J Syst Sci Complex (2011) 24: 401–412

NEURAL NETWORKS AND THE BEST TRIGOMOMETRIC APPROXIMATION*

Jianjun WANG · Zongben XU

DOI: 10.1007/s11424-011-8080-x
Received: 21 February 2008 / Revised: 23 July 2008

Abstract With the best trigonometric polynomial approximation as a metric, the rate of approximation of the one-hidden-layer feedforward neural networks to approximate an integrable function is estimated by using a constructive approach in this paper. The obtained result shows that for any 2π-periodic integrable function, a neural networks with sigmoidal hidden neuron can be constructed to approximate the function, and that the rate of approximation do not exceed the double of the best trigonometric polynomial approximation of function.

Key words Approximation, best trigonometric approximation, neural networks.

1 Introduction

As we know, artificial neural networks have been extensively applied in various fields of science and engineering[1−2]. In recent years, many researchers have done a number of researches on the problems in the artificial neural networks and have gained a series of results. Nowadays, artificial neural networks have been widely involved in studying problems in a variety of field, such as biology, mechanical engineering, electrical and computer engineering, computer science, and physics, etc. This is mainly because feedforward neural networks (FNNs) have the universal approximation capability[3−4]. A typical example of such universal approximation assertions states that for any given continuous function defined on a compact set $\boldsymbol{K}$ of $\mathcal{R}^d$, there exists a three-layer of FNNs so that it can approximate the function arbitrarily well. A three-layer of FNNs with one hidden layer, d inputs and one output can be mathematically expressed as

$$\mathcal{N}(x) = \sum_{i=1}^{m} c_i \phi\left(\sum_{j=1}^{d} w_{ij} x_j + \theta_i\right), \quad \boldsymbol{x} \in \mathcal{R}^d, \quad d \geq 1, \tag{1}$$

Jianjun WANG
School of Mathematics & Statistics, Southwest University, Chongqing 400715, *China*; *Institute for Information and System Science, Xi'an Jiaotong University, Xi'an* 710049, *China*. Email: wjj@swu.edu.cn.
Zongben XU
Institute for Information and System Science, Xi'an Jiaotong University, Xi'an 710049, *China*.
*This paper was supported by the National Basic Research Program of China (973 Program) under Grant No. 2007CB311000, the Natural Science Foundation of China under Grant Nos. 11001227, 60972155, 10701062, the Key Project of Chinese Ministry of Education under Grant No. 108176, Natural Science Foundation Project of CQ CSTC Nos. CSTC 2009BB2306, CSTC2009BB2305, the Fundamental Research Funds for the Central Universities under Grant No. XDJK2010B005, XDJK2010C023.
◇ *This paper was recommended for publication by Editor Jinhu LÜ.*

where $1 \leq i \leq m$, $\theta_i \in \mathcal{R}$ are the thresholds, $w_i = (w_{i1}, w_{i2}, \cdots, w_{id})^{\mathrm{T}} \in \mathcal{R}^d$ are connection weights of neuron i in the hidden layer with the input neurons, $c_i \in \mathcal{R}$ are the connection strength of neuron i with the output neuron, and ϕ is the activation function used in the network. The activation function is normally taken as sigmoid type, that is, it satisfies $\phi(t) \to 1$ as $t \to +\infty$ and $\phi(t) \to 0$ as $t \to -\infty$.

Later, many results about density and complexity were obtained by using different methods for more or less general situations[5–7]. However, such research works have only involved in the properties of convergence. In applied viewpoint, people care more for the quantitative problem. The works of related quantitative analysis have recently attracted intensive attention of people, especially in the topic of relationship between the converge rate of approximation and topology structure of hidden layer. In the past few years, the relationship between the rate of approximation and the number of hidden layer units have been studied and some useful results have been established, especially in characterizing accurately the topology selection and approximation order of neural networks[8–11].

In this paper, we take the best trigonometric polynomial approximation as a metric, establish the relationship between approximation speed of neural networks with a one-hidden-layer and the best trigonometric polynomial approximation, and explicitly calculate the number of hidden neurons needed for guaranteeing the predetermined approximation precision. The results characterize approximation ability of the constructive networks and clarify the relationship among the rate of approximation, the number of hidden-layer units and the properties of approximated function.

The remainder of this paper is organized as follows. In Section 2, we present some notations, basic concepts, and give the main result and remarks. Some fundamental lemmas are given in Section 3. In Section 4, we prove our main result and give remarks. And two numerical examples for illustrating theoretical results are given in Section 5, and in Section 6, we briefly summarize this study and indicate further study.

2 Notations and Main Results

We use the following notations. The symbols N, R, stand for the sets of natural and real numbers, respectively. Let $N_0 = N \cup \{0\}$ and $e_i = (0, \cdots, \overset{i}{1}, \cdots, 0) \in N_0^m$, And let $|\boldsymbol{r}| = \sum_{i=1}^m |r_i|$ for $\boldsymbol{r} = (r_i)_{i=1}^m \in N_0^m$, and $\boldsymbol{rt} = \sum_{i=1}^m r_i t_i$. For $p \geq 1$, we denote by $L_{2\pi}^p(R^m)$ the space of 2π-periodic (on each R of domain R^m) pth-order Lebesgue integrable functions on R^m to R with

$$\|f\|_p = (2\pi)^{-m}\left\{\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} |f(x)|^p dx\right\}^{\frac{1}{p}}.$$

Let

$$E_n(f) = \inf_{P \in T_n^m} \|f - P\|_p$$

denote the best trigonometric polynomial approximation of function f, where T_n^m is the set of trigonometric polynomial in $x_1, x_2, \cdots, x_m$ of degree n with the form $\sum\limits_{|r| \leq n} (a_r \sin(rx) + b_r \cos(rx))$. If there exists $t_n^* \in T_n^m$, such that

$$\|f - t_n^*\| = \inf_{t_n \in T_n^m} \|f - t_n\|_p,$$

then t_n^* is called trigonometric polynomial of the best approximation of function f.

By the above notations, our main result can be summarized as the following theorem.

Theorem 1 *For $f \in L^p_{2\pi}(R^m)(1 \le p \le \infty)$, then there exist FNN*

$$N(x) = \sum_{i=1}^{\aleph} C_i \phi(W_i \cdot x + \theta), \quad x \in R^m$$

with the sigmoidal hidden neuron number $\aleph = 2n^2(n+1)\rho$ (ρ is any integer not less than the reciprocal of $E_n(f)$) sigmoidal hidden-layer units, such that

$$\|N(x) - f\|_p \le 2E_n(f),$$

where $C_i \in R$, $W_i = (w_{i1}, w_{i2}, \cdots, w_{im}) \in R^m$.

Remark 1 This theorem reveals two things:

i) For any multivariate functions $f \in L^p_{2\pi}(R^m)$, there is feedforward neural network of the form (1) that approximates f arbitrarily well in $L^p_{2\pi}$, that is, the feedforward neural networks can be used as the universal approximator of functions in $L^p_{2\pi}(R^m)$;

ii) Quantitatively, the approximation accuracy of a network of the form (1) can attain the order of $2E_n(f)$, where $E_n(f)$ denotes the best approximation trigonometric polynomial of function f.

3 Lemmas

First, we show constructive approximation results to a multivariate trigonometric function by networks with piecewise linear and sigmoidal hidden-layer units of reference [12], which are more accurate than the corresponding results of reference [12].

Lemma 1 *For $\sigma \in N$, three-layer networks $PS_\sigma(\boldsymbol{rx})$ and $PC_\sigma(\boldsymbol{rx})$, which respectively approximate $\sin(\boldsymbol{rx})$ and $\cos(\boldsymbol{rx})$ and have $4|\boldsymbol{r}|\sigma$ piecewise linear hidden-layer units based on $PL_{\sigma,k}$, are constructed by*

$$PS_\sigma(\boldsymbol{rx}) = 2(-1)^{|\boldsymbol{r}|} \sin\frac{\pi}{4\sigma} \sum_{k=0}^{4|\boldsymbol{r}|\sigma-1} \cos\frac{(2k+1)\pi}{4\sigma} PL_{\sigma,k}(\boldsymbol{rx}) \tag{2}$$

and

$$PC_\sigma(\boldsymbol{rx}) = (-1)^{|\boldsymbol{r}|} - 2(-1)^{|\boldsymbol{r}|} \sin\frac{\pi}{4\sigma} \sum_{k=0}^{4|\boldsymbol{r}|\sigma-1} \sin\frac{(2k+1)\pi}{4\sigma} PL_{\sigma,k}(\boldsymbol{rx}), \tag{3}$$

where

$$PL_{\sigma,k}(\boldsymbol{rx}) = \begin{cases} 0, & \boldsymbol{rx} \le -|\boldsymbol{r}|\pi + \dfrac{k\pi}{2\sigma}, \\ \dfrac{2\sigma}{\pi}\boldsymbol{rx} + 2|\boldsymbol{r}|\sigma - k, & -|\boldsymbol{r}|\pi + \dfrac{k\pi}{2\sigma} < \boldsymbol{rx} < -|\boldsymbol{r}|\pi + \dfrac{(k+1)\pi}{2\sigma}, \\ 1, & \boldsymbol{rx} \ge -|\boldsymbol{r}|\pi + \dfrac{(k+1)\pi}{2\sigma}. \end{cases}$$

Then we have the following estimation

$$\|\sin(\boldsymbol{rx}) - PS_\sigma(\boldsymbol{rx})\|_p = \|\cos(\boldsymbol{rx}) - PC_\sigma(\boldsymbol{rx})\|_p \le \frac{1}{4\sigma^2}.$$

Proof we have the polygonal line with the vertex $(-|\boldsymbol{r}|\pi+\frac{k\pi}{2\sigma},\sin(-|\boldsymbol{r}|\pi+\frac{k\pi}{2\sigma}))$ and $(-|\boldsymbol{r}|\pi+\frac{(k+1)\pi}{2\sigma},\sin(-|\boldsymbol{r}|\pi+\frac{(k+1)\pi}{2\sigma}))$, $k=0,1,\cdots,4|\boldsymbol{r}|\sigma-1$. Using the vertex of polygonal line, $PL_{\sigma,k}(\boldsymbol{rx})$ and trigonometric formula, we have the network

$$\begin{aligned} PS_\sigma(\boldsymbol{rx}) &= \sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left\{\sin\left(-|\boldsymbol{r}|\pi+\frac{(k+1)\pi}{2\sigma}\right)-\sin\left(-|\boldsymbol{r}|\pi+\frac{k\pi}{2\sigma}\right)\right\}PL_{\sigma,k}(\boldsymbol{rx}) \\ &= 2(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\cos\frac{(2k+1)\pi}{4\sigma}PL_{\sigma,k}(\boldsymbol{rx}), \end{aligned}$$

which approximates $\sin(\boldsymbol{rx})$, where σ is a partition number of a quarter period of $\sin(\boldsymbol{rx})$. Obviously, $|PS_\sigma(\boldsymbol{rx})-\sin(\boldsymbol{rx})|$ has the maximum value about $\boldsymbol{rx}$ in the interval $[-|\boldsymbol{r}|\pi,|\boldsymbol{r}|\pi]$. We denote the point of arriving maximum value by $\boldsymbol{rx_0}$, and then we choose the $j\in N$ such that $\boldsymbol{rx_0}\in[-|\boldsymbol{r}|\pi+\frac{j\pi}{2\sigma},-|\boldsymbol{r}|\pi+\frac{(j+1)\pi}{2\sigma}]$. For convenience of consideration, we supposed $\boldsymbol{x_0}\le -|\boldsymbol{r}|\pi+\frac{(k+1/2)\pi}{2\sigma}$, then choose $|\boldsymbol{r}|h=\boldsymbol{rx_0}+|\boldsymbol{r}|\pi-\frac{k}{2\sigma}\pi$, so $\boldsymbol{rx_0}\pm h\in[-|\boldsymbol{r}|\pi+\frac{j\pi}{2\sigma},-|\boldsymbol{r}|\pi+\frac{(j+1)\pi}{2\sigma}]$ and because $PS_\sigma(\boldsymbol{rx})$ is linear, hence

$$\begin{aligned} &\sin(\boldsymbol{rx_0}+|\boldsymbol{r}|h)+\sin(\boldsymbol{rx_0}-|\boldsymbol{r}|h)-2\sin(\boldsymbol{rx_0}) \\ =&\sin(\boldsymbol{rx_0}+|\boldsymbol{r}|h)-PS_\sigma(\boldsymbol{rx_0}+|\boldsymbol{r}|h)+\sin(\boldsymbol{rx_0}-|\boldsymbol{r}|h) \\ &-PS_\sigma(\boldsymbol{rx_0}-|\boldsymbol{r}|h)-2\sin(\boldsymbol{rx_0})+2PS_\sigma(\boldsymbol{rx_0}). \end{aligned}$$

So by $\sin(\boldsymbol{rx_0}-|\boldsymbol{r}|h)=PS_\sigma(\boldsymbol{rx_0}-|\boldsymbol{r}|h)$ and the maximum value $|PS_\sigma(\boldsymbol{rx_0})-\sin(\boldsymbol{rx_0})|$ of $|PS_\sigma(\boldsymbol{rx})-\sin(\boldsymbol{rx})|$, we have

$$\begin{aligned} \|\sin(\boldsymbol{rx})-PS_\sigma(\boldsymbol{rx})\|_p &\le \|\sin(\boldsymbol{rx_0})-PS_\sigma(\boldsymbol{rx_0})\|_p \\ &\le \|2(\sin(\boldsymbol{rx_0})-PS_\sigma(\boldsymbol{rx_0}))-(\sin(\boldsymbol{rx_0}+|\boldsymbol{r}|h)-PS_\sigma(\boldsymbol{rx_0}+|\boldsymbol{r}|h)\|_p \\ &= \|\sin(\boldsymbol{rx_0}+|\boldsymbol{r}|h)+\sin(\boldsymbol{rx_0}-|\boldsymbol{r}|h)-2\sin(\boldsymbol{rx_0})\|_p \\ &\le (|\boldsymbol{r}|h)^2 \\ &\le \frac{1}{4\sigma^2}. \end{aligned}$$

From the Equation (2), the networks $PS_\sigma(\boldsymbol{rx})$ has $4|\boldsymbol{r}|\sigma$ hidden-layer units. We can construct $PC_\sigma(\boldsymbol{rx})$ in the same manner and prove the corresponding conclusion.

Lemma 2 *For $\sigma\in N$, three-layer networks $SS_\sigma(\boldsymbol{rx})$ and $SC_\sigma(\boldsymbol{rx})$, which respectively approximate $\sin(\boldsymbol{rx})$ and $\cos(\boldsymbol{rx})$ and have $4|\boldsymbol{r}|\sigma$ sigmoidal hidden-layer units based on $SG_{\sigma,k}$, are constructed by*

$$SS_\sigma(\boldsymbol{rx})=2(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\cos\frac{(2k+1)\pi}{4\sigma}SG_{\sigma,k}(\boldsymbol{rx}) \tag{4}$$

and

$$SC_\sigma(\boldsymbol{rx})=(-1)^{|\boldsymbol{r}|}-2(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\sin\frac{(2k+1)\pi}{4\sigma}SG_{\sigma,k}(\boldsymbol{rx}), \tag{5}$$

where $SG_{\sigma,k}(\boldsymbol{rx}) = [1+\exp\{-(\frac{8\sigma}{\pi}\boldsymbol{rx}+8|\boldsymbol{r}|\sigma-4k-2)\}]^{-1}$. *Then we have the following estimation*

$$\|\sin(\boldsymbol{rx}) - SS_\sigma(\boldsymbol{rx})\|_p = \|\cos(\boldsymbol{rx}) - SC_\sigma(\boldsymbol{rx})\|_p \le \left\{\frac{|\boldsymbol{r}|}{2\sigma}\ln\frac{4e^3+4e}{2e^2+e^4+1} + \frac{1}{4\sigma^2}\right\}^{\frac{1}{p}}. \tag{6}$$

Proof We replace $PL_{\sigma,k}$ in (2) with $SG_{\sigma,k}(\boldsymbol{rx})$, and obtain the networks denoted by $SS_\sigma(\boldsymbol{rx})$. Then it is a network with $4|\boldsymbol{r}|\sigma$ sigmoidal hidden-layer units based on $SG_{\sigma,k}(\boldsymbol{rx})$. If $\boldsymbol{r} = \boldsymbol{0}$, the result is obvious. Since

$$\begin{aligned}
&\|SS_\sigma(\boldsymbol{rx}) - PS_\sigma(\boldsymbol{rx})\|_1 \\
&= 2\sin\frac{\pi}{4\sigma}\left|\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\cos\frac{(2k+1)\pi}{4\sigma}\frac{1}{(2\pi)^m}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}(PL_{\sigma,k}(\boldsymbol{rx}) - SG_{\sigma,k}(\boldsymbol{rx}))d\boldsymbol{x}\right| \\
&\le 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|\cdot\left|\frac{1}{(2\pi)^m}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}(PL_{\sigma,k}(\boldsymbol{rx}) - SG_{\sigma,k}(\boldsymbol{rx}))d\boldsymbol{x}\right| \\
&= 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|\cdot\frac{1}{(2\pi)^m}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\int_{\frac{\boldsymbol{rx}-r_mx_m-r_m}{|\boldsymbol{r}|}}^{\frac{\boldsymbol{rx}-r_mx_m+r_m}{|\boldsymbol{r}|}}\frac{|\boldsymbol{r}|}{r_m} \\
&\quad |PL_{\sigma,k}(|\boldsymbol{r}|y) - SG_{\sigma,k}(|\boldsymbol{r}|y)|dx_1dx_2\cdots dx_{x_{m-1}}dy \\
&\le 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|\cdot\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|\boldsymbol{r}|}{r_m}\left|PL_{\sigma,k}(|r|y) - SG_{\sigma,k}(|r|y)\right|dy\right| \\
&= 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|\cdot\left|\frac{1}{2\pi r_m}\int_{-|\boldsymbol{r}|\pi}^{|\boldsymbol{r}|\pi}\left|PL_{\sigma,k}(t) - SG_{\sigma,k}(t)\right|dt\right|.
\end{aligned}$$

The following we will estimate $\frac{1}{2\pi r_m}\int_{-|\boldsymbol{r}|\pi}^{|\boldsymbol{r}|\pi}|PL_{\sigma,k}(t) - SG_{\sigma,k}(t)|dt$.

$$\begin{aligned}
&\frac{1}{2\pi r_m}\int_{-|\boldsymbol{r}|\pi}^{|\boldsymbol{r}|\pi}|PL_{\sigma,k}(t) - SG_{\sigma,k}(t)|dt \\
&= \frac{1}{2\pi r_m}\left\{\int_{-|\boldsymbol{r}|\pi}^{-\pi+\frac{k\pi}{2\sigma}}SG_{\sigma,k}(t)dt + \int_{-\pi+\frac{k\pi}{2\sigma}}^{-\pi+\frac{(k+1)\pi}{2\sigma}}|SG_{\sigma,k}(t) - PL_{\sigma,k}(t)|dt\right. \\
&\qquad\left. + \int_{-\pi+\frac{(k+1)\pi}{2\sigma}}^{|\boldsymbol{r}|\pi}(1 - SG_{\sigma,k}(t))dt\right\} \\
&= \frac{1}{2\pi r_m}\left\{\int_{-|\boldsymbol{r}|\pi}^{-\pi+\frac{k\pi}{2\sigma}}\frac{dt}{1+e^{-8\sigma t/\pi-8\sigma+4k+2}} + \int_{\frac{k\pi}{2\sigma}}^{\frac{(k+1)\pi}{2\sigma}}\left|\frac{2\sigma}{\pi}(t-\pi) + 2\sigma - k -\right.\right. \\
&\qquad\left.\left.\frac{1}{1+e^{-8\sigma(t-\pi)/\pi-8\sigma+4k+2}}\right|dt + \int_{-\pi+\frac{(k+1)\pi}{2\sigma}}^{|\boldsymbol{r}|\pi}\frac{e^{-8\sigma t/\pi-8\sigma+4k+2}}{1+e^{-8\sigma t/\pi-8\sigma+4k+2}}dt\right\} \\
&\doteq I_1 + I_2 + I_3.
\end{aligned}$$

For I_1, we have

$$\begin{aligned} I_1 &= \frac{1}{2\pi r_m}\int_{(-|\boldsymbol{r}|+1)\pi}^{\frac{k\pi}{2\sigma}} \frac{1}{1+\mathrm{e}^{-8\sigma(t-\pi)/\pi-8\sigma+4k+2}}dt \\ &= \frac{1}{2\pi r_m}\int_{(-|\boldsymbol{r}|+1)\pi}^{\frac{k\pi}{2\sigma}} \frac{1}{1+\mathrm{e}^{-8\sigma t/\pi-8\sigma+4k+2}}dt \\ &= \frac{1}{2\pi r_m}\int_{(-|\boldsymbol{r}|+1)(-8\sigma)}^{-4k} \frac{1}{1+\mathrm{e}^{t+4k+2}}dt \\ &= -\frac{1}{16\sigma r_m}\int_{(-|\boldsymbol{r}|+1)(-8\sigma)+4k}^{0} \frac{1}{1+\mathrm{e}^{t+2}}dt \\ &\le \frac{1}{16\sigma r_m}(\ln(1+\mathrm{e}^2)-2) \end{aligned}$$

and

$$\begin{aligned} I_2 &= \frac{1}{2\pi r_m}\int_{\frac{k\pi}{2\sigma}}^{\frac{(k+1)\pi}{2\sigma}} \left|\frac{2\sigma}{\pi}t-k-\frac{1}{1+\mathrm{e}^{-8\sigma t/\pi+4k+2}}\right|dt \\ &= \frac{1}{2\pi r_m}\int_{k}^{k+1} \left|t-k-\frac{1}{1+\mathrm{e}^{-4t+4k+2}}\right|dt \\ &= \frac{1}{2\sigma r_m}\int_{0}^{1} \left|x-\frac{1}{1+\mathrm{e}^{-4x+2}}\right|dt \\ &\le \frac{1}{8\sigma r_m}[2\ln 2-\ln(2+\mathrm{e}^{-2}+\mathrm{e}^2)+1]. \end{aligned}$$

Similarly, we have

$$I_3 \le \frac{1}{16\sigma r_m}\ln(1+\mathrm{e}^{-2}).$$

So we have

$$\begin{aligned} \|SS_\sigma(\boldsymbol{rx})-PS_\sigma(\boldsymbol{rx})\|_1 &\le 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|(I_1+I_2+I_3) \\ &\le 4|\boldsymbol{r}|\frac{1}{8\sigma r_m}\ln\frac{4\mathrm{e}^3+4e}{2\mathrm{e}^2+\mathrm{e}^4+1} \\ &= \frac{|\boldsymbol{r}|}{2\sigma r_m}\ln\frac{4\mathrm{e}^3+4e}{2\mathrm{e}^2+\mathrm{e}^4+1}. \end{aligned} \tag{7}$$

By using the Lemma 1 and (7), the result can be derived in the case of $p=1$. According to the construction of network $SS_\sigma(\boldsymbol{rx})$, we have $0\le|SS_\sigma(\boldsymbol{rx})-PS_\sigma(\boldsymbol{rx})|\le 1$. So we obtain (6) from $\|SS_\sigma(\boldsymbol{rx})-PS_\sigma(\boldsymbol{rx})\|_p\le\|SS_\sigma(\boldsymbol{rx})-PS_\sigma(\boldsymbol{rx})\|_1^{1/p}$.

From Lemma 2, we can easily obtain the following Corollary.

Corollary 1 *For all $\varepsilon>0$, for $t(x)=\sum_{|r|\le n}(a_r\sin(rx)+b_r\cos(rx))$, there exist FNNs*

$$\begin{aligned} N(x) = \sum_{|r|\le n} b_r(-1)^{|\boldsymbol{r}|} + 2\sum_{|r|\le n}(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\Big\{a_r\cos\frac{(2k+1)\pi}{4\sigma} \\ -b_r\sin\tfrac{(2k+1)\pi}{4\sigma}\Big\}SG_{\sigma,k}(\boldsymbol{rx}) \end{aligned} \tag{8}$$

which have $2n^2(n+1)\tau$ *(*τ *is any integer not less than the reciprocal of the preset approximation precision* ε*) sigmoidal hidden-layer units, such that* $\|N-t\|_p<\varepsilon$.

4 Proof of Theorem 1

The following we prove Theorem 1. Let $t_n^*(x)$ be the *n-th* best trigonometric approximation polynomial of function f, then from corollary, we can construct a one-hidden-layer neural networks

$$N(x)=\sum_{|r|\le n}b_r(-1)^{|\boldsymbol{r}|}+2\sum_{|r|\le n}(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\left\{a_r\cos\frac{(2k+1)\pi}{4\sigma}\right.$$
$$\left.-b_r\sin\tfrac{(2k+1)\pi}{4\sigma}\right\}SG_{\sigma,k}(\boldsymbol{rx}) \tag{9}$$

such that

$$\|t_n^*-N\|_p\le\varepsilon.$$

By triangle inequality

$$\|f-N\|_p\le\|t_n^*-N\|_p+\|f-t_n^*\|_p\le\varepsilon+E_n(f),$$

and taking $\varepsilon\le E_n(f)$, we can deduce our results.

Remark 2 The activation function of neural networks is the standard sigmoidal function $\phi(x)=1/(1+\mathrm{e}^{-x})$ in Theorem 1. It is not hard to see from the proof of Lemma 2 that the result can in fact be generalized to activation functions obeying the following conditions.

1) There is a constant C_ϕ such that $|\ \phi^{(k)}(x)\ |\ge C_\phi>0, k=0,1,\cdots$

2) For each finite k, there is a finite constant l_k such that$|\ \phi^{(k)}(x)\ |\le l_k$.

5 Numerical Examples

In this section, we present some numerical experiments to demonstrate the validity of the obtained results and suggest the error bound of neural networks approximation.

We select a continuous function $f_1(x)=\sin(5x)+0.25\cos(5x), x\in[-\pi,\pi]$, and a 2π- periodic integrable function $f_2(x)\in L^2(R)$

$$f_2(x)=\begin{cases}-1, & -\pi\le x<0;\\ 1, & 0\le x<\pi.\end{cases}$$

as target functions and investigate the approximate networks with sigmoid activation function $\phi(x)=1/(1+\mathrm{e}^{-x})$ over the interval $[-\pi,\pi]$.

Clearly, $f_1(x)\in C[-\pi,\pi]$, $f_2(x)\in L^2([-\pi,\pi])$. The best approximation trigonometric polynomial of target functions are $t_n(x)=0(n\le 4)$ and $t_n(x)=\frac{2}{\pi}\sum_{k=1}^n(1-(-1)^k)\frac{\sin(kx)}{k}$, respectively. So we can obtain the best approximation of target function

$$E_n(f_1)=\sqrt{1^2+0.25^2}=\frac{\sqrt{17}}{4},\quad n\le 4$$

and

$$E_n(f_2)=\sqrt{2\pi-\frac{4}{\pi}\sum_{k=1}^n\frac{(1-(-1)^k)^2}{k^2}}.$$

From Theorem 1, it follows that

$$\|f_1 - N(x)\| \le \frac{\sqrt{17}}{2}$$

and

$$\|f_2 - N\| \le 2\sqrt{2\pi - \frac{4}{\pi}\sum_{k=1}^{n}\frac{(1-(-1)^k)^2}{k^2}}.$$

The following Table 1 shows some approximation errors of the target function $f_1(x)$, and Figures 1–6 then demonstrate that the target function $f_1(x)$ is well approximated. The Table 2, Figures 7–12 show the corresponding ones of the target function $f_2(x)$.

Table 1 Approximation error for target function $f_1(x)$

σ	1	2	3	5	10	20
$\max_{x\in[-\pi,\pi]} \|N(x) - f_1(x)\|$	0.2733	0.0917	0.0450	0.0142	0.0041	0.0008
$2E_n(f_1)(n \le 4)$	2.0616	2.0616	2.0616	2.0616	2.0616	2.0616

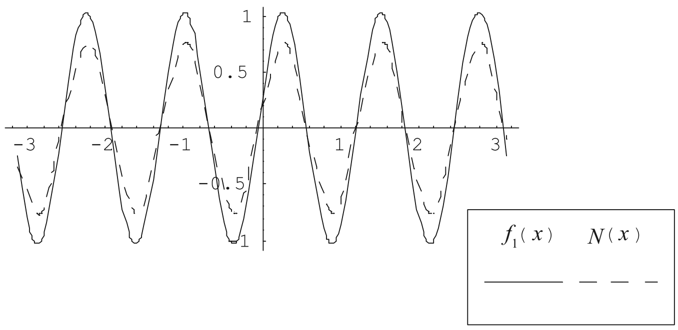

Figure 1 $\sigma = 1$ case

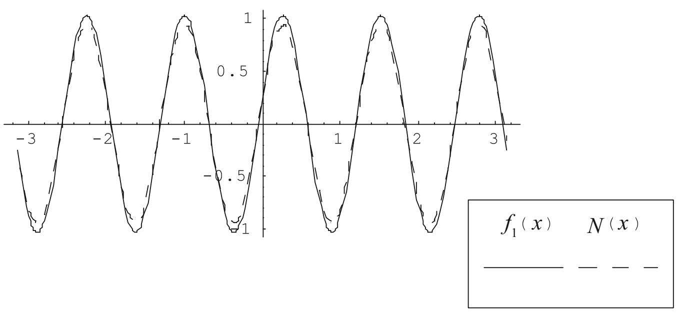

Figure 2 $\sigma = 2$ case

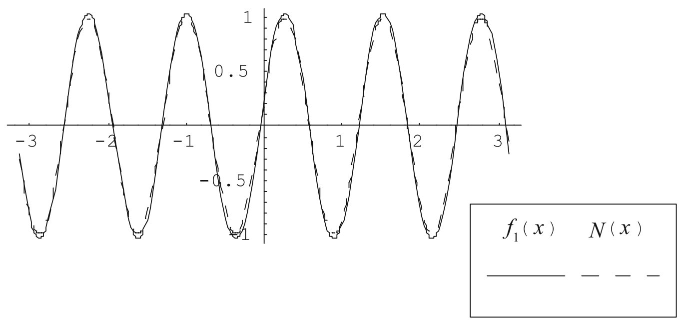

Figure 3 $\sigma = 3$ case

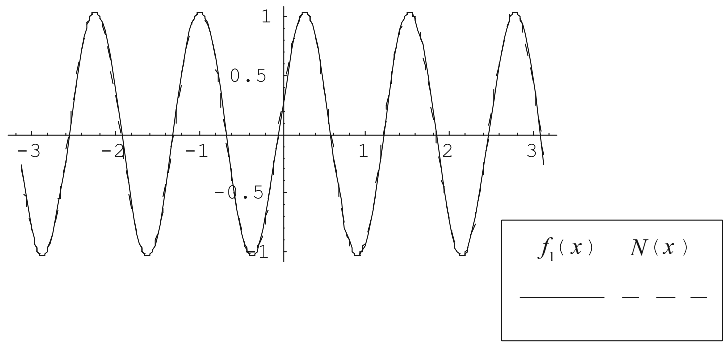

Figure 4 $\sigma = 5$ case

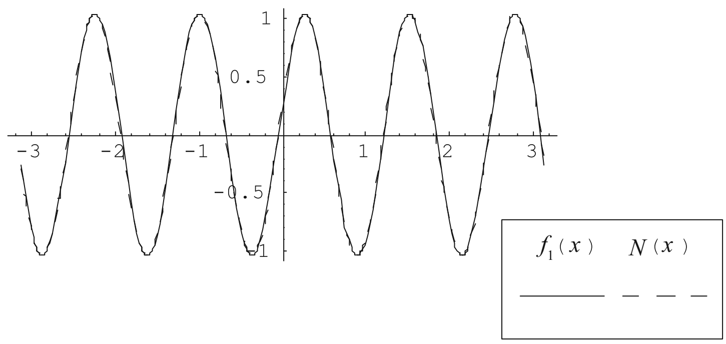

Figure 5 $\sigma = 10$ case

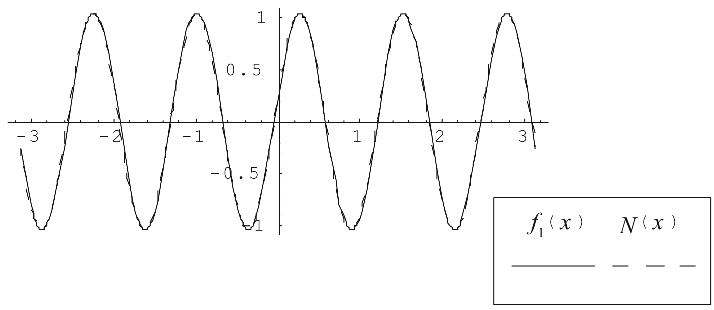

Figure 6 $\sigma = 20$ case

Table 2 Approximation error for target function $f_2(x)$

σ	1	2	5	10	15	20
n	1	10	20	15	25	30
$\|N - f_2\|_2$	0.75847	0.205624	0.139885	0.156562	0.124095	0.115725
$2E_n(f_2)$	2.18195	1.00759	0.713353	0.797368	0.625758	0.582585

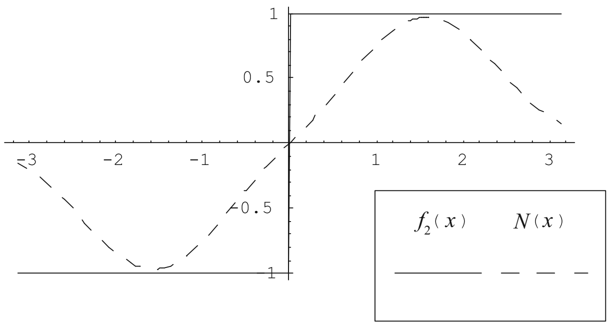

Figure 7 $\sigma = n = 1$ case

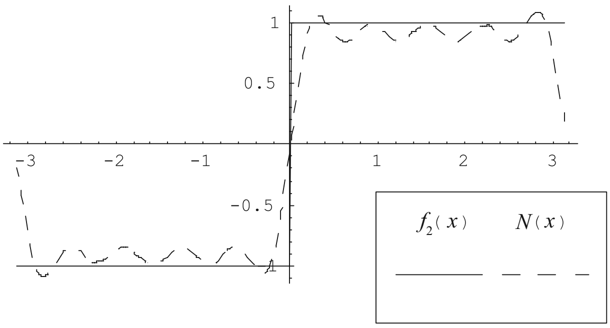

Figure 8 $\sigma = 2, n = 10$ case

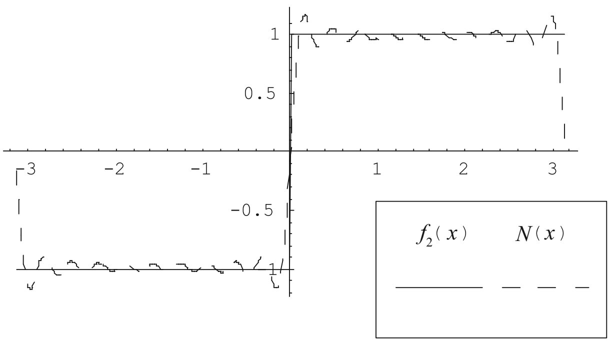

Figure 9 $\sigma = 5, n = 20$ case

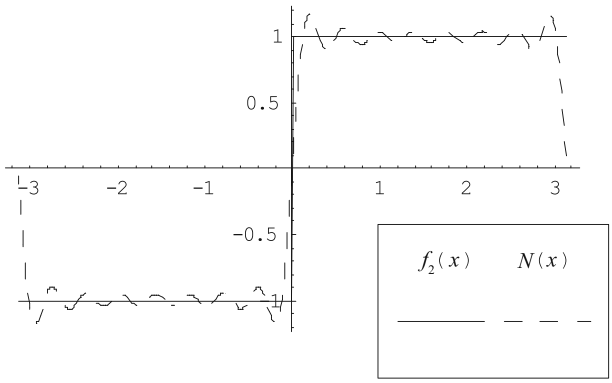

Figure 10 $\sigma = 10, n = 15$ case

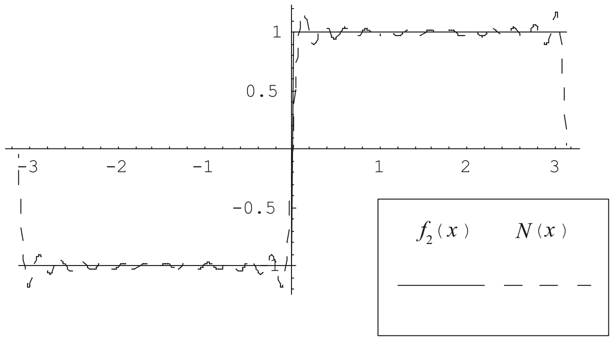

Figure 11 $\sigma = 15, n = 25$ case

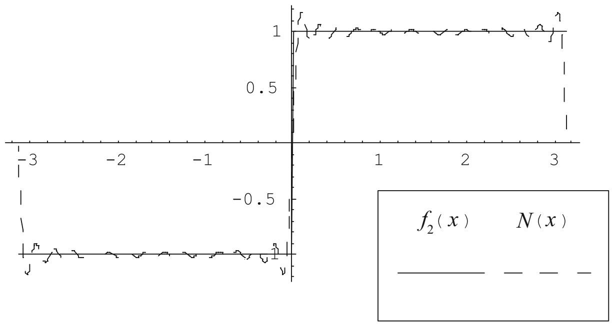

Figure 12 $\sigma = 20, n = 30$ case

6 Conclusions

In this paper, approximation estimations of the neural networks have been studied. In terms of the best trigonometric approximation of a function, an upper bound estimations on approximation precision and speed of the neural networks are simultaneously developed. Our research reveals that the approximation precision and speed of the neural networks depend not only on the number of hidden neurons used, but also the best trigonometric approximation of functions to be approximated. We have explicitly given a lower bound estimation on the number of hidden neurons of the network in order to attain a predetermined approximation precision. The results obtained are helpful in understanding the approximation capability and topology construction of the neural networks.

References

[1] C. G. Looney, *Pattern Recognition Using Neural Networks*, Oxford University Press, New York, 1997.

[2] C. R. Zhang and B. D. Zheng, Hopf bifurcation in numerical approximation of an n-dimension neural network model with multi-delays, *Chaos, Solitons & Fractals*, 2005, **25**: 129–146.

[3] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals System*, 1989, **2**: 303–314.

[4] K. Hornik, M. Stinchombe, and H. White, Multilayer feedforward networks are universal approximator, *Neural Networks*, 1989, **2**: 359–366.

[5] T. P. Chen, Approximation problems in system identification with neural networks, *Science in China Series A*, 1994, **24(1)**: 1–7.

[6] 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 Transaction on Neural Networks*, 1995, **6**: 911–917.

[7] J. G. Attali and G. Pages, Approximation of functions by a multilayer perceptron: A new approach, *Neural Networks*, 1997, **10**: 1069–1081.

[8] F. L. Cao, Z. B. Xu, and J. Y. Liang, Approximation of polynomial functions by neural networks: Construction of network and algorithm of approximation, *Chinese Journal of Computers*, 2003, **26(8)**: 906–912.

[9] Z. B. Xu and J. J. Wang, The essential order of approximation for nearly exponential type neural networks, *Science in China Series F*, 2006, **49(4)**: 446–460.

[10] Z. B. Xu and F. L. Cao, The essential order of approximation for neural networks, *Science in China Series F*, 2004, **47**: 97–112.

[11] F. L. Cao, Y. Q. Zhang, and W. G. Zhang, Neural Networks with single hidden layer and the best polynomial approximation, *Acta Mathematica Sinica*, 2007, **50(2)**: 385–392.

[12] S. Suzuki, Constructive function approximation by three-layer artificial neural networks, *Neural Networks*, 1998, **11**: 1049–1058.