Bull. Korean Math. Soc. **46** (2009), No. 4, pp. 701–712
DOI 10.4134/BKMS.2009.46.4.701

THE SIMULTANEOUS APPROXIMATION ORDER BY NEURAL NETWORKS WITH A SQUASHING FUNCTION

Nahmwoo Hahm

Abstract. In this paper, we study the simultaneous approximation to functions in $C^m[0,1]$ by neural networks with a squashing function and the complexity related to the simultaneous approximation using a Bernstein polynomial and the modulus of continuity. Our proofs are constructive.

1. Introduction

It is well known that an approximation by neural networks is based on superpositions of a transfer function. A feedforward neural network with a transfer function $\sigma : \mathbb{R} \to \mathbb{R}$ and n neurons in the hidden layer is given by

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

where b_i's are weights and c_i's are thresholds. Many researchers have proved the density results [1, 5, 7] and the complexity results [11, 12, 13] by neural networks. Especially, simultaneous approximations of functions and their derivatives have been studied in recent years by many authors [2, 4, 6, 8, 9] since they have many applications in the fields of science and engineering. Most results in [2, 4, 6, 9] investigated the density results of simultaneous approximations by neural networks. Li and Xu [8] showed the simultaneous approximation order by neural networks with a trigonometric function. The goal of this paper is to obtain the simultaneous approximation order of functions in $C^m[0,1]$ and their derivatives by neural networks with a sigmoidal function. A sigmoidal function is a function $\sigma : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{x\to-\infty} \sigma(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} \sigma(x) = 1.$$

Received September 8, 2008.
2000 *Mathematics Subject Classification.* Primary 41A10, 41A29, 65D15.
Key words and phrases. simultaneous approximation, Bernstein polynomial, squashing function, neural network.

The following functions are some sigmoidal functions.

$$\text{Heaviside function :} \qquad \sigma_h(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

$$\text{Squashing function :} \qquad \sigma_c(x) = \frac{1}{1+ce^{-x}}, \quad c \text{ is a positive constant.}$$

It is well known that continuous functions on a compact subset of $\mathbb{R}$ can be approximated well by neural networks with the Heaviside function. But, the Heaviside function is not differentiable. Therefore, in this paper, we choose a squashing function $\sigma(x) = 1/(1+e^{-x})$ as a transfer function of neural networks. In addition, neural networks with a squashing function are important due to historical reasons [14] as a suitable model for the response characteristics of a biological neuron.

2. Preliminaries

Throughout the paper, m denotes a fixed positive integer, $c, c_1, c_2, c_3, \ldots$ denote positive constants and their values may be different at different occurrences.

In order to get the main result, we use a Bernstein polynomial. By the binomial theorem and the properties of a Bernstein polynomial [10], we have the following result.

Lemma 2.1. *Let $n \in \mathbb{N}$ be given. Then, for any integer r with $0 \leq r \leq n$ and $x \in [0,1]$, the followings hold.*

(a) $\sum_{k=0}^{n-r} \binom{n-r}{k} x^k(1-x)^{n-r-k} = 1.$

(b) $\sum_{k=0}^{n-r} \binom{n-r}{k} kx^k(1-x)^{n-r-k} = (n-r)x.$

(c) $\sum_{k=0}^{n-r} \binom{n-r}{k} k^2x^k(1-x)^{n-r-k} = (n-r)x\big((n-r-1)x+1\big).$

The following lemma can be directly obtained from Lemma 2.1.

Lemma 2.2. *Let $n \in \mathbb{N}$ be given. Then, for any integer r with $0 \leq r \leq n$, we have*

$$\|T_{n,r}\|_{\infty,[0,1]} \leq \frac{n-r}{4},$$

where $T_{n,r}(x) = \sum_{k=0}^{n-r} \big(k-(n-r)x\big)^2 \binom{n-r}{k} x^k(1-x)^{n-r-k}$ for $x \in [0,1]$.

Proof. By Lemma 2.1, we have

$$
\begin{aligned}
(1)\quad T_{n,r}(x) &= \sum_{k=0}^{n-r}(k-(n-r)x)^2\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&= \sum_{k=0}^{n-r}\Big\{k^2-2(n-r)kx+(n-r)^2x^2\Big\}\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&= (n-r)x\Big\{(n-r-1)x+1-(n-r)x\Big\}\\
&= (n-r)(x-x^2).
\end{aligned}
$$

Therefore $|T_{n,r}(x)| \le (n-r)/4$ for any $x \in [0,1]$. □

For $f \in C^m[0,1]$ and $\delta > 0$, we define the modulus of continuity of $f^{(r)}$ for $r \in \mathbb{Z}$ with $0 \le r \le m$ as follows.

$$
\omega(f^{(r)},\delta) := \sup\Big\{|f^{(r)}(x)-f^{(r)}(y)| : |x-y| < \delta\Big\},
$$

where $x, y \in [0,1]$. Note that $\omega(f^{(r)},\delta)$ satisfies the followings.

(a) $\omega(f^{(r)},\delta)$ is a non-decreasing function.

(b) $\omega(f^{(r)},\alpha\delta) \le (\alpha+1)\omega(f^{(r)},\delta)$ for a positive real number α.

(c) $\omega(f^{(r)},\alpha+\beta) \le \omega(f^{(r)},\alpha)+\omega(f^{(r)},\beta)$ for positive real numbers α,β.

For any nonnegative integer n, the difference of a function f with a step h is defined by

$$
\Delta_h^n f(x) := \Delta_h(\Delta_h^{n-1} f(x)),
$$

where $\Delta_h^0 f(x) := f(x)$ and $\Delta_h f(x) := \Delta_h^1 f(x) = f(x+h)-f(x)$.

3. Main results

In this section, first, we obtain the simultaneous approximation order of functions in $C^m[0,1]$ by Bernstein polynomials. Then we show that Bernstein polynomials are simultaneously approximated by neural networks with a squashing function using the properties of a squashing function.

Theorem 3.1. *Let $f \in C^m[0,1]$ and $n \in \mathbb{N}$ with $m < n$. Then, for any integer r with $0 \le r \le m$, we have*

$$
\|f^{(r)} - B_n^{(r)}(f)\|_{\infty,[0,1]} \le c_1\omega\Big(f^{(r)},\frac{1}{\sqrt{n-r}}\Big)+\frac{c_2}{n},
$$

where c_1 and c_2 are positive constants independent of n.

Proof. For $r = 0$, it is proved in [10] that

$$
(2)\qquad \|f - B_n(f)\|_{\infty,[0,1]} \le c\omega\Big(f,\frac{1}{\sqrt{n}}\Big),
$$

where c is a positive constant independent of n. Note that

$$(3)\quad B_n^{'}(f,x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}\left[kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-1-k}\right]$$
$$= n\sum_{k=0}^{n-1}\left[f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right]\binom{n-1}{k}x^k(1-x)^{n-1-k}$$
$$= n\sum_{k=0}^{n-1}\Delta_{1/n} f\left(\frac{k}{n}\right)\binom{n-1}{k}x^k(1-x)^{n-1-k}.$$

By differentiating $B_n^{'}(f,x)$ and simplifying the computation, we get

(4)

$$B_n^{''}(f,x)$$
$$= n(n-1)\sum_{k=0}^{n-2}\left[f\left(\frac{k+2}{n}\right) - 2f\left(\frac{k+1}{n}\right) + f\left(\frac{k}{n}\right)\right]\binom{n-2}{k}x^k(1-x)^{n-2-k}$$
$$= n(n-1)\sum_{k=0}^{n-2}\Delta_{1/n}^2 f\left(\frac{k}{n}\right)\binom{n-2}{k}x^k(1-x)^{n-2-k}.$$

Inductively, we can easily obtain

$$(5)\quad B_n^{(r)}(f,x) = n(n-1)\cdots(n-r+1)\sum_{k=0}^{n-r}\Delta_{1/n}^r f\left(\frac{k}{n}\right)\binom{n-r}{k}x^k(1-x)^{n-r-k}$$

for r with $1 \le r \le m$. According to the properties of the divided difference [3], there exists $\theta_{k,r} \in [k/n, (k+r)/n]$ such that

$$(6)\qquad \Delta_{1/n}^r f\left(\frac{k}{n}\right) = \left(\frac{1}{n}\right)^r f^{(r)}(\theta_{k,r}).$$

Thus, by (6), (5) can be rewritten as

$$(7)\qquad B_n^{(r)}(f,x) = \prod_{j=0}^{r-1}\left(1-\frac{j}{n}\right)\sum_{k=0}^{n-r} f^{(r)}(\theta_{k,r})\binom{n-r}{k}x^k(1-x)^{n-r-k}$$

for r with $1 \le r \le m$. Now we estimate the simultaneous approximation order by a Bernstein polynomial. For $x \in [0,1]$, we have

$$(8)\qquad \begin{aligned} &|f^{(r)}(x) - B_n^{(r)}(f,x)| \\ \le &\prod_{j=0}^{r-1}\left(1-\frac{j}{n}\right)\sum_{k=0}^{n-r}|f^{(r)}(x) - f^{(r)}(\theta_{k,r})|\binom{n-r}{k}x^k(1-x)^{n-r-k} \\ &+ \left[1 - \prod_{j=0}^{r-1}(1-\frac{j}{n})\right]|f^{(r)}(x)|. \end{aligned}$$

We compute the first part of the right side of (8). Note that

$$
\begin{aligned}
&\prod_{j=0}^{r-1}\left(1-\frac{j}{n}\right)\sum_{k=0}^{n-r}|f^{(r)}(x)-f^{(r)}(\theta_{k,r})|\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&\le \sum_{k=0}^{n-r}|f^{(r)}(x)-f^{(r)}(\theta_{k,r})|\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&\le \sum_{k=0}^{n-r}\left|f^{(r)}(x)-f^{(r)}\left(\frac{k}{n-r}\right)\right|\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&\quad+\sum_{k=0}^{n-r}\left|f^{(r)}\left(\frac{k}{n-r}\right)-f^{(r)}(\theta_{k,r})\right|\binom{n-r}{k}x^k(1-x)^{n-r-k}.
\end{aligned}
\tag{9}
$$

Now we compute an upper bound of

$$\sum_{k=0}^{n-r}\left|f^{(r)}(x)-f^{(r)}(\frac{k}{n-r})\right|\binom{n-r}{k}x^k(1-x)^{n-r-k}$$

in (9). Let $\delta > 0$ be given. For $x_0, y_0 \in [0,1]$, we set $\alpha := \alpha(x_0, y_0, \delta)$ an integer $\big[|y_0 - x_0|/\delta\big]$, where $[\cdot]$ is the Gauss function. For $i = 0, 1, 2, \dots, \alpha+1$, we define $q_i = x_0 + \big((y_0 - x_0)/(\alpha+1)\big)i$. Then

$$|f^{(r)}(x_0)-f^{(r)}(y_0)|\le\sum_{i=0}^{\alpha}|f^{(r)}(q_{i+1})-f^{(r)}(q_i)|\le(\alpha+1)\omega(f^{(r)},\delta) \tag{10}$$

for r with $1 \le r \le m$. By Lemma 2.2 and (10), we have

$$
\begin{aligned}
&\sum_{k=0}^{n-r}\left|f^{(r)}(x)-f^{(r)}\left(\frac{k}{n-r}\right)\right|\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&\le\omega(f^{(r)},\delta)\sum_{k=0}^{n-r}\left[\alpha\left(x,\frac{k}{n-r},\delta\right)+1\right]\binom{n-r}{k}x^k(1-x)^{n-r-k}\\
&\le\omega(f^{(r)},\delta)\left[\sum_{k=0}^{n-r}\alpha\left(x,\frac{k}{n-r},\delta\right)\binom{n-r}{k}x^k(1-x)^{n-r-k}+1\right]\\
&\le\omega(f^{(r)},\delta)\left[\frac{1}{\delta^2}\sum_{k=0}^{n-r}\left(\frac{k}{n-r}-x\right)^2\binom{n-r}{k}x^k(1-x)^{n-r-k}+1\right]\\
&\le\omega(f^{(r)},\delta)\left(\frac{1}{4(n-r)\delta^2}+1\right).
\end{aligned}
\tag{11}
$$

If we choose $\delta = 1/\sqrt{n-r}$, then, by (11), we have

$$\sum_{k=0}^{n-r}\left|f^{(r)}(x)-f^{(r)}\left(\frac{k}{n-r}\right)\right|\binom{n-r}{k}x^k(1-x)^{n-r-k}\le\frac{5}{4}\omega\Big(f^{(r)},\frac{1}{\sqrt{n-r}}\Big). \tag{12}$$

Finally, we compute an upper bound of

$$\sum_{k=0}^{n-r} \left| f^{(r)}\left(\frac{k}{n-r}\right) - f^{(r)}(\theta_{k,r}) \right| \binom{n-r}{k} x^k (1-x)^{n-r-k}$$

in (9). Note that $k/n \le k/(n-r) \le (k+r)/n$ for r with $1 \le r \le m < n$ and any integer k with $0 \le k \le n-r$. Since $\theta_{k,r} \in [k/n, (k+r)/n]$, we have

$$\begin{aligned} &\sum_{k=0}^{n-r} \left| f^{(r)}\left(\frac{k}{n-r}\right) - f^{(r)}(\theta_{k,r}) \right| \binom{n-r}{k} x^k (1-x)^{n-r-k} \\ &\le \sum_{k=0}^{n-r} \omega\left(f^{(r)}, \frac{r}{n}\right) \binom{n-r}{k} x^k (1-x)^{n-r-k}. \end{aligned} \tag{13}$$

By the properties of the modulus of continuity,

$$\omega\left(f^{(r)}, \frac{r}{n}\right) \le (r+1)\omega\left(f^{(r)}, \frac{1}{n}\right) \le (r+1)\omega\left(f^{(r)}, \frac{1}{\sqrt{n-r}}\right). \tag{14}$$

From (12), (13) and (14), we have an upper bound of (9) such that

$$\begin{aligned} &\prod_{j=0}^{r-1}\left(1-\frac{j}{n}\right) \sum_{k=0}^{n-r} |f^{(r)}(x) - f^{(r)}(\theta_{k,r})| \binom{n-r}{k} x^k (1-x)^{n-r-k} \\ &\le \frac{5}{4}\omega\left(f^{(r)}, \frac{1}{\sqrt{n-r}}\right) + (r+1)\omega\left(f^{(r)}, \frac{1}{\sqrt{n-r}}\right) \\ &:= c_1 \omega\left(f^{(r)}, \frac{1}{\sqrt{n-r}}\right), \end{aligned} \tag{15}$$

where c_1 is a constant independent of n.

Now we compute the second part of the right side of (8). Since

(16)

$$\begin{aligned} 1 - \prod_{j=0}^{r-1}\left(1-\frac{j}{n}\right) &= \left[1 - \left(1-\frac{1}{n}\right)\right] + \left[\left(1-\frac{1}{n}\right) - \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\right] + \cdots \\ &\quad + \left[\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-2}{n}\right) - \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)\right] \\ &= \sum_{i=1}^{r-1} \frac{i}{n} \prod_{j=0}^{i-1}\left(1-\frac{j}{n}\right) \\ &\le \sum_{i=1}^{r-1} \frac{i}{n} = \frac{r(r-1)}{2n}, \end{aligned}$$

we have

$$\left[1 - \prod_{j=0}^{r-1}\left(1-\frac{j}{n}\right)\right] |f^{(r)}(x)| \le \frac{r(r-1)}{2n} M_r := \frac{c_2}{n}, \tag{17}$$

where $M_r := \|f^{(r)}\|_{\infty,[0,1]}$ for r with $1 \le r \le m$ and c_2 is a constant independent of n. From (2), (15) and (17), we get

$$||f^{(r)} - B_n^{(r)}(f)||_{\infty,[0,1]} \le c_1\omega\Big(f^{(r)}, \frac{1}{\sqrt{n-r}}\Big) + \frac{c_2}{n} \tag{18}$$

for r with $0 \le r \le m$, where c_1 and c_2 are positive constants independent of n. □

If n is sufficiently large in Theorem 3.1, we are able to obtain the improved simultaneous approximation order as follows.

Theorem 3.2. *Let $f \in C^m[0,1]$ and $n \in \mathbb{N}$ with $2m < n$. Then, for any integer r with $0 \le r \le m$, we have*

$$\|f^{(r)} - B_n^{(r)}(f)\|_{\infty,[0,1]} \le c_1\omega\Big(f^{(r)}, \frac{1}{\sqrt{n}}\Big) + \frac{c_2}{n},$$

where c_1 and c_2 are positive constants independent of n.

Proof. Since $2m < n$, we have $n/2 = n - n/2 < n - m \le n - r$ and so

$$\omega\Big(f^{(r)}, \frac{1}{\sqrt{n-r}}\Big) < \omega\Big(f^{(r)}, \frac{\sqrt{2}}{\sqrt{n}}\Big) \le (\sqrt{2}+1)\omega\Big(f^{(r)}, \frac{1}{\sqrt{n}}\Big). \tag{19}$$

By Theorem 3.1 and (19), we complete the proof. □

Note that a squashing function $\sigma(x) = 1/(1+e^{-x})$ is a nonlinear, monotone increasing and differentiable sigmoidal function. In addition, it has a *non-vanishing* point in $[0,1]$ by the following two lemmas.

Lemma 3.3. *Suppose that $R_n(\sigma) := R_n(\sigma, x)$ denotes a polynomial of degree $\le n$ with respect to σ for $n \in \mathbb{N}$. Then*

$$\sigma^{(n)}(x) = \sigma(x)(1-\sigma(x))R_{n-1}(\sigma) \tag{20}$$

for any $n \in \mathbb{N}$.

Proof. We prove it by the mathematical induction.
For $n = 1$, it is clear that $\sigma'(x) = \sigma(x)(1-\sigma(x))$.
Assume that (20) is true for $n = k$. That is, $\sigma^{(k)}(x) = \sigma(x)(1-\sigma(x))R_{k-1}(\sigma)$.
Then, for $n = k+1$, we have

$$\begin{aligned} &\sigma^{(k+1)}(x)\\ &= \sigma'(x)\big[(1-\sigma(x))R_{k-1}(\sigma) - \sigma(x)R_{k-1}(\sigma) + \sigma(x)(1-\sigma(x))R'_{k-1}(\sigma)\big]\\ &= \sigma(x)(1-\sigma(x))R_k(\sigma), \end{aligned} \tag{21}$$

since $(1-\sigma(x))R_{k-1}(\sigma), \sigma(x)R_{k-1}(\sigma)$ and $\sigma(x)(1-\sigma(x))R'_{k-1}(\sigma)$ are polynomials of degree $\le k$ with respect to σ. This completes the proof. □

Lemma 3.4. *Let $\sigma(x) = 1/(1+e^{-x})$. Then there exists $x_0 \in [0,1]$ such that $\sigma^{(n)}(x_0) \ne 0$ for any $n \in \mathbb{N}$.*

Proof. Note that $\sigma(x) \neq 0$ and $1 - \sigma(x) \neq 0$ for all $x \in [0,1]$. Since $\sigma(x)$ is monotone increasing on $\mathbb{R}$, $R_k(\sigma) = 0$ has at most k roots in $[0,1]$. Thus $\sigma^{(k+1)}(x) = 0$ has at most k roots by (21). So $\{x \in [0,1] : \bigcup_{n=1}^{\infty} \sigma^{(n)}(x) = 0\}$ is countable and hence there exists $x_0 \in [0,1]$ such that $x_0 \in [0,1] - \{x \in [0,1] : \bigcup_{n=1}^{\infty} \sigma^{(n)}(x) = 0\}$. □

Similarly, we can easily show that any squashing function $\sigma_c(x) = 1/(1 + ce^{-x})$ with a positive constant c also has a *non-vanishing* point. Using Lemma 3.3, Lemma 3.4 and the divided difference, we now approximate monomials simultaneously by neural networks with a squashing function.

Lemma 3.5. *Let σ be a squashing function and $k \in \mathbb{N} \cup \{0\}$. If $b \in [0,1]$ such that $\sigma^{(j)}(b) \neq 0$ for any $j \in \mathbb{N}$ and $h > 0$, there exists a neural network*

$$N_{k,h}(\sigma, x) = \frac{1}{h^k \sigma^{(k)}(b)} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} \sigma(hjx + b)$$

such that

$$||(x^k)^{(r)} - N_{k,h}^{(r)}||_{\infty,[0,1]} = \mathcal{O}(h)$$

for any integer r with $0 \leq r \leq m$.

Proof. For $r = 0$, $N_{k,h}(\sigma, x)$ represents the divided difference for x^k and so

$$||x^k - N_{k,h}||_{\infty,[0,1]} = \mathcal{O}(h) \tag{22}$$

holds. In order to compute the derivatives of $N_{k,h}(\sigma, x)$, we define

$$[p]_q = \prod_{i=0}^{q-1} (p - i) \tag{23}$$

for $p, q \in \mathbb{N}$. For $p, s \in \mathbb{N}$, we choose $a_{q,s} \in \mathbb{R}$ for $q = 1, 2, \ldots, s$ so that

$$p^s = \sum_{q=1}^{s} a_{q,s} [p]_q. \tag{24}$$

From (24), we get

$$\frac{[j]_q}{[k]_q} \binom{k}{j} = \frac{j(j-1)\cdots(j-(q-1))}{k(k-1)\cdots(k-(q-1))} \frac{k!}{j!(k-j)!} = \binom{k-q}{j-q} \tag{25}$$

for $q, j, k \in \mathbb{N}$ with $q \le j \le k$. Using (24) and (25), we compute the r-th derivative of $N_{k,h}(x)$ for r with $1 \le r \le m$.

$$
\begin{aligned}
(26)\qquad & N_{k,h}^{(r)}(\sigma, x) \\
&= \frac{1}{h^k \sigma^{(k)}(b)} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} \sigma^{(r)}(hjx+b) h^r j^r \\
&= \frac{1}{h^{k-r} \sigma^{(k)}(b)} \sum_{j=1}^{k} \binom{k}{j} (-1)^{k-j} \sum_{q=1}^{r} a_{q,r}[j]_q \sigma^{(r)}(hjx+b) \\
&= \frac{1}{h^{k-r} \sigma^{(k)}(b)} \sum_{q=1}^{r} a_{q,r}[k]_q \sum_{j=q}^{k} \binom{k-q}{j-q} (-1)^{k-j} \sigma^{(r)}(hjx+b) \\
&= \frac{1}{h^{k-r} \sigma^{(k)}(b)} \sum_{q=1}^{r} a_{q,r}[k]_q \sum_{i=0}^{k-q} \binom{k-q}{i} (-1)^{k-i-q} \sigma^{(r)}(hix+hqx+b).
\end{aligned}
$$

By Taylor's theorem for an integer p with $p > k - r$, we have

$$
\begin{aligned}
\sigma^{(r)}(hix+hqx+b) = \sigma^{(r)}(hix+b) + \sum_{l=1}^{p-1} \frac{\sigma^{(r+l)}(hix+b)}{l!}(hqx)^l \\
+ \frac{\sigma^{(r+p)}(\xi_{i,q})}{p!}(hqx)^p,
\end{aligned}
\tag{27}
$$

where $\xi_{i,q}$ is a point between $hix + b$ and $hix + hqx + b$. Note that

$$
\begin{aligned}
&\frac{1}{h^{k-r} \sigma^{(k)}(b)} \sum_{q=1}^{r} a_{q,r}[k]_q \sum_{i=0}^{k-q} \binom{k-q}{i} (-1)^{k-i-q} \sigma^{(r)}(hix+b) \\
&= \sum_{q=1}^{r} a_{q,r}[k]_q h^{r-q} N_{k-q,h}(\sigma^{(r)}, x) \\
&= a_{r,r}[k]_r N_{k-r,h}(\sigma^{(r)}, x) + \sum_{q=1}^{r-1} a_{q,r}[k]_q h^{r-q} N_{k-q,h}(\sigma^{(r)}, x) \\
&= [k]_r N_{k-r,h}(\sigma^{(r)}, x) + \mathcal{O}(h),
\end{aligned}
\tag{28}
$$

since $a_{r,r} = 1$ by comparing the leading coefficients in (24). Since $l + r - q \ge 1$, we have

(29)

$$
\frac{1}{h^{k-r} \sigma^{(k)}(b)} \sum_{q=1}^{r} a_{q,r}[k]_q \sum_{i=0}^{k-q} \binom{k-q}{i} (-1)^{k-i-q} \left[\sum_{l=1}^{p-1} \frac{\sigma^{(r+l)}(hix+b)}{l!} (hqx)^l \right]
$$

$$= \sum_{l=1}^{p-1}\sum_{q=1}^{r} a_{q,r}[k]_q h^{l+r-q}\left[\frac{1}{h^{k-q}\sigma^{(k)}(b)}\sum_{i=0}^{k-q}(-1)^{k-i-q}\sigma^{(r+l)}(hix+b)\right]$$
$$= \sum_{l=1}^{p-1}\sum_{q=1}^{r} a_{q,r}[k]_q h^{l+r-q}\left[x^{k-q}+\mathcal{O}(h)\right]$$
$$= \mathcal{O}(h).$$

Moreover, since $\sigma^{(r+p)}(\xi_{i,q})$ is bounded for $\xi_{i,q} \in [0,1]$ and $p-k+r \geq 1$, we have

(30)
$$\frac{1}{h^{k-r}\sigma^{(k)}(b)}\sum_{q=1}^{r} a_{q,r}[k]_q \sum_{i=0}^{k-q}\binom{k-q}{i}(-1)^{k-i-q}\frac{\sigma^{(r+p)}(\xi_{i,q})}{p!}(hqx)^p = \mathcal{O}(h).$$

From (28), (29) and (30), we have

(31) $$N_{k,h}^{(r)}(\sigma,x) = [k]_r N_{k-r,h}(\sigma^{(r)},x) + \mathcal{O}(h).$$

Therefore

(32) $$||(x^k)^{(r)} - N_{k,h}^{(r)}||_{\infty,[0,1]} = ||[k]_r x^{k-r} - [k]_r N_{k-r,h}(\sigma^{(r)},x)||_{\infty,[0,1]} = \mathcal{O}(h)$$

for r with $1 \leq r \leq m$. By (22) and (32), we complete the proof. □

The next theorem follows from Lemma 3.5 immediately.

Theorem 3.6. *Let $\epsilon > 0$ be given and let σ be a squashing function. If $b \in [0,1]$ such that $\sigma^{(j)}(b) \neq 0$ for any $j \in \mathbb{N}$ and $P_n(x) = \sum_{k=0}^{n} a_k x^k$ for $n \in \mathbb{N}$, there exists a neural network*

$$N_n(\sigma,x) := \sum_{k=0}^{n} a_k N_{k,h}(\sigma,x) = \sum_{k=0}^{n} a_k\left[\frac{1}{h^k\sigma^{(k)}(b)}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}\sigma(hjx+b)\right]$$

such that
$$||P_n^{(r)} - N_n^{(r)}(\sigma)||_{\infty,[0,1]} < \epsilon$$
for sufficiently small $h > 0$ and any integer r with $0 \leq r \leq m$.

By combining Theorem 3.2 and Theorem 3.6, we get the following theorem that is the main result of this paper.

Theorem 3.7. *Let $f \in C^m[0,1]$ and $n \in \mathbb{N}$ with $2m < n$. If σ is a squashing function and $b \in [0,1]$ such that $\sigma^{(j)}(b) \neq 0$ for any $j \in \mathbb{N}$, there exists a neural network*

$$N_n(\sigma,x) = \sum_{k=0}^{n} a_k\left[\frac{1}{h^k\sigma^{(k)}(b)}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}\sigma(hjx+b)\right]$$

such that
$$||f^{(r)} - N_n^{(r)}(\sigma)||_{\infty,[0,1]} \leq c_1\omega\Big(f^{(r)},\frac{1}{\sqrt{n}}\Big) + \frac{c_2}{n},$$

where a_k's are the coefficients of the Bernstein polynomial with resect to f, and constants c_1 and c_2 are independent of n for sufficiently small $h > 0$ and any integer r with $0 \leq r \leq m$.

Proof. By Theorem 3.2, we have

$$\|f^{(r)} - B_n^{(r)}(f)\|_{\infty,[0,1]} \leq c_1\omega\Big(f^{(r)}, \frac{1}{\sqrt{n}}\Big) + \frac{c_2}{n} \tag{33}$$

for r with $0 \leq r \leq m$, where c_1 and c_2 are positive constants independent of n. For a given $\epsilon > 0$, we get, by Theorem 3.6,

$$||B_n^{(r)}(f) - N_n^{(r)}(\sigma)||_{\infty,[0,1]} < \epsilon \tag{34}$$

for sufficiently small $h > 0$ and r with $0 \leq r \leq m$. Therefore

$$\begin{aligned} &\|f^{(r)} - N_n^{(r)}(\sigma)\|_{\infty,[0,1]} \\ &\leq \|f^{(r)} - B_n^{(r)}(f)\|_{\infty,[0,1]} + \|B_n^{(r)}(f) - N_n^{(r)}(\sigma)\|_{\infty,[0,1]} \\ &\leq c_1\omega\Big(f^{(r)}, \frac{1}{\sqrt{n}}\Big) + \frac{c_2}{n} + \epsilon. \end{aligned} \tag{35}$$

Since $\epsilon > 0$ is arbitrarily small, we complete the proof. □

References

[1] R. M. Burton and H. G. Dehling, *Universal approximation in p-mean by neural networks*, Neural Networks **11** (1998), 661–667.

[2] P. Cardaliaguet and G. Euvrard, *Approximation of a function and its derivative with a neural network*, Neural Networks **5** (1992), 207–220.

[3] R. A. DeVore and G. G. Lorentz, *Constructive Approximation*, Springer Verlag, Heidelberg, 1993.

[4] A. R. Gallant and H. White, *On learning the derivatives of an unknown mapping with multilayer feedforward networks*, Lett. Math. Phys. **5** (1992), 129–138.

[5] B. Gao and Y. Xu, *Univariant approximation by superpositions of a sigmoidal function*, J. Math. Anal. Appl. **178** (1993), no. 1, 221–226.

[6] Y. Ito, *Simultaneous L_p-approximations of polynomials and dervatives on the whole space*, Arti. Neural Networks Conf. (1999), 587–592.

[7] G. Lewicki and G. Marino, *Approximation of functions of finite variation by superpositions of a sigmoidal function*, Appl. Math. Lett. **17** (2004), no. 10, 1147–1152.

[8] F. Li and Z. Xu, *The essential order of simultaneous approximation for neural networks*, Appl. Math. Comput. **194** (2007), no. 1, 120–127.

[9] X. Li, *Simultaneous approximation of a multivariate functions and their derivatives by neural networks with one hidden layer*, Neurocomputing **12** (1996), 327–343.

[10] G. G. Lorentz, *Bernstein Polynomials*, Chelsea, Engelwood Cliffs, 1986.

[11] B. Malakooti and Y. Q. Zhou, *Approximation polynomial functions by feedforward artificial neural networks : capacity analysis and design*, Appl. Math. and Comp. **90** (1998), 27–51.

[12] M. V. Medvedeva, *On sigmoidal functions*, Moscow Univ. Math. Bull. **53** (1998), no. 1, 16–19.

[13] H. N. Mhaskar and N. Hahm, *Neural networks for functional approximation and system identification*, Neural Computation **9** (1997), no. 1, 143–159.

[14] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Parallel Distributed Processing : explorations in the microstructure of cognition*, MIT Press, Massachusetts, 1986.

Department of Mathematics
University of Incheon
Incheon 402-749, Korea
E-mail address: nhahm@incheon.ac.kr