International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 10, No. 3 (2012) 1250021 (18 pages)

DOI: 10.1142/S021969131250021X

CONSTRUCTIVE ESTIMATION OF APPROXIMATION FOR TRIGONOMETRIC NEURAL NETWORKS

JIANJUN WANG*

School of Mathematics and Statistics
Southwest University
Chongqing 400715, P. R. China
wjj@swu.edu.cn

WEIHUA XU

School of Mathematics and Physics
Chongqing University of Technology
Chongqing 400054, P. R. China
xuweihua@cqut.edu.cn

BIN ZOU

Faculty of Mathematics and Computer Science
Hubei University, Wuhan 430062, P. R. China
zoubin0502@hubu.edu.cn

Received 16 August 2010
Revised 4 November 2011
Published 18 May 2012

For the three-layer artificial neural networks with trigonometric weights coefficients, the upper bound and lower bound of approximating 2π-periodic pth-order Lebesgue integrable functions $L^p_{2\pi}$ are obtained in this paper. Theorems we obtained provide explicit equational representations of these approximating networks, the specification for their numbers of hidden-layer units, the lower bound estimation of approximation, and the essential order of approximation. The obtained results not only characterize the intrinsic property of approximation of neural networks, but also uncover the implicit relationship between the precision (speed) and the number of hidden neurons of neural networks.

Keywords: Order of approximation; neural networks; trigonometric; modulus of smoothness.

AMS Subject Classification: 41A36, 41A25

*Corresponding author.

1. Introduction

In recent years, researchers have done many researches on the problems of artificial neural networks and gain a series of results. Nowadays, the artificial neural networks have widely been applied in variety of fields, such as biology, mechanical engineering, electrical and computer engineering, computer science, and physics, etc.[1,3,5,17] The application problems have been occasionally converted into the problems of utilizing an underlying artificial neural network as an approximation function.[14,15,4,3,8,16,10] Although the approximation ability of artificial neural networks has been sufficiently discussed in some earlier articles,[3,8,16,10] the works of related quantitative analysis is recently gave rise to the strong attention of the people, especially on the topic of relationship between the converge rate of approximation and the structural topology of hidden layer.[2,11,9,13] In fact, the estimate of the approximation upper bound with an underlying network which is bounded by the upper bound of converge rate is clearly expressed.[2,12] Having only the estimation of upper bound towards the forecast ability of an underlying network is still unsatisfactory. The upper bound reflect ability level of the underlying network. Hence, the work of consecutive estimation of the lower bound is necessary. In addition, the degree of approximation bound, regarding the estimation precision of the approximation if and only if the underlying network can achieve, had better be given out to guarantee the exact level of the ability for the underlying network. In order to get such an essential order of approximation of a neural network, besides upper bound estimation, a lower bound estimation that characterizes the worst approximation precision of the network can also be expected. The essential order of approximation can be obtained when and only when the upper and the lower bound estimations are of the same order. Clearly, obtaining the essential order of a neural network is not easy, but very important and is of significance. In Refs. 14 and 15, such are tackled for the neural networks.

Among the previous researches, Suzuki[13] obtained an upper bound estimation on approximation of the neural networks. In this paper, we not only give the upper bound estimation, which sharpens the result in Ref. 13, but also give the lower bound estimation of the approximation and explicitly calculated the number of hidden neurons needed for guaranteeing the predetermined approximation precision, the obtained results clarify the relationship between the approximation speed (precision) and the number of hidden neurons needed for the neural networks.

The remainder of this paper is organized as follows. In Sec. 2, we present some notations, basic concepts, and give the main result and remarks. Some fundamental lemmas are given in Sec. 3. In Sec. 4, we prove our main result and give remarks. Section 5 briefly summarizes our conclusions and indicates further study.

2. Notations and Main Results

The following notations are used through the paper. The symbols $\mathbf{N}, \mathbf{R}$, stand for the sets of natural and real numbers, respectively. Let $N_0 = N \cup \{0\}$,

$e_i = \big(0, \dots, \overset{i}{1}, \dots, 0\big) \in N_0^m$, $|\mathbf{r}| = \sum_{i=1}^m |r_i|$, $\mathbf{rt} = \sum_{i=1}^m r_i t_i$ and $\|t\| = (\sum_{i=1}^m t_i^2)^{1/2}$ for $\mathbf{r} = (r_1, r_2, \dots, r_m) \in N_0^m$, $\mathbf{t} = (t_1, t_2, \dots, t_m) \in R^m$. 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

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

and by $C_{2\pi}(R^m)$ the space of 2π-periodic continuous functions on R^m to R with

$$\|f\|_\infty = \sup_{|x_i| \leq \pi} |f(x)|, \quad x \in R^m.$$

For convenience, we denote by $L_{2\pi}^\infty(R^m)$ the space $C_{2\pi}(R^m)$. So for $f, g \in L_{2\pi}^p(R^m)$, we define the convolution

$$(f * g)(x) = (2\pi)^{-m} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} f(\mathbf{t}) g(\mathbf{x} - \mathbf{t}) d\mathbf{t},$$

and the Fourier transformation

$$\hat{f}(\mathbf{r}) = \langle f, e^{-i\mathbf{rt}} \rangle,$$

where

$$\langle f, g \rangle = \frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} f(\mathbf{t}) g(\mathbf{t}) d\mathbf{t}.$$

In order to characterize the approximation ability of neural networks, the higher-order modulus of continuity and Lipschitz condition for a multivariate function are defined as follows, which measure the variation and the smoothness of a function.

Definition 2.1 (Higher-order modulus of continuity and Lipschitz condition). Let $f \in L_{2\pi}^p(R^m)$ and $\delta > 0$, the modulus of continuity of f in $L_{2\pi}^p(R^m)$ is

$$\omega(f, \delta) = \sup_{\|x_1 - x_2\| \leq \delta} \|f(x_1) - f(x_2)\|_p,$$

the rth-order modulus of continuity of f is defined by

$$\omega_r(f, t) = \sup_{\mathbf{x} \pm \frac{h\mathbf{e}}{2} \in Q \subseteq R^m, \|h\| \leq t} \|\Delta_h^r f(\mathbf{x})\|_p,$$

where $\Delta_h^r f(\mathbf{x}) = \sum_{i=0}^r (-1)^{r-i} \binom{r}{i} f(x + (\frac{r}{2} - i) h \mathbf{e_i})$ is the rth-order symmetric difference of f along direction e_i and with step-length h. A function f is said to belong to the α-Lipschitz class, denoted by $f \in \mathrm{Lip}(\alpha)_2$, if the second-order modulus of continuity $\omega_2(f, t) = O(t^\alpha)$, where α is a positive real parameter.

For all $f_i \in L_{2\pi}^p(R)$, we say $F = (f_1, f_2, \dots, f_m)$ is a $F \in L_{2\pi}^p(R^m)$ function on $R^m \to R^m$. For $F = (f_1, f_2, \dots, f_m)$, Suzuki[13] constructed a three-layer feedforward neural networks

$$PN_{\lambda,\sigma}[F] = (PN_{\lambda,\sigma}[f_1], \dots, PN_{\lambda,\sigma}[f_m])^T,$$

here,

$$PN_{\lambda,\sigma}[f_i](x) = \theta_{\lambda,\sigma}[f_i] + \sum_{\text{Combinations of } \mathbf{p}\neq\mathbf{q}}^{0\leq p_u, q_v\leq\lambda} \times \sum_{k=0}^{4|\mathbf{p}-\mathbf{q}|\sigma-1} \alpha_{\lambda,\sigma,\mathbf{p},\mathbf{q},k}[f_i] PL_{\sigma,k}((\mathbf{p}-\mathbf{q})\mathbf{x}),$$

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

Note this summation is over combinations of $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ and $q = (q_1, q_2, \ldots, q_m) \in N_0^m$ such that $\mathbf{p} \neq \mathbf{q}, 0 \leq p_u, q_v \leq \lambda$.

In addition, Suzuki constructed another three-layer feedforward neural networks

$$SN_{\lambda,\sigma}[F] = (SN_{\lambda,\sigma}[f_1], \ldots, SN_{\lambda,\sigma}[f_m])^T,$$

here

$$SN_{\lambda,\sigma}[f_i](x) = \theta_{\lambda,\sigma}[f_i] + \sum_{\text{Combinations of } \mathbf{p}\neq\mathbf{q}}^{0\leq p_u, q_v\leq\lambda} \times \sum_{k=0}^{4|\mathbf{p}-\mathbf{q}|\sigma-1} \alpha_{\lambda,\sigma,\mathbf{p},\mathbf{q},k}[f_i] SG_{\sigma,k}((\mathbf{p}-\mathbf{q})\mathbf{x}),$$

$$SG_{\sigma,k}(\mathbf{rx}) = \left[1 + \exp\left\{-\left(\frac{8\sigma}{\pi}\mathbf{rx} + 8|\mathbf{r}|\sigma - 4k - 2\right)\right\}\right]^{-1},$$

$$\theta_{\lambda,\sigma}[f_i] = \langle f_i, 1\rangle + 2\left(\frac{2}{\lambda+2}\right)^m \sin\frac{\pi}{4\sigma} \sum_{\text{Combinations of } \mathbf{p}\neq\mathbf{q}}^{0\leq p_u, q_v\leq\lambda} \times (-1)^{|\mathbf{p}-\mathbf{q}|} b_{\lambda,\mathbf{p}} b_{\lambda,\mathbf{q}} \langle f_i, \cos(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle,$$

and

$$\alpha_{\lambda,\sigma,\mathbf{p},\mathbf{q},k} = 4(-1)^{|\mathbf{p}-\mathbf{q}|}\left(\frac{2}{\lambda+2}\right)^m b_{\lambda,\mathbf{p}} b_{\lambda,\mathbf{q}} \sin\frac{\pi}{4\sigma} \times \left\{\langle f_i, \sin(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle \cos\frac{(2k+1)\pi}{4\sigma} - \langle f_i, \cos(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle \sin\frac{(2k+1)\pi}{4\sigma}\right\}.$$

Suzuki obtained the upper bound of the above two neural networks. Our approach will be based on establishing upper bound (more exact than the corresponding result in Ref. 13) and lower bound estimation on the approximation order of this networks. The main results we obtain can be summarized as the following theorems.

Theorem 2.1. *Let $f \in L^p_{2\pi}$, for independent λ and $\sigma \in N$, there is a three-layer network $PN_{\lambda,\sigma}[f]$ with $2m\sigma\lambda(\lambda+1)(2\lambda+1)^{m-1}$ piecewise linear hidden-layer units of $PL_{\sigma,k}$ such that*

$$\|f_i - PN_{\lambda,\sigma}[f_i]\|_p \leq \frac{1}{2}\left(1+\sqrt{\frac{m}{2}}\pi^2\right)^2 \omega_2\left(f_i, \frac{1}{\lambda+2}\right) + \frac{1}{2\sigma^2}\|f_i\|_p\left(\left(\frac{8(\lambda+2)}{\pi}\right)^m - \left(\frac{2}{\lambda+2}\right)^m\right), \tag{2.1}$$

$$\omega_2\left(f_i, \frac{1}{\lambda+2}\right) \leq C\left\{\frac{1}{\lambda^2}\sum_{k=1}^{\lambda} k\|PN_{k,\sigma}[f_i] - f_i\|_p + \left(\frac{\lambda^m}{\sigma^2} + \frac{1}{\lambda^2}\right)\|f_i\|_p\right\}, \tag{2.2}$$

and for $0 < \alpha \leq 2$, there holds the following essential order estimation:

$$\|f_i - PN_{\lambda,\sigma}[f_i]\|_p = O(\lambda^{-\alpha}) \Leftrightarrow \sigma^2 = O(\lambda^{m+\alpha}) \quad \text{and} \quad f_i \in \mathrm{Lip}(\alpha)_2. \tag{2.3}$$

Here and hereafter, C, C_1, C_2, C_3 are positive constants independent of n, f, and x (its value, however, may be different in different contexts).

Theorem 2.2. *Let $f \in L^p_{2\pi}$, for independent λ and $\sigma \in N$, there is a three-layer network $SN_{\lambda,\sigma}[f]$ with $2m\sigma\lambda(\lambda+1)(2\lambda+1)^{m-1}$ sigmoidal hidden-layer units of $SG_{\sigma,k}$ such that*

$$\|f_i - SN_{\lambda,\sigma}[f_i]\|_p \leq \frac{1}{2}\left(1+\sqrt{\frac{m}{2}}\pi^2\right)^2 \omega_2\left(f_i, \frac{1}{(\lambda+2)}\right) + \|f_i\|_p\left(\left(\frac{8(\lambda+2)}{\pi}\right)^m - \left(\frac{2}{\lambda+2}\right)^m\right) \times \left\{\frac{|r|}{2\sigma}\ln\frac{4e^3+4e}{2e^2+e^4+1} + \frac{1}{4\sigma^2}\right\}, \tag{2.4}$$

$$\omega_2\left(f_i, \frac{1}{\lambda+2}\right) \leq C\left\{\frac{1}{\lambda^2}\sum_{k=1}^{\lambda} k\|SN_{k,\sigma}[f_i] - f_i\|_p + \left(\frac{\lambda^m}{\sigma^2} + \frac{\lambda^m}{\sigma} + \frac{1}{\lambda^2}\right)\|f_i\|_p\right\}, \tag{2.5}$$

and for $0 < \alpha \leq 2$, there holds the following essential order estimation:

$$\|f_i - SN_{\lambda,\sigma}[f_i]\|_p = O(\lambda^{-\alpha}) \Leftrightarrow \sigma = O(\lambda^{m+\alpha}) \quad \text{and} \quad f_i \in \mathrm{Lip}(\alpha)_2. \tag{2.6}$$

Remark 2.1. (1) If σ is a higher-order infinity than $\lambda^{\frac{m}{2}}$, then $\|f_i - PN_{\lambda,\sigma}[f_i]\|_p \to 0$ as $\lambda \to \infty$. (2) If σ is a higher-order infinity than λ^m, then $\|f_i - SN_{\lambda,\sigma}[f_i]\|_p \to 0$

as $\lambda \to \infty$. They therefore also show that any 2π-periodic on each R of domain R^m Lebesgue integrable functions f can be approximated arbitrarily well by the above networks.

Remark 2.2. In Ref. 13, the author obtained the following upper bounds:

$$\|f_i - PN_{\lambda,\sigma}[f_i]\|_p \le \left(1 + \frac{\sqrt{m}\pi^2}{2}\right)\omega\left(f_i, \frac{1}{\lambda+2}\right) + 2\|f_i\|_p \left\{\left(\frac{8(\lambda+2)}{\pi^2}\right)^m - 1\right\} \times \left\{\frac{\pi\sqrt{m}}{(2\pi)^{m-1}\sigma}\left(\frac{4\sigma}{\pi} - \cot\frac{\pi}{4\sigma}\right)\right\}^{\frac{1}{p}}, \tag{2.7}$$

$$\|f_i - SN_{\lambda,\sigma}[f_i]\|_p \le \left(1 + \frac{\sqrt{m}\pi^2}{2}\right)\omega\left(f_i, \frac{1}{\lambda+2}\right) + 2\|f_i\|_p \left\{\left(\frac{8(\lambda+2)}{\pi^2}\right)^m - 1\right\} \times \left\{\frac{\pi\sqrt{m}}{(2\pi)^{m-1}\sigma}\left(\ln 2 - \frac{1}{2} + \frac{4\sigma}{\pi} - \cot\frac{\pi}{4\sigma}\right)\right\}^{\frac{1}{p}}. \tag{2.8}$$

From Eqs. (2.1) and (2.4), the approximation error we obtained is more accurate than Eqs. (2.7) and (2.8) in Ref. 13, respectively. For any $\lambda \in N$, if σ is large enough for λ, the functions constructed by networks $PN_{\lambda,\sigma}$ and $SN_{\lambda,\sigma}$ become almost the same, and their approximate errors become almost the same value, which can be estimated mainly by the same formulation based on the second-order modulus of continuity $\omega_2(f,\cdot)$ in terms of λ, i.e. the first terms of the right sides of Eqs. (2.1) and (2.4), while the second terms of the right sides of Eqs. (2.1) and (2.4) are negligible.

Remark 2.3. The assertions (2.2) and (2.5) of Theorems 1 and 2 provide us lower bound estimations on approximation accuracy of the networks $PN_{\lambda,\sigma}$ and $SN_{\lambda,\sigma}$, respectively, the results imply that the average of the networks $PN_{\lambda,\sigma}$ and $SN_{\lambda,\sigma}$ over parameters λ, σ or, equivalently, over different number of neurons, is lower controlled by the second-order modulus of smoothness of f and λ, σ (actually, we can see this through rewriting (2.2) as

$$\omega_2\left(f_i, \frac{1}{\lambda+2}\right) - \left(\frac{\lambda^m}{\sigma^2} + \frac{1}{\lambda^2}\right)\|f_i\|_p \le C\left(\frac{1}{2} + \frac{1}{\lambda}\right)\left\{\frac{2}{\lambda(\lambda+1)}\sum_{k=1}^{\lambda} k\|PN_{\lambda,\sigma}[f_i] - f_i\|_p\right\},$$

and rewriting (2.5) as

$$\omega_2\left(f_i, \frac{1}{\lambda+2}\right) - \left(\frac{\lambda^m}{\sigma} + \frac{\lambda^m}{\sigma^2} + \frac{1}{\lambda^2}\right)\|f_i\|_p \le C\left(\frac{1}{2} + \frac{1}{\lambda}\right)\left\{\frac{2}{\lambda(\lambda+1)}\sum_{k=1}^{\lambda} k\|SN_{\lambda,\sigma} - f_i\|_p\right\}.$$

Hence for any $\lambda \in N$, if σ is large enough for λ, (or, equivalently, the number of hidden neurons) is sufficiently large, the above inequalities and Eqs. (2.2) and (2.5) show that the networks $PN_{\lambda,\sigma}$ and $SN_{\lambda,\sigma}$ can achieve the highest approximation accuracy and the accuracy is found to be $\omega_2\big(f_i, \frac{1}{\lambda+2}\big)$.

Remark 2.4. The assertions (2.3) and (2.6) give essential order estimation of the networks $PN_{\lambda,\sigma}[f]$ and $SN_{\lambda,\sigma}[f]$. Equation (2.3) shows that whenever f belongs to the α-Lipschitz class and $\sigma^2 = O(\lambda^{m+\alpha})$, the essential order of approximation of the networks $PN_{\lambda,\sigma}[f]$ is $\bigcirc(\lambda^{-\alpha})$; Eq. (2.6) shows that whenever f belongs to the α-Lipschitz class and $\sigma = O(\lambda^{m+\alpha})$, the essential order of approximation of the networks $SN_{\lambda,\sigma}[f]$ is $\bigcirc(\lambda^{-\alpha})$. They show also that the higher the smoothness of the function to be approximated, the faster the networks can approximate, and vice versa.

3. Preliminaries

First, we prove an approximation theorem by multidimensional trigonometric polynomials, which is a multidimensional extension of Jackson's theorem[6] of approximations by trigonometric polynomials. This provides an explicit equational representation of an approximating multidimensional trigonometric polynomial and an explicit formulation of the corresponding approximation-error estimation for the order of the polynomial.

Lemma 3.1. *Let $f \in L^p_{2\pi}$, $\delta > 0$, and $k_\lambda \in L^1_{2\pi}(R^m)$ be non-negative and even function for $\lambda \in N$. Then the convolution $k_\lambda * f$ approximates f such that*

$$\|k_\lambda * f - f\|_p \le |\hat{k}_\lambda(0) - 1|\|f\|_p + \frac{1}{2}\omega_2(f,\delta) \times \left\{\left(1 + \pi\delta^{-1}(m\hat{k}_\lambda(0) - \sum_{i=1}^{m}\mathbf{Re}(\hat{k}_\lambda(1_i))^{\frac{1}{2}}\right)^2\right\}.$$

Proof. Since $t \le \pi \sin\frac{t}{2}, 0 \le t \le \pi; \pi\sin\frac{t}{2} \le t, -\pi \le t \le 0$, then $t^2 \le \pi^2 \sin^2\frac{t}{2}$, $-\pi \le t \le \pi$. Then

$$(2\pi)^{-m}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\|t\|^2 k_\lambda(\mathbf{t})d\mathbf{t} \le \pi^2\left(m\hat{k}_\lambda(0) - \sum_{i=1}^{m}\mathbf{Re}(\hat{k}_\lambda(1_i))\right).$$

Since $k_\lambda(t)$ be even function, so

$$k_\lambda * f(\mathbf{x}) = \frac{1}{(2\pi)^m}\int_{-\pi}^{\pi}\frac{1}{2}(f(\mathbf{x}+\mathbf{t}) + f(\mathbf{x}-\mathbf{t}))k_\lambda(\mathbf{t})d\mathbf{t},$$

and by Cauchy–Swartz inequality, we have

$$
\begin{aligned}
\|k_\lambda * f - \hat{k}_\lambda(0) f\|_p &= \left\| \frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} \frac{1}{2}(f(\mathbf{x}+\mathbf{t}) + f(\mathbf{x}-\mathbf{t}) - 2f(\mathbf{x})) K_\lambda(\mathbf{t}) d\mathbf{t} \right\|_p \\
&\leq \frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} \frac{1}{2} K_\lambda(\mathbf{t}) \|\triangle_t^2 f(\mathbf{x})\|_p d\mathbf{t} \\
&\leq \frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} \frac{1}{2} K_\lambda(\mathbf{t}) \omega_2(f, \|\mathbf{t}\|)_p d\mathbf{t} \\
&\leq \frac{1}{(2\pi)^m} \omega_2(f,\delta)_p \int_{-\pi}^{\pi} \frac{1}{2} K_\lambda(\mathbf{t})(1+\delta^{-1}\|\mathbf{t}\|)^2 d\mathbf{t} \\
&\leq \omega_2(f,\delta) \left\{ \frac{1}{2} + \delta^{-2} \frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} \frac{1}{2} K_\lambda(\mathbf{t}) \|\mathbf{t}\|^2 d\mathbf{t} \right. \\
&\quad \left. + \delta^{-1} \left(\frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} K_\lambda(\mathbf{t}) d\mathbf{t} \right)^{\frac{1}{2}} \left(\frac{1}{(2\pi)^m} \int_{-\pi}^{\pi} K_\lambda(\mathbf{t}) \|\mathbf{t}\|^2 d\mathbf{t} \right)^{\frac{1}{2}} \right\} \\
&\leq \omega_2(f,\delta) \left\{ \frac{1}{2} + \frac{1}{2} \delta^{-2} \pi^2 \left(m\hat{k}_\lambda(0) - \sum_{i=1}^m \mathbf{Re}(\hat{k}_\lambda(1_i)) \right) \right. \\
&\quad \left. + \delta^{-1} \left(\pi^2 \left(m\hat{k}_\lambda(0) - \sum_{i=1}^m \mathbf{Re}(\hat{k}_\lambda(1_i)) \right) \right)^{\frac{1}{2}} \right\} \\
&= \frac{1}{2} \omega_2(f,\delta) \left(1 + \pi\delta^{-1} \left(m\hat{k}_\lambda(0) - \sum_{i=1}^m \mathbf{Re}(\hat{k}_\lambda(1_i)) \right)^{\frac{1}{2}} \right)^2 .
\end{aligned} \tag{3.1}
$$

So from (3.1) and triangle inequality,

$$
\begin{aligned}
\|k_\lambda * f - f\|_p &\leq \|\hat{k}_\lambda(0) f - f\|_p + \|k_\lambda * f - \hat{k}_\lambda(0) f\|_p \\
&\leq |\hat{k}_\lambda(0) - 1| \|f\|_p \\
&\quad + \frac{1}{2} \omega_2(f,\delta) \left(1 + \pi\delta^{-1} \left(m\hat{k}_\lambda(0) - \sum_{i=1}^m \mathbf{Re}(\hat{k}_\lambda(1_i)) \right)^{\frac{1}{2}} \right)^2 . \qquad \square
\end{aligned}
$$

We introduce the following multivariate Fejér–Korovkin kernel.[13] Let $\mathbf{r} = (r_1, r_2, \ldots, r_m) \in N_0^m, \lambda \in N$, the Fejér–Korovkin kernel is defined by $K_\lambda(\mathbf{t}) = B_\lambda |\sum_{0 \leq r_i \leq \lambda} b_{\lambda,\mathbf{r}} e^{i\mathbf{r}\mathbf{t}}|^2$, here $b_{\lambda,\mathbf{r}} = \prod_{i=1}^m \frac{r_i+1}{\lambda+2}\pi, B_\lambda = (\sum_{0 \leq r_i \leq \lambda} (b_{\lambda,\mathbf{r}})^2)^{-1}$, from

Ref. 13, we know

$$B_\lambda = \left(\frac{2}{\lambda+2}\right)^m, \quad K_\lambda(\mathbf{t}) = 1 + 2B_\lambda \sum_{\mathbf{p}\neq\mathbf{q}\in N_0^m}^{0\le p_u,q_v\le\lambda} b_{\lambda,\mathbf{p}} b_{\lambda,\mathbf{q}} \cos(\mathbf{p}-\mathbf{q})\mathbf{t},$$
$$\hat{K}_\lambda(\mathbf{t}) = B_\lambda \sum_{\mathbf{p}-\mathbf{q}=\mathbf{r},\mathbf{p},\mathbf{q}\in N_0^m}^{0\le p_u,q_v\le\lambda} b_{\lambda,\mathbf{p}} b_{\lambda,\mathbf{q}}, \quad \hat{K}_\lambda(\mathbf{0}) = 1, \hat{K}_\lambda(\mathbf{1}_i) = \cos\frac{\pi}{\lambda+2}. \tag{3.2}$$

By using of multivariate Fejér–Korovkin kernel, we give a constructive approximation theorem by trigonometric polynomial as follows.

Lemma 3.2. *Let $f \in L^p_{2\pi}$ and K_λ be the m-dimensional Fejér–Korovkin kernel. The convolution $K_\lambda * f$ is an m-dimensional trigonometric polynomial which approximates f such that*

$$(K_\lambda * f)(x) = \langle f, 1\rangle + 2B_\lambda \sum_{\mathbf{p}\neq\mathbf{q}\in N_0^m}^{0\le p_u,q_v\le\lambda} b_{\lambda,\mathbf{p}} b_{\lambda,\mathbf{q}} \{\langle f, \cos(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle \cos(\mathbf{p}-\mathbf{q})\mathbf{x}$$
$$+ \langle f, \sin(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle \sin(\mathbf{p}-\mathbf{q})\mathbf{x}\}. \tag{3.3}$$

The approximation upper bound is estimated by

$$\|K_\lambda * f - f\|_p \le \frac{1}{2}\omega_2\left(f, \frac{1}{\lambda+2}\right)\left(1 + \sqrt{\frac{m}{2}}\pi^2\right)^2, \tag{3.4}$$

and the lower bound is

$$\omega_2\left(f, \frac{1}{\lambda+2}\right) \le \frac{C}{\lambda^2}\left\{\sum_{k=1}^{\lambda} k\|K_k * f - f\|_p + \|f\|_p\right\}. \tag{3.5}$$

Proof. Let m-dimensional Fejér–Korovkin kernel K_λ instead k_λ in Lemma 3.1, and using (3.2), the upper bound is directly inferred. Now we prove the lower bound estimation (3.5). Recalling that the convolution $K_\lambda * f$ of two continuous 2π-periodic functions K_λ and f, obviously,

$$D^\nu K_\lambda * f = (2\pi)^{-d} \int_{[-\pi,\pi]^d} D^\nu f(x-t) K_\lambda(t) dt.$$

For all $1 \le q \le \infty$, since $(2\pi)^{-d} \int_{[-\pi,\pi]^d} K_\lambda(x) dx = 1$ and $\|K_\lambda * f\|_q \le \|f\|_q \|K_\lambda\|_1$, we have

$$\|D^{|\nu|} K_\lambda * f\|_q \le \|D^\nu f\|_q. \tag{3.6}$$

By using the Bernstein inequality, we then obtain

$$\|D^{|\nu|} K_\lambda * f\|_q \le C\lambda^{|\nu|} \|D^\nu f\|_q. \tag{3.7}$$

Let $a_\lambda = \frac{1}{\lambda^2}\|D^{|\nu|}K_\lambda * f\|_q$, $|\nu| = 2$, $b_\lambda = \|K_\lambda * f - f\|_q$. From Eqs. (3.6) and (3.7), we obtain

$$\begin{aligned} a_\lambda &\le \frac{1}{\lambda^2}\|D^{|\nu|}K_\lambda * f(K_k * f)\|_q + \frac{1}{\lambda^2}\|D^{|\nu|}K_\lambda * f(f - K_k * f)\|_q \\ &\le \frac{1}{\lambda^2}\|D^{|\nu|}K_k * f\|_q + C\|f - K_k * f\|_q \\ &= \left(\frac{k}{\lambda}\right)^2 a_k + Cb_k. \end{aligned}$$

Applying Lemma 1 in Ref. 15, thus gives

$$a_\lambda \le C_2\lambda^{-2}\left\{\sum_{k=1}^{\lambda} kb_k + a_1\right\}.$$

So we obtain

$$\sup_{|\nu|=2}\|D^{|\nu|}K_\lambda * f\|_q \le C_2\left\{\sum_{k=1}^{\lambda} k\|K_k * f - f\|_q + \|f\|_q\right\}. \tag{3.8}$$

On the other hand, for $\lambda \ge 2$, there exists $m \in N$ such that $\frac{\lambda}{2} \le m \le \lambda$ and

$$\|f - K_m * f\|_q \le \|f - K_k * f\|_q, \quad \frac{\lambda}{2} \le k \le \lambda. \tag{3.9}$$

Based on Eqs. (3.6)–(3.9), we now can define a K-function as follows:

$$K_2(f, t^2) = \inf_{g \in A.C.loc}\left\{\|f - g\| + t^2 \sup_{|m|=2}\|D^{|m|}g\|\right\}.$$

From Ref. 7, there exists a positive constant C_1 such that

$$C_1^{-1}K_2(f, t^2) \le \omega_2(f, t) \le C_1 K_2(f, t^2). \tag{3.10}$$

Hence,

$$\begin{aligned} K_2\left(f, \frac{1}{(\lambda+2)^2}\right) &\le \|f - K_m * f\|_q + \frac{1}{(\lambda+2)^2}\sup_{|\nu|=2}\|D^{|\nu|}K_\lambda * f\|_q \\ &\le \frac{4}{\lambda^2}\sum_{\frac{\lambda}{2} \le k \le \lambda} k\|f - K_k * f\|_q \\ &\quad + \frac{C_2}{(\lambda+2)^2}\left\{\sum_{k=1}^{\lambda} k\|K_k * f - f\|_q + \|f\|_q\right\} \\ &\le C_3\frac{1}{\lambda^2}\left\{\sum_{k=1}^{\lambda} k\|K_k * f - f\|_q + \|f\|_q\right\}. \end{aligned}$$

Using (3.10), we deduce that

$$\omega_2\left(f, \frac{1}{\lambda+2}\right) \le \frac{C}{\lambda^2}\left\{\sum_{k=1}^{\lambda} k\|f - K_k * f\|_q + \|f\|_q\right\}. \qquad \square$$

Remark 3.1. From Theorem 5 of Ref. 13, we know the author had also obtained the upper bound:

$$\|K_\lambda * f - f\|_p \le \omega\left(f, \frac{1}{\lambda+2}\right)\left(1 + \frac{\sqrt{m}\pi^2}{2}\right). \tag{3.11}$$

Comparing (3.4) and (3.11), we take the second-order modulus of smoothness instead of the modulus of first-order to deduce a more accurate upper bound estimation on approximation of the $K_\lambda * f$, which generalize and sharpen Theorem 5 of Ref. 13. (For example, for the polynomial of nth-order $P(x) = \sum_{i=1}^n a_i x^i$, its modulus of first-order $\omega(P, \delta) = O(\delta)$, and the second-order modulus of smoothness $\omega_2(P, \delta) = \delta^2$. Obviously, the second-order modulus of smoothness is more accurate than the modulus of first-order. In general, the relationship between them can be found by the following expression

$$\omega(f, t) = O(\omega_2(f, \sqrt{t})); \quad \omega_2(f, t) = O(\omega(f, t)).$$

So for characterizing the error of approximation of the function, the second-order modulus of smoothness is more sharpen than the first-order modulus.) With Lemma 3.2, we also develop a lower bound estimation of approximation accuracy of $K_\lambda * f$.

The following we show two constructive approximation theorems to a multivariate trigonometric function by networks with piecewise linear and sigmoidal hidden-layer units, which are more accurate than the corresponds to that result of Ref. 13; In addition, we can obtain the lower bound estimation of the networks. Applying this to the polynomial obtained in Lemma 3.2, we derive our results.

Lemma 3.3. *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*

$$PS_\sigma(\boldsymbol{rx}) = 2(-1)^{|\boldsymbol{r}|} \sin\frac{\pi}{4\sigma} \sum_{k=0}^{4|\boldsymbol{r}|\sigma-1} \cos\frac{(2k+1)\pi}{4\sigma} PL_{\sigma,k}(\boldsymbol{rx}) \tag{3.12}$$

and

$$PC_\sigma(\boldsymbol{rx}) = (-1)^{|\boldsymbol{r}|} - 2(-1)^{|\boldsymbol{r}|} \sin\frac{\pi}{4\sigma} \sum_{k=0}^{4|\boldsymbol{r}|\sigma-1} \sin\frac{(2k+1)\pi}{4\sigma} PL_{\sigma,k}(\boldsymbol{rx}) \tag{3.13}$$

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

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

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

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

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

$$\begin{aligned} &\|\sin(\mathbf{rx}) - PS_\sigma(\mathbf{rx})\|_p \\ &\quad \le \|\sin(\mathbf{rx_0}) - PS_\sigma(\mathbf{rx_0})\|_p \\ &\quad \le \|2(\sin(\mathbf{rx_0}) - PS_\sigma(\mathbf{rx_0})) - (\sin(\mathbf{rx_0} + |\mathbf{r}|h) - PS_\sigma(\mathbf{rx_0} + |\mathbf{r}|h)\|_p \\ &\quad = \|\sin(\mathbf{rx_0} + |\mathbf{r}|h) + \sin(\mathbf{rx_0} - |\mathbf{r}|h) - 2\sin(\mathbf{rx_0})\|_p \\ &\quad \le \omega_2(\sin, |\mathbf{r}|h) \le (|\mathbf{r}|h)^2 \le \frac{1}{4\sigma^2}. \end{aligned}$$

From Eq. (3.13), the networks $PS_\sigma(\mathbf{rx})$ has $4|\mathbf{r}|\sigma$ hidden-layer units. We can construct $PC_\sigma(\mathbf{rx})$ in the same manner and prove the corresponding conclusion. □

Lemma 3.4. *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*

$$SS_\sigma(\boldsymbol{rx}) = 2(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\cos\frac{(2k+1)\pi}{4\sigma}SG_{\sigma,k}(\boldsymbol{rx}) \tag{3.14}$$

and

$$SC_\sigma(\boldsymbol{rx}) = (-1)^{|\boldsymbol{r}|} - 2(-1)^{|\boldsymbol{r}|}\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\boldsymbol{r}|\sigma-1}\sin\frac{(2k+1)\pi}{4\sigma}SG_{\sigma,k}(\boldsymbol{rx}). \tag{3.15}$$

Then we have the following estimation

$$\begin{aligned}\|\sin(\boldsymbol{rx}) - SS_\sigma(\boldsymbol{rx})\|_p &= \|\cos(\boldsymbol{rx}) - SC_\sigma(\boldsymbol{rx})\|_p \\ &\le \frac{|\mathbf{r}|}{2\sigma}\ln\frac{4e^3+4e}{2e^2+e^4+1} + \frac{1}{4\sigma^2}.\end{aligned} \tag{3.16}$$

Proof. We change $PL_{\sigma,k}$ in (3.13) by $SG_{\sigma,k}(\mathbf{rx})$, and obtain the networks denoted by $SS_\sigma(\mathbf{rx})$, then it is a network with $4|\mathbf{r}|\sigma$ sigmoidal hidden-layer units based on $SG_{\sigma,k}(\mathbf{rx})$. If $\mathbf{r} = \mathbf{0}$, the result is obvious; When $\mathbf{r} \neq \mathbf{0}$, for convenience we supposed $r_m \neq 0$, Since

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

$$\leq 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\mathbf{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|\mathbf{r}|}{r_m}|PL_{\sigma,k}(|r|y)-SG_{\sigma,k}(|r|y)|dy\right.$$

$$= 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\mathbf{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|\left|\frac{1}{2\pi r_m}\int_{-|\mathbf{r}|\pi}^{|\mathbf{r}|\pi}\right|PL_{\sigma,k}(t)-SG_{\sigma,k}(t)|dt.$$

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

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

The following we estimate I_1, I_2, I_3, we have

$$\begin{aligned}
I_1&=\frac{1}{2\pi r_m}\int_{(-|\mathbf{r}|+1)\pi}^{\frac{k\pi}{2\sigma}}\frac{1}{1+e^{-8\sigma(t-\pi)/\pi-8\sigma+4k+2}}dt\\
&=-\frac{1}{16\sigma r_m}\int_{(-|\mathbf{r}|+1)(-8\sigma)+4k}^{0}\frac{1}{1+e^{t+2}}dt\\
&\leq\frac{1}{16\sigma r_m}(\ln(1+e^2)-2),\\
I_2&=\frac{1}{2\pi r_m}\int_{\frac{k\pi}{2\sigma}}^{\frac{(k+1)\pi}{2\sigma}}\left|\frac{2\sigma}{\pi}t-k-\frac{1}{1+e^{-8\sigma t/\pi+4k+2}}\right|dt\\
&=\frac{1}{2\sigma r_m}\int_0^1\left|x-\frac{1}{1+e^{-4x+2}}\right|dt\\
&\leq\frac{1}{8\sigma r_m}[2\ln 2-\ln(2+e^{-2}+e^2)+1].
\end{aligned}$$

Similarly, we have $I_3 \leq \frac{1}{16\sigma r_m}\ln(1+e^{-2})$. As $r_m \geq 1$, so we have

$$\begin{aligned}\|SS_\sigma(\mathbf{rx}) - PS_\sigma(\mathbf{rx})\|_1 &\leq 2\sin\frac{\pi}{4\sigma}\sum_{k=0}^{4|\mathbf{r}|\sigma-1}\left|\cos\frac{(2k+1)\pi}{4\sigma}\right|(I_1+I_2+I_3)\\ &\leq 4|\mathbf{r}|\frac{1}{8\sigma r_m}\ln\frac{4e^3+4e}{2e^2+e^4+1} = \frac{|\mathbf{r}|}{2\sigma}\ln\frac{4e^3+4e}{2e^2+e^4+1}. \qquad (3.17)\end{aligned}$$

Using the Lemma 3.3, (3.17) is derived in the case of $p=1$. As for $1<p<\infty$, we can easily verify by use of the result of $p=1$ and the following proposition. □

Proposition 3.1. *For $a_i \geq 0$ $(0 \leq i \leq n)$, $1 \leq p \leq q$, then we have*

$$\left(\sum_{i=0}^{n} a_i^q\right)^{\frac{1}{q}} \leq \left(\sum_{i=0}^{n} a_i^p\right)^{\frac{1}{p}}.$$

Proof. The result is obvious for $a_i \equiv 0$ $(i = 0\cdots n)$. Let $\tau = \{\sum_{i=0}^{n} a_k^p\}^{\frac{1}{p}}$, we have

$$\sum_{i=0}^{n}\left(\frac{a_i}{\tau}\right)^q = \sum_{i=0}^{n}\left(\left(\frac{a_i}{\tau}\right)^p\right)^{\frac{q}{p}} \leq \sum_{i=0}^{n}\left(\frac{a_i}{\tau}\right)^p = 1.$$

Thus $(\sum_{i=0}^{n} a_i^q)^{\frac{1}{q}} \leq (\sum_{i=0}^{n} a_i^p)^{\frac{1}{p}}$. □

4. Proof of the Main Results

Proof (Network construction). We denote, by (3.3) of Lemma 3.2 replacing $\cos(\mathbf{p}-\mathbf{q})\mathbf{x}$ and $\sin(\mathbf{p}-\mathbf{q})\mathbf{x}$ respectively with $PC_\sigma(\mathbf{rx})$ and $PS_\sigma(\mathbf{rx})$ of Lemma 3.3 and obtain the networks $PN_{\lambda,\sigma}[F]$.

Upper bound estimation:

$$\begin{aligned}\sum_{\text{Combinations of }\mathbf{p}\neq\mathbf{q}}^{0\leq p_u,q_v\leq\lambda} B_\lambda b_{\lambda,\mathbf{p}} b_{\lambda,\mathbf{q}} &= \frac{1}{2}B_\lambda\left\{\sum_{0\leq p_u,q_v\leq\lambda} b_{\lambda,\mathbf{p}}b_{\lambda,\mathbf{q}} - \sum_{0\leq p_u\leq\lambda} b_{\lambda,\mathbf{p}}^2\right\}\\ &\leq \frac{1}{2}\left(\left(\frac{8(\lambda+2)}{\pi}\right)^m - \left(\frac{2}{\lambda+2}\right)^m\right). \qquad (4.1)\end{aligned}$$

Hence, from (3.4), Lemma 3.3, $|\langle f_i, \cos(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle| \leq \|f_i\|_p, |\langle f_i, \sin(\mathbf{p}-\mathbf{q})\mathbf{t}\rangle| \leq \|f_i\|_p$, and

$$\|K_\lambda * f_i - PN_{\lambda,\sigma}[f_i]\| \leq \frac{1}{2\sigma^2}\|f_i\|_p\left(\left(\frac{8(\lambda+2)}{\pi}\right)^m - \left(\frac{2}{\lambda+2}\right)^m\right), \qquad (4.2)$$

the upper bound is obtained.

Lower bound estimation: From Eq. (3.5), we have

$$\begin{aligned}
\omega_2\left(f_i, \frac{1}{\lambda+2}\right) &\leq \frac{C}{\lambda^2}\left\{\sum_{k=1}^{\lambda} k\|K_k{*}f_i - f_i\|_p + \|f_i\|_p\right\} \\
&\leq \frac{C}{\lambda^2}\left\{\sum_{k=1}^{\lambda} k\|K_k{*}f_i - PN_{k,\sigma}[f_i]\|_p\right. \\
&\quad \left. + \sum_{k=1}^{\lambda} k\|PN_{k,\sigma}[f_i] - f_i\|_p + \|f_i\|_p\right\} \\
&\leq \frac{C}{\lambda^2}\left\{\frac{\lambda(\lambda+1)}{2}\frac{1}{2\sigma^2}\|f_i\|_p\left(\left(\frac{8(\lambda+2)}{\pi}\right)^m - \left(\frac{2}{\lambda+2}\right)^m\right)\right\} \\
&\quad + \frac{C}{\lambda^2}\left\{\sum_{k=1}^{\lambda} k\|PN_{k,\sigma}[f_i] - f_i\|_p + \|f_i\|_p\right\} \\
&\leq C\left\{\frac{1}{\lambda^2}\sum_{k=1}^{\lambda} k\|PN_{k,\sigma}[f_i] - f_i\|_p + \left(\frac{\lambda^m}{\sigma^2} + \frac{1}{\lambda^2}\right)\|f_i\|_p\right\}. \qquad (4.3)
\end{aligned}$$

Hidden-layer unit number: Each $PN_{\lambda,\sigma}[f_i]$ has hidden units based on the $PL_{\sigma,k}$. Let $\mathbf{r} = \mathbf{p} - \mathbf{q}$, then the number is given by

$$\begin{aligned}
\frac{1}{2}\sum_{\mathbf{r}\neq\mathbf{0}\in N_0^m}^{-\lambda\leq r_i\leq\lambda}\sum_{k=0}^{4|\mathbf{r}|\sigma-1} &= 2\sigma\sum_{\mathbf{r}\in N_0^m}^{-\lambda\leq r_i\leq\lambda}|\mathbf{r}| \\
&= 2\sigma\left\{\lambda(\lambda+1)(2\lambda+1)^{m-1}\right. \\
&\quad \left. +\ (2\lambda+1)\sum_{-\lambda\leq r_2,r_3,\ldots,r_m\leq\lambda}(|r_2|+|r_3|+\cdots+|r_m|)\right\} \\
&= 2m\sigma\lambda(\lambda+1)(2\lambda+1)^{m-1}.
\end{aligned}$$

This finishes the proof of Theorem 2.1. Similarly, we can prove Theorem 2.2. □

5. Conclusion

In this work, approximation estimations of the neural networks have been studied. In terms of second-order modulus of smoothness of a function, an upper bound and lower 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 on the smoothness of the 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.

Acknowledgments

The research in this paper is supported by the Natural Science Foundation of China (Nos. 11001227, 61105041), the Natural Science Foundation Project of CQ CSTC (CSTC, No. 2009BB2306), the Fundamental Research Funds for the Central Universities (No. XDJK2010B005).

References

1. B. Abibullaev, H. D. Seo and M. S. Kim, Epileptic spike detection using continuous wavelet transforms and artificial neural networks, *Int. J. Wavelets, Multiresolut. Inf. Process.* **8** (2010) 33–48.
2. F. L. Cao and Z. B. Xu, Neural network approximation for multivariate periodic functions: Estimates on approximation order, *Chinese J. Comput.* **24** (2001) 903–908 (in Chinese).
3. 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 Networks* **6** (1995) 911–917.
4. C. K. Chui and X. Li, Approximation by ridge functions and neural networks with one hidden layer, *J. Approx. Theory* **70** (1992) 131–141.
5. A. Golbabai and S. Seifollahi, Numerical solution of the second kind integral equations using radial basis function networks, *Appl. Math. Comput.* **174** (2006) 877–883.
6. D. Jackson, On the approximation by trigonometric sums and polynomials, *Trans. Amer. Math. Soc.* **13** (1912) 491–515.
7. H. Johnen and K. Scherer, On the equivalence of the K-functional and modulus of continuity and some applications, in *Constructive Theory of Functions of Several Variable*, eds. W. Schempp and K. Zeller (Springer-Verlag, Berlin, 1977), pp. 119–140.
8. X. Li, On simultaneous approximations by radical basis function neural networks, *Appl. Math. Comput.* **95** (1998) 75–89.
9. V. Maiorov and R. S. Meir, Approximation bounds for smooth functions in $C(\mathcal{R}^d)$ by neural and mixture networks, *IEEE Trans. Neural Networks* **9** (1998) 969–978.
10. H. N. Mhaskar, Approximation properties of a multilayered feed forward artificial neural networks, *Adv. Comput. Math.* **1** (1993) 61–80.
11. H. N. Mhaskar and C. A. Micchelli, Approximation by superposition of a sigmoidal function, *Adv. Appl. Math.* **13** (1992) 350–373.
12. G. Ritter, Efficient estimation of neural weights by polynomial approximation, *IEEE Trans. Inf. Theory* **45** (1999) 1541–1550.
13. S. Suzuki, Constructive function approximation by three-layer artificial neural networks, *Neural Networks* **11** (1998) 1049–1058.
14. Z. B. Xu and F. L. Cao, The essential order of approximation for neural networks, *Sci. China (Ser. F)* **47** (2004) 97–112.
15. Z. B. Xu and J. J. Wang, The essential order of approximation for nearly exponential type neural networks, *Sci. China (Ser. F)* **49** (2006) 446–460.

16. Y. Itô, Approximation of functions on a compact set by finite sums of sigmoid function without scaling, *Neural Networks* **4** (1991) 817–826.
17. M. Zaied, S. Said, O. Jemai and C. B. Amar, A novel approach for face recognition based on fast learning algorithm and wavelet network theory, *Int. J. Wavelets, Multiresolut. Inf. Process.* Doi: No.: 10.1142/S0219691311004389.