+1}^{\infty} f\left(\frac{k_0}{n}\right) (\phi(nx + n\Delta x - k) - \phi(nx - k)) \\
&= f\left(\frac{k_0}{n}\right) \sum_{k=-\infty}^{+\infty} (\phi(nx + n\Delta x - k) - \phi(nx - k)) \\
&= 0,
\end{aligned}$$

which shows that for given n, $G_n(f, x)$ is a monotone increasing function in x. The proof of Theorem 4 is completed. □

References

[1] F. Brauer, C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, Springer-Verlag, New York, 2001, pp. 8–9.
[2] N. Hritonenko, Y. Yatsenko, Mathematical Modeling in Economics, Ecology and the Environment, Science Press, Beijing, 2006, pp. 92–93, Reprint.
[3] G. Cybenko, Approximation by superpositions of sigmoidal function, Math. of Control Signals, and System 2 (1989) 303–314.
[4] K.I. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Netw. 2 (1989) 183–192.
[5] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximation, Neural Netw. 2 (1989) 359–366.
[6] K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Netw. 3 (1990) 551–560.
[7] M. Leshno, V.Y. Lin, A. Pinks, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Netw. 6 (1993) 861–867.
[8] H.N. Mhaskar, C.A. Micchelli, Approximation by superposition of a sigmoidal function, Adv. Appl. Math. 13 (1992) 350–373.
[9] C.K. Chui, X. Li, Approximation by ridge functions and neural networks with one hidden layer, J. Approx. Theory 70 (1992) 131–141.
[10] T.P. Chen, 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.
[11] N. Hahm, B.I. Hong, An approximation by neural networks with a fixed weight, Comput. & Math. Appl. 47 (2004) 1897–1903.

[12] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inform. Theory 39 (1993) 930–945.
[13] H.N. Mhaskar, C.A. Miccheli, Degree of approximation by neural networks with a single hidden layer, Adv. Appl. Math. 16 (1995) 151–183.
[14] S. Suzuki, Constructive function approximation by three-layer artificial neural networks, Neural Netw. 11 (1998) 1049–1058.
[15] V. Maiorov, R.S. Meir, Approximation bounds for smooth functions in $C(R^d)$ by neural and mixture networks, IEEE Trans. Neural Networks 9 (1998) 969–978.
[16] Y. Makovoz, Uniform approximation by neural networks, J. Approx. Theory 95 (1998) 215–228.
[17] S. Ferrari, R.F. Stengel, Smooth function approximation using neural networks, IEEE Trans. Neural Networks 16 (2005) 24–38.
[18] Z.B. Xu, F.L. Cao, The essential order of approximation for neural networks, Sci. China Ser. F 47 (2004) 97–112.
[19] G.A. Anastassiou, Rate of convergence of some neural network operators to the unit-univariate case, J. Math. Anal. Appl. 212 (1997) 237–262.
[20] F.L. Cao, T.F. Xie, Z.B. Xu, The estimate for approximation error of neural networks: A constructive approach, Neurocomputing 71 (2008) 626–630.
[21] G.G. Lorentz, Approximation of Functions, Rinehart and Winston, New York, 1966.
[22] T.F. Xie, S.P. Zhou, Approximation Theory of Real Functions, Hangzhou University Press, Hangzhou, 1998.
[23] A. Zygmund, Trigonometric Series, Cambridge University Press, Cambridge, 1959.