APPROXIMATION ORDER TO
A FUNCTION IN L_p SPACE BY
GENERALIZED TRANSLATION NETWORKS

NAHMWOO HAHM[†] AND BUM IL HONG[*]

Abstract. We investigate the approximation order to a function in $L_p[-1, 1]$ for $0 \le p < \infty$ by generalized translation networks. In most papers related to neural network approximation, sigmoidal functions are adapted as an activation function. In our research, we choose an infinitely many times continuously differentiable function as an activation function. Using the integral modulus of continuity and the divided difference formula, we get the approximation order to a function in $L_p[-1, 1]$.

1. Introduction

Recently, approximation of functions from spaces generated by the linear combinations of translates and dilates of one function have been widely studied in many different disciplines such as mathematics, computer science and engineering[1, 4, 7].

A feedforward network with one hidden layer is of the form

$$\sum_{i=1}^{n} c_i \sigma(a_i x + b_i) \tag{1.1}$$

where the weight a_i, the threshold b_i and c_i are real numbers for $1 \le i \le n$ and σ is an activation function. There are some examples of activation functions.

$$\sigma(x) = (1 + e^{-x})^{-1}, \qquad \text{(The squashing function)}$$

$$\sigma(x) = (1 + x^2)^\alpha, \quad \alpha \notin \mathbb{Z}, \qquad \text{(Generalized multiquadrics)}$$

$$\sigma(x) = |x|^{2q-1}, \quad q \in \mathbb{N}, \qquad \text{(Thin plate splines)}$$

and

$$\sigma(x) = \exp(-|x|^2). \qquad \text{(The Gaussian function)}$$

In neural network approximation, we have two main problems. The first is the density problem and the second is the complexity problem. The complexity problem in the theory of neural networks is almost the same as the problem of degree of approximation. This problem has been studied extensively in recent years[2, 3, 5, 6].

In [6], Mhaskar and Hahm introduced the generalized translation networks. A generalized translation network with n neurons is of the form

$$\sum_{i=1}^{n} c_i \psi(a_i \cdot x + b_i) \qquad (1.2)$$

where a_i, b_i and c_i are real numbers for $1 \le i \le n$ and ψ is a real-valued function defined on $\mathbb{R}$. For a fixed natural number n, $\Pi_{n,\psi}$ denotes the set of such functions. In the generalized translation network, the activation function ψ doesn't have a special properties around ∞ and $-\infty$ like a sigmoidal function.

In [2], we obtained the approximation order to a function in $\bar{C}(\mathbb{R})$ by neural networks with a sigmoid activation function. In this paper, we investigate the approximation order to a function in $L_p[-1,1]$ by generalized translation networks. In general, a function in $L_p[-1,1]$ may not be continuous and so the approximation order is represented by the integral modulus of continuity which is defined in Section 2. In order to use the integral modulus of continuity, we assume that the target function is periodic.

2. Preliminaries

We introduce some notations. If f is a Lebesgue measurable function on $[-1, 1]$ and $1 \leq p \leq \infty$, we define the p norms of f as follows.

$$\|f\|_{p,[-1,1]} := \begin{cases} \left\{ \int_{-1}^{1} |f(x)|^p dx \right\}^{\frac{1}{p}} & \text{if } 1 \leq p < \infty \\ \text{ess sup}_{x \in [-1,1]} |f(x)| & \text{if } p = \infty \end{cases}$$

In order to get a main result, we need the following.

Definition 2.1. *If the measurable function $f(x)$ of period $b - a$ belongs to $L_p[a, b]$, then*

$$\omega(f, t)_{p,[a,b]} := \sup_{|h| \leq t} \left\{ \int_{a}^{b} |f(x + h) - f(x)|^p dx \right\}^{\frac{1}{p}}. \qquad (2.1)$$

Note that the integral modulus of continuity satisfies the following properties directly from the definition.

(1) $\omega(f, 0)_{p,[a,b]} = 0$

(2) $\omega(f, t)_{p,[a,b]}$ is a nondecreasing function

(3) $\omega(f, \alpha t)_{p,[a,b]} \leq (\alpha + 1)\omega(f, t)_{p,[a,b]}$ for each non-negative real number α.

Throughout the paper, Π_n denotes the set of all algebraic polynomials of degree at most n and Π_n^* denotes the set of all trigonometric polynomials of degree at most n. For every $f \in L_p[-1, 1]$ and $f^* \in L_p[-\pi, \pi]$, we define

$$E_{n,p}(f) := \inf_{P \in \Pi_n} \|f - P\|_{p,[-1,1]} \quad \text{and} \quad E_{n,p}^*(f^*) := \inf_{P^* \in \Pi_n^*} \|f^* - P^*\|_{p,[-\pi,\pi]}.$$

Note that there is a connection between functions defined on $[-1, 1]$ and functions of period 2π. If $f : [-1, 1] \to \mathbb{R}$, then $f^*(\theta) = f(\cos \theta)$ defines an even periodic function. Conversely, $f(\cos \theta) = f^*(\theta)$ defines a function f on $[-1, 1]$ for an even, 2π-periodic function f^*. From these relations, we can easily obtain

$$E_{n,p}(f) = E_{n,p}^*(f^*). \qquad (2.2)$$

3. Main Results

First of all, we want to show the approximation order to a function in $L_p[-1, 1]$ by algebraic polynomials. For this, we need some well-known results. From the identity

$$\left(\frac{\sin \frac{nt}{2}}{\sin \frac{t}{2}}\right)^2 = n + [(n-1)\cos t + (n-2)\cos 2t + \cdots + \cos(n-1)t] \quad (3.1)$$

we conclude that $\left(\frac{\sin \frac{nt}{2}}{\sin \frac{t}{2}}\right)^4$ is a trigonometric polynomial of degree $2n-2$ and it is also an even function. The computations of following lemmas are in [8,9].

Lemma 3.1.

$$\int_0^{\frac{\pi}{2}} \left(\frac{\sin nt}{\sin t}\right)^4 dt = \frac{n\pi(2n^2 + 1)}{6}. \quad (3.2)$$

Lemma 3.2.

$$\int_0^{\frac{\pi}{2}} t\left(\frac{\sin nt}{\sin t}\right)^4 dt < \frac{n^2\pi^2}{4}. \quad (3.3)$$

In order to investigate the approximation order to a function in $L_p[-1, 1]$ by algebraic polynomials, we use even trigonometric polynomials and the equation (2.2).

Theorem 3.3. *For 2-periodic function* $f \in L_p[-1, 1]$ *and* $m \in \mathbb{N}$, *we have*

$$E_{2m}(f) \le M \cdot \omega(f, \frac{1}{m})_{p,[-1,1]} \quad (3.4)$$

where M is a positive constant depending on f.

Proof. We use the idea stated in Section 2. We set $f^*(x) = f(\cos x)$ for $x \in [-\pi, \pi]$. We define

$$P^*_{2m-2}(x) = \frac{3}{2m\pi(2m^2 + 1)} \int_{-\pi}^{\pi} f^*(x+t) \left(\frac{\sin \frac{mt}{2}}{\sin \frac{t}{2}}\right)^4 dt.$$

Since f^* is even, $P^*_{2n-2}(x)$ is an even trigonometric polynomial of degree at most $2m - 2$ and so $P^*_{2m-2} \in \Pi^*_{2m}$. By some suitable substitutions, we have

$$P^*_{2m-2}(x) = \frac{3}{2m\pi(2m^2+1)} \int_{-\pi}^{\pi} f^*(x+t) \left(\frac{\sin \frac{mt}{2}}{\sin \frac{t}{2}} \right)^4 dt$$

$$= \frac{3}{2m\pi(2m^2+1)} \int_0^{\pi} [f^*(x+t) - f^*(x-t)] \left(\frac{\sin \frac{mt}{2}}{\sin \frac{t}{2}} \right)^4 dt$$

$$= \frac{3}{2m\pi(2m^2+1)} \int_0^{\frac{\pi}{2}} [f^*(x+2t) - f^*(x-2t)] \left(\frac{\sin mt}{\sin t} \right)^4 dt$$

By (3.3) and the generalized Minkowski inequality, we have

$$\|f^* - P^*_{2m-2}\|_{p,[-\pi,\pi]}$$

$$= \left\{ \frac{3}{2m\pi(2m^2+1)} \int_{-\pi}^{\pi} | \int_0^{\frac{\pi}{2}} [f^*(x+2t) - f^*(x-2t) \right.$$

$$\left. - 2f^*(x)] \left(\frac{\sin mt}{\sin t} \right) dt |^p dx \right\}^{\frac{1}{p}}$$

$$\leq \frac{6}{2m\pi(2m^2+1)} \int_0^{\frac{\pi}{2}} \omega(f^*, 2t)_{p,[-\pi,\pi]} \left(\frac{\sin mt}{\sin t} \right)^4 dt$$

$$\leq \frac{6}{2m\pi(2m^2+1)} \omega(f^*, \frac{1}{m})_{p,[-\pi,\pi]} \int_0^{\frac{\pi}{2}} (2m+1+t) \left(\frac{\sin mt}{\sin t} \right)^4 dt$$

$$\leq M^* \cdot \omega(f^*, \frac{1}{m})_{p,[-\pi,\pi]}.$$

Therefore, we have

$$E^*_{2m}(f^*) \leq E^*_{2m-2}(f^*) \leq M^* \cdot \omega(f^*, \frac{1}{m})_{p,[-\pi,\pi]} \qquad (3.5)$$

Since $P^*_{2m-2}(x)$ is an even trigonometric polynomial, it can be written as

$$P^*_{2m-2}(x) = \sum_{k=0}^{2m-2} a_k \cos kx. \qquad (3.6)$$

For each k with $0 \leq k \leq 2m - 2$, let $T_k(x)$ be the Chebyshev polynomial such that $T_k(x) = \cos kx$ for $x \in [-\pi, \pi]$. Now, we set

$$P_{2m-2}(x) := \sum_{k=0}^{2m-2} a_k T_k(x). \qquad (3.7)$$

Then $P_{2m-2}(x) \in \Pi_{2m}$. By (3.5) and (2.2), we have

$$E_{2m}(f) \leq E_{2m-2}(f) \leq \|f - P_{2m-2}\|_{p,[-1,1]} \leq M \cdot \omega(f, \frac{1}{m})_{p,[-1,1]} \quad (3.8)$$

where M is a positive constant depending on f. This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Now we approximate polynomials on $[-1,1]$ by the generalized translation networks. The proof of Theorem 3.4 is the modified proof in [5].

Theorem 3.4. *Let ψ be a real-valued function on $\mathbb{R}$ which is infinitely many times differentiable in some open interval $(b - \delta, b + \delta)$ in $\mathbb{R}$. Assume that*

$$D^{(r)}\psi(b) \neq 0.$$

for $r \in \mathbb{Z}$ and let $T_k(x)$ be the Chebyshev polynomial of degree k. Then for any $\epsilon > 0$, there exists a generalized translation network $G_{k,m} \in \Pi_{6m+1,\psi}$ such that

$$\|T_k - G_{k,m}\|_{p,[-1,1]} \leq \frac{1}{2^p}\epsilon \qquad\qquad (3.9)$$

for $1 \leq p < \infty$.

Proof. Let $\epsilon > 0$. Note that $T_k(x) = \sum_{p=0}^{k} c_p x^p$ for some $c_p \in \mathbb{R}$. From the formula

$$\psi_p(a, x) := \frac{\partial^p}{\partial a^p}\psi(a \cdot x + b) = x^p \cdot \psi^{(p)}(a \cdot x + b),$$

we have

$$x^p = (\psi(b))^{-1} \cdot \psi_p(0, x).$$

For any $h > 0$, the formula

$$\Psi_{p,h}(x) := \frac{1}{h^p} \sum_{i=0}^{p} (-1)^{(p-i)} \binom{p}{i} \psi(h \cdot (2i - p) \cdot x + b)$$

represents a divided difference for $\psi_p(0, x)$. In addition, we have

$$\|\Psi_{p,h} - \psi_p(0, \cdot)\|_{\infty, [-1,1]} \leq M \cdot h$$

where M is a positive constant.

Now, we choose

$$h_m := \min\left\{ \frac{\delta}{3m}, \min_{0 \leq k \leq 2m} \left(\frac{\epsilon}{M \cdot \sum_{p=0}^{k} (\psi^{(p)}(b))^{-1} |c_p|} \right) \right\}$$

Then, the generalized translation network $G_{k,m}$ defined by

$$G_{k,m}(x) := \sum_{p=0}^{k} c_p (\psi^{(p)}(b))^{-1} \cdot \Psi_{p,h_m}(x) \tag{3.10}$$

belongs to $\Pi_{6m+1,\psi}$ since the weights and thresholds in $G_{k,m}$ are selected from the set $\{(h_m j, b) : |j| \leq 3m\}$. Hence

$$\|T_k - G_{k,m}\|_{p,[-1,1]} \leq \frac{1}{2^p} \|T_k - G_{k,m}\|_{\infty, [-1,1]} \leq \frac{1}{2^p} \epsilon.$$

This completes the proof. □

For a fixed m, we can easily obtain from (3.9) that

$$\left\| \sum_{k=0}^{2m-2} a_k T_k - \sum_{k=0}^{2m-2} a_k G_{k,m} \right\|_{p,[-1,1]} \leq \frac{C}{2^p} \epsilon \tag{3.11}$$

where C is a positive constant depending on f and m.

From Theorem 3.3 and Theorem 3.4, we get the main result.

Theorem 3.5. *Let ψ be a real-valued function on $\mathbb{R}$ which is infinitely many times differentiable in some open interval $(b - \delta, b + \delta)$ in $\mathbb{R}$. Assume that*

$$D^{(r)} \psi(b) \neq 0.$$

For any 2-periodic function $f \in L_p[-1, 1]$ and a given natural number n, there exists a generalized translation network $N_n \in \Pi_{n,\psi}$ such that

$$\|f - N_n\|_{p,[-1,1]} < M \cdot \omega(f, \frac{1}{n})_{p,[-1,1]}$$

Where M is a positive constant depending on f.

Proof. Let $n > 13$ and $\epsilon > 0$ be given. Also, let m be the largest natural number such that $12m + 1 \leq n$. Then $n \leq 13m$ and

$$\omega(f, \frac{1}{m})_p = \omega(f, \frac{13}{13m})_p \leq 13\omega(f, \frac{1}{13m})_p \leq 13\omega(f, \frac{1}{n})_p \qquad (3.12)$$

Note that $P_{2m-2}(x) = \sum_{k=0}^{2m-2} a_k T_k(x)$ as in (3.7) is an algebraic polynomials of degree at most $2m - 2$. We define a neural network

$$N_n(x) := \sum_{k=0}^{2m-2} a_k G_{k,2m}(x)$$

where $G_{k,m}(x) = \sum_{p=0}^{k} c_p (\psi^{(p)}(b))^{-1} \cdot \Psi_{p,h_m}(x)$. Then $N_n \in \Pi_{12m+1,\psi} \subset \Pi_{n,\psi}$. (see (3.10)) From Theorem 3.3, (3.11) and (3.12), we have

$$\|f - N_n\|_{p,[-1,1]} \leq \|f - \sum_{k=0}^{2m-2} P_{2m-2}\|_{p,[-1,1]} + \|\sum_{k=0}^{2m-2} P_{2m-2} - N_n\|_{p,[-1,1]}$$

$$\leq M \cdot \omega(f, \frac{1}{n})_{p,[-1,1]} + \frac{C}{2^p}\epsilon.$$

Since $\epsilon > 0$ is arbitrary, we get the result. This completes the proof. $\square$

References

[1] H. Amin, K.M. Curtis and B. R. Hayes-Gill, *Piecewise Linear Approximation Applied to Nonlinear Function of a Neural Network*, IEEE Proc. Circuits Devices Syst. **144** (1997), 313-317.

[2] N. Hahm and B. I. Hong, *Approximation Order to a Function in $\bar{C}(\mathbb{R})$ by Superposition of a Sigmoidal Function*, Applied Mathematics Letters **15** (2002), 591 - 597.

[3] N. Hahm, M. Yang and B. I. Hong, *Degree of Approximation to a Smooth Function by Generalized Translation Networks*, Honam Mathematical Journal **27** (2005), 225-232.

[4] M. Leshno, V. Y. Lin, A. Pinkus and S. Schoken, *Multilayer Feedforward Networks with a Nonpolynomial Activation Function Can Approximate Any Function*, Neural Networks **7** (1995), 861-867.

[5] H. N. Mhaskar, *Neural Networks for Optimal Approximation of Smooth and Analytic Functions*, Neural Computation **8** (1996), 164-177.

[6] H. N. Mhaskar and N. Hahm, *Neural Networks for Functional Approximation and System Identification*, Neural Computation **9(1)** (1997), 143-159.

[7] H. N. Mhaskar and C. A. Micchelli, *Degree of Approximation by Neural and Translation Networks with a Single Hidden Layer*, Advances in Applied Mathematics **16** (1994), 151-183.

[8] I. P. Natanson, *Constructive Function Theory*, Frederick Ungar Publ., 1964.

[9] A. F. Timan, *Theory of Approximation of Functions of a Real Variable*, Dover, 1994.

Nahmwoo Hahm
Department of Mathematics
University of Incheon
Incheon 402-749, Korea
Email : nhahm@incheon.ac.kr

Bum Il Hong
Department of Mathematics and
Institute of Natural Sciences
Kyung Hee University
Yongin, Kyunggi 449-701, Korea
Email : bihong@khu.ac.kr