Opuscula Mathematica • Vol. 27 • No. 1 • 2007

Michał A. Nowak

APPROXIMATION PROPERTIES OF SOME TWO-LAYER FEEDFORWARD NEURAL NETWORKS

Abstract. In this article, we present a multivariate two-layer feedforward neural networks that approximate continuos functions defined on $[0,1]^d$. We show that the L_1 error of approximation is asymptotically proportional to the modulus of continuity of the underlying function taken at $\sqrt{d}/n$, where n is the number of function values used.

Keywords: neural networks, approximation of functions, sigmoidal function.

Mathematics Subject Classification: Primary 41A35, 41A63; Secondary 41A25, 92B20.

1. INTRODUCTION

Currently analyzed issues related to neural networks have already raised unusually vast interdisciplinary interest. In the scope of neural network construction, one of the classical problems is the ability to recognize patterns and to approximate functions. And as vast and multi-faceted is the neural network issue itself, so multiple are the approaches to the function approximation issue.

In the late 1980s and early 1990s, several important papers discussing neural networks as universal approximators were published ([3, 5–8, 12]). However, these results were of purely extentional nature. In following years, the search for constructive results was launched.

The construction of Kolmogorov networks, for which the initial problem concerned the non-triviality of an activation function (see [13]) deserves a special emphasis here. Another research direction deals with sigma-pi networks (see [10, 11]). Examples of such constructions include the Cardaliaguet-Euvrard operator, utilizing a bell-shaped function as an activation function (see [1, 2, 4]). The majority of research into the constructive approach has been dealing with feedforward networks. (Certain similar results are included in, e.g., [9, 13, 14]).

The uniformity of a neural network seems a natural requirement for the construction thereof. It is a common belief that if a network is to be of any use, it should consist of thousands of interconnected basic elements. The uniformity condition may be understood as the applicability of the same non-decreasing and bounded activation function to all network elements.

To the best of the author's knowledge, there is no construction of approximation functions under the assumptions stated above. In the paper, a construction of a two-layer feedforward neural network is presented. The rate of convergence in the L_1 norm is also estimated.

2. PRELIMINARIES AND A DEFINITION OF A MULTILAYERS FEEDFORWARD UNIFORM NEURAL NETWORKS

We have to begin with the following notations. In $\mathbb{R}^d$, we consider a natural partial order, i.e., for $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$ and $\beta = (\beta_1, \ldots, \beta_d) \in \mathbb{R}^d$, we write $\alpha \leq \beta$ if and only if $\alpha_i \leq \beta_i$ for every $i = 1, \ldots, d$.

Throughout the paper, $[a]$ stands for the integral part of $a \in \mathbb{R}$. For $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$ we set $|\alpha| := \alpha_1 + \ldots + \alpha_d$, $\alpha + 1 := (\alpha_1 + 1, \ldots, \alpha_n + 1)$, $\alpha - 1 := (\alpha_1 - 1, \ldots, \alpha_n - 1)$, $[\alpha] := ([\alpha_1], \ldots, [\alpha_n])$, $J_n := \{0, 1, \ldots, n-1\}^d$ and $J_n^* := \{1, \ldots, n-1\}^d$.

In what follows, we use the symbol ω_f to denote the continuity modulus of f, i.e.

$$\omega_f(\delta) = \sup_{x,y \in [0,1]^d; \|x-y\|_2 \leq \delta} |f(x) - f(y)|,$$

where $\| \cdot \|_2$ is the standard Euclidean norm in $\mathbb{R}^d$.

For two non-negative sequences, we will use the Landau symbol: $a_n = O(b_n)$ to mean that there exists a positive constant C such that $a_n \leq C b_n$ for large n. We write $a_n \sim b_n$ if and only if $a_n = O(b_n)$ and $b_n = O(a_n)$.

Now we recall and introduce definitions of some classes of neural networks.

Definition 2.1. *A non-decreasing function* $\sigma\colon \mathbb{R} \to [0,1]$ *is called* a sigmoidal function, *if* $\lim_{x\to-\infty} \sigma(x) = 0$ *and* $\lim_{x\to+\infty} \sigma(x) = 1$.

Definition 2.2. *Fix* $m, n, l \in \mathbb{N}$. *Let* σ_i *for* $i = 1, \ldots, l$ *be a sigmoidal function and let* a_{ki}, θ_i, b_{kj} *be real numbers for* $i = 1, \ldots, l$, $k = 1, \ldots, n$ *and* $j = 1, \ldots, m$. *Let* $\Phi\colon \mathbb{R}^m \to \mathbb{R}^l \to \mathbb{R}^n$ *be given by*

$$\Phi_k(x) = \sum_{i=1}^{l} a_{ki}\sigma_i\left(\theta_i + \sum_{j=1}^{m} b_{ij}x_j\right), \quad x = (x_1, \ldots, x_m) \in \mathbb{R}^m,$$

where Φ_k *is the* k*-th coordinate of* Φ. *We call any function* Φ *so defined* a one-layer feedforward neural network. *We also say that* m *is* the amount of the input units, l *is* the number of the hidden units *and* n *is* the number of the output units.

Using a matrix notation, Definition 2.2 can be rewritten as follows:

$$\Phi(x) = A\Sigma(\Theta + Bx), \quad x \in \mathbb{R}^m,$$

where $\Sigma = [\sigma_1, \dots, \sigma_l]^T$ is a vector-valued function with sigmoidal coordinates, $\Theta = [\theta_1, \dots, \theta_l]^T$ is a vector of activation coefficients (biases), $A \in M_{n\times l}(\mathbb{R})$ and $B \in M_{l\times m}(\mathbb{R})$.

Definition 2.3. *By a* feedforward two-layer neural network *we mean any superposition of two feedforward one-layer neural networks.*

Likewise, one can define a feedforward n-layer neural network as a superposition of n one-layer neural networks.

Remark 2.4. *We can write any two-layer network in the following way:*

$$\Phi(x) = A\Sigma_2(\Theta_2 + B\Sigma_1(\Theta_1 + Cx)). \tag{1}$$

Indeed, assume that $\Phi_1 \colon \mathbb{R}^{m_1} \to \mathbb{R}^{l_1} \to \mathbb{R}^{n_1}$ *and* $\Phi_2 \colon \mathbb{R}^{m_2} \to \mathbb{R}^{l_2} \to \mathbb{R}^{n_2}$ *are two feedforward one-layer neural networks given by:*

$$\Phi_1(x) = A_1\Sigma_1(\Theta_1 + B_1x), \quad x \in \mathbb{R}^{m_1},$$
$$\Phi_2(x) = A_2\Sigma_2(\Theta_2 + B_2x), \quad x \in \mathbb{R}^{m_2}.$$

Then

$$\Phi_2 \circ \Phi_1(x) = A_2\Sigma_2\big(\Theta_2 + B_2A_1\Sigma_1(\Theta_1 + B_1x)\big).$$

Now, to obtain desired equality (1)*, it suffices to set* $A_2 := A$*,* $B_1 := C$ *and* $B := B_2A_1$*.*

Let us note that $\Phi_2 \circ \Phi_1$ *calls for some comment. This superposition makes sense in the case of* $n_1 = m_2$*, which means that the number of output units of the first network is equal to the number of input units of the second network. We can write*

$$\Phi_2 \circ \Phi_1 \colon \mathbb{R}^{m_1} \to \mathbb{R}^{l_1} \to \mathbb{R}^{l_2} \to \mathbb{R}^{n_2},$$

where m_1 *is the number of input units,* l_1 *is the number of units of first hidden layer,* l_2 *the number of units of second hidden layer and* n_2 *is the number of output units.*

Definition 2.5. *Assume that the same sigmoidal function* σ *appears in every layer* Σ_k $(k = 1, \dots, n)$ *of a feedforward* n*-layer neural network, i.e.,*

$$\Sigma_1 = (\sigma, \dots, \sigma), \Sigma_2 = (\sigma, \dots, \sigma), \dots, \Sigma_n = (\sigma, \dots, \sigma).$$

We say that such network is a uniform neural network.

3. THE MAIN RESULT

Take a sigmoidal function σ. Let us fix natural numbers n, s, r such that $s \geq 1$, $r > d$ and $\alpha \in J_n$. We now define the function $\sigma_\alpha^n \colon [0,1]^d \to \mathbb{R}$ by

$$\sigma_\alpha^n(x) := \sigma\left(n^s\left(\frac{1}{2} - d + \sum_{k=1}^{d} \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big)\right)\right), \quad x = (x_1, \dots, x_d) \in [0,1]^d.$$

Now we are in a position to construct the desired operators. We use the symbol $\mathcal{C}([0,1]^d)$ to denote the set of all real-valued continuous functions defined on $[0,1]^d$.

Definition 3.1. *Fix a sigmoidal function σ. Let us define a sequence of operators $\{B_n\}_{n=1}^{\infty}$ on $\mathcal{C}([0,1]^d)$. For every $n \in \mathbb{N}$, $f \in \mathcal{C}([0,1]^d)$ and $x \in [0,1]^d$ we set*

$$B_n f(x) := \sum_{\alpha \in J_n} \sigma_\alpha^n(x) \sum_{\beta \in J_n; \alpha - 1 \le \beta \le \alpha} (-1)^{|\alpha-\beta|} f\Big(\frac{\beta}{n}\Big). \tag{2}$$

The above sequence is in fact a sequence of correspondent two layers feedforward uniform neural networks with a $1-d$ outputs.

Remark 3.2. *One can easily deduce from* (2) *that*

$$B_n f(x) = \sum_{\beta \in J_n} f\Big(\frac{\beta}{n}\Big) \sum_{\alpha \in J_n; \beta \le \alpha \le \beta+1} (-1)^{|\alpha-\beta|} \sigma_\alpha^n(x).$$

The formula for B_n closely resembles the Bernstein operator for polynomials.

The main result of this paper is the following theorem.

Theorem 3.3. *Let f be a continuous function on $[0,1]^d$, and let σ be a sigmoidal function such that $1 - \sigma(n^s) = O(\frac{1}{n^r})$ and $\sigma(-n^s) = O(\frac{1}{n^r})$. Then there exists a sequence $\xi_n(\sigma) \ge 0$ such that*

$$\|B_n f - f\|_{L^1} \le \omega_f\Big(\frac{\sqrt{d}}{n}\Big) + \|f\|_\infty \xi_n(\sigma)$$

and $\lim_{n\to\infty} \xi_n(\sigma) = 0$. *Moreover,* $\xi_n(\sigma) = O(\frac{1}{n^{r-d}})$.

4. THE PROOF OF THE MAIN RESULT

Before we go into the proof of the main theorem, let us start with some usual lemmas and remarks.

Remark 4.1.

$$\sum_{\alpha \in J_n; \alpha \le nx} \sum_{\beta \in J_n; \alpha-1 \le \beta \le \alpha} (-1)^{|\alpha-\beta|} f\Big(\frac{\beta}{n}\Big) = f\Big(\frac{[nx]}{n}\Big).$$

Proof of Remark.

$$\sum_{\alpha \in J_n; \alpha \le nx} \sum_{\beta \in J_n; \alpha-1 \le \beta \le \alpha} (-1)^{|\alpha-\beta|} f\Big(\frac{\beta}{n}\Big) = \sum_{\beta \in J_n; \beta \le nx} f\Big(\frac{\beta}{n}\Big) \sum_{\alpha \in J_n; \alpha \le nx; \beta \le \alpha \le \beta+1} (-1)^{|\alpha-\beta|} =$$
$$= \sum_{\beta \in J_n; \beta \le [nx]} f\Big(\frac{\beta}{n}\Big) \sum_{\beta+\gamma \in J_n; \beta+\gamma \le [nx]; 0 \le \gamma \le 1} (-1)^{|\gamma|}. \tag{3}$$

Notice that if $\beta < [nx]$ ($\beta \leq [nx]$ but $\beta \neq [nx]$), then there exists $i \in \{1, \dots, d\}$ such that $\beta_i < [nx_i]$ and this implies that $\beta_i + 1 \leq [nx_i]$, too. So the sum

$$\sum_{\beta+\gamma \in J_n; \beta+\gamma \leq [nx]; 0 \leq \gamma \leq 1} (-1)^{|\gamma|}$$

consists of elements of the form

$$(-1)^{|(\gamma_1, \dots, \gamma_{i-1}, 0, \gamma_{i+1}, \dots, \gamma_d)|} + (-1)^{|(\gamma_1, \dots, \gamma_{i-1}, 1, \gamma_{i+1}, \dots, \gamma_d)|} = 0.$$

Therefore, sum (3) is reduced to a single component for $\beta = [nx]$. Then there is only one $\gamma = (0, \dots, 0)$ that satisfies all conditions $\beta + \gamma \in J_n$, $\beta + \gamma \leq [nx]$ and $0 \leq \gamma \leq 1$. Thus (3) is equal to $f(\frac{[nx]}{n})$, which completes the proof of the remark. $\square$

Lemma 4.2. *Let σ be a sigmoidal function such that $1 - \sigma(n^s) = O(\frac{1}{n^r})$. Then there exists a sequence $\xi_n^1(\sigma) \geq 0$ such that*

$$\int_{[0,1]^d} \sum_{\alpha \in J_n; \alpha \leq nx} \big(1 - \sigma_\alpha^n(x)\big) dx \leq \xi_n^1(\sigma)$$

and $\lim_{n\to\infty} \xi_n^1(\sigma) = 0$. Moreover, $\xi_n^1(\sigma) = O(\frac{1}{n^{r-d}})$.

Proof. Let $[\frac{\beta}{n}, \frac{\beta+1}{n}]^d = [\frac{\beta_1}{n}, \frac{\beta_1+1}{n}] \times \dots \times [\frac{\beta_d}{n}, \frac{\beta_d+1}{n}]$. Observe that $[0,1]^d = \bigcup_{\beta \in J_n} [\frac{\beta}{n}, \frac{\beta+1}{n}]^d$. Now

$$\begin{aligned}
&\int_{[0,1]^d} \sum_{\alpha \in J_n; \alpha \leq nx} (1 - \sigma_\alpha^n(x)) dx = \int_{[0,1]^d} \sum_{\substack{\alpha \in J_n; \\ \alpha \leq nx}} \Big(1 - \sigma\Big(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big)\Big)\Big)\Big) dx = \\
&= \sum_{\beta \in J_n} \int_{[\frac{\beta}{n}, \frac{\beta+1}{n}]^d} \sum_{\alpha \in J_n; \alpha \leq [nx] = \beta} \Big(1 - \sigma\Big(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big)\Big)\Big)\Big) dx = \\
&= \sum_{\beta \in J_n} \sum_{\alpha \in J_n; \alpha \leq \beta} \int_{[\frac{\beta}{n}, \frac{\beta+1}{n}]^d} \Big(1 - \sigma\Big(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big)\Big)\Big)\Big) dx = \\
&= \sum_{\beta \in J_n} \sum_{\alpha \in J_n; \alpha \leq \beta} \int_{[\frac{\beta-\alpha}{n}, \frac{\beta-\alpha+1}{n}]^d} \Big(1 - \sigma\Big(n^s\big(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_k)\big)\Big)\Big) dt = \\
&= \sum_{\gamma \in J_n} \prod_{l=1}^d (n - \gamma_l) \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \Big(1 - \sigma\Big(n^s\big(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_k)\big)\Big)\Big) dt = \\
&= \underbrace{\sum_{\gamma \in J_n^*} \prod_{l=1}^d (n - \gamma_l) \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \Big(1 - \sigma\Big(n^s\big(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_k)\big)\Big)\Big) dt}_{S_1} + \\
&\quad + \underbrace{\sum_{\gamma \in J_n \setminus J_n^*} \prod_{l=1}^d (n - \gamma_l) \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \Big(1 - \sigma\Big(n^s\big(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_k)\big)\Big)\Big) dt}_{S_2}.
\end{aligned}$$

Denote

$$T_1 := \max\Big\{0, \sup\Big\{x : \sigma(x) < 1 - \frac{1}{4d}\Big\}\Big\}.$$

Therefore, if $x > T_1$, then $\sigma(x) \geq 1 - \frac{1}{4d}$.

Let $\gamma = (\gamma_1, \ldots, \gamma_d) \in J_n^*$ then $\gamma_j \in \{1, \ldots, n-1\}$ for $j = 1, \ldots, d$. Assume that $t \in [\frac{\gamma}{n}, \frac{\gamma+1}{n}]$; then $n^r t_j \in [\gamma_j n^{r-1}, (\gamma_j + 1)n^{r-1}]$, so $n^r t_k \geq n^{r-1}$ and, because σ is non-decreasing, we get: $\sigma(n^r t_k) \geq \sigma(n^{r-1}) \geq 1 - \frac{1}{4d}$ for $n > T_1^{\frac{1}{r-1}}$. Therefore, if $n > T_1^{\frac{1}{r-1}}$, then we can estimate $\sum_{k=1}^d \sigma(n^r t_j) \geq d(1 - \frac{1}{4d}) = d - \frac{1}{4}$ and we obtain

$$1 - \sigma\left(n^s\left(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_j)\right)\right) \leq 1 - \sigma\left(n^s\left(\frac{1}{2} - d + d - \frac{1}{4}\right)\right) = 1 - \sigma\left(\frac{n^s}{4}\right). \quad (4)$$

Then by (4), for every $n > T_1^{\frac{1}{r-1}}$, there is

$$\begin{aligned} S_1 &= \sum_{\gamma \in J_n^*} \prod_{l=1}^d \Big(n - \gamma_l\Big) \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \left(1 - \sigma\left(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma\big(n^r t_k\big)\Big)\right)\right) dt \leq \\ &\leq (n-1)^d n^d \frac{1}{n^d} \left(1 - \sigma\left(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma\big(n^r\big)\Big)\right)\right) \leq \\ &\leq (n-1)^d \left(1 - \sigma\left(n^s\left(\frac{1}{2} - d + d\Big(1 - \frac{1}{4d}\Big)\right)\right)\right) = (n-1)^d \left(1 - \sigma\Big(\frac{n^s}{4}\Big)\right) =: \xi_n^{11}(\sigma). \end{aligned}$$

Note a simple fact.

We get $\lim_{n\to\infty} \xi_n^{11}(\sigma) = 0$. Moreover, there is

$$n^{r-d} \xi_n^{11}(\sigma) = n^{r-d}(n-1)^d \left(1 - \sigma\left(\frac{n^s}{4}\right)\right) \leq n^r \left(1 - \sigma\left(\frac{n^s}{4}\right)\right).$$

The assumption of Lemma 4.2 implies that
$\lim_{n\to\infty} n^r(1 - \sigma(\frac{n^s}{4})) = 0$, so $\lim_{n\to\infty} n^{r-d}\xi_n^{11}(\sigma) = 0$ and $\xi_n^{11}(\sigma) = O(\frac{1}{n^{r-d}})$.

Next we estimate

$$\begin{aligned} S_2 &= \sum_{\gamma \in J_n \setminus J_n^*} \prod_{l=1}^d (n - \gamma_l) \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \left(1 - \sigma\left(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_k)\Big)\right)\right) dt \leq \\ &\leq \sum_{\gamma \in J_n \setminus J_n^*} \prod_{l=1}^d (n - \gamma_l) \int_{[0, \frac{1}{n}]^d} \left(1 - \sigma\left(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^d \sigma(n^r t_k)\Big)\right)\right) dt. \end{aligned}$$

The last inequality is satisfied, because the function σ is non-decreasing. Notice that the number of components of the sum S_2 is equal to $n^d-(n-1)^d$ and $\prod_{l=1}^d(n-\gamma_l)\le n^d$ for every $\gamma=(\gamma_1,\dots,\gamma_d)\in J_n$. Therefore,

$$\begin{aligned}S_2 &\le (n^d-(n-1)^d)n^d\int_{[0,\frac{1}{n}]^d}\left(1-\sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^d\sigma(n^r t_k)\big)\Big)\right)dt=\\ &=(n^d-(n-1)^d)n^d n^{-dr}\int_{[0,n^{r-1}]^d}\left(1-\sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^d\sigma(y_k)\big)\Big)\right)dy=\\ &=(n^d-(n-1)^d)n^{-d(r-1)}\left(\int_{[0,n^{r-1}]^d\setminus[T_1,n^{r-1}]^d}\left(1-\sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^d\sigma(y_k)\big)\Big)\right)dy+\right.\\ &\qquad\left.+\int_{[T_1,n^{r-1}]^d}\left(1-\sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^d\sigma(y_k)\big)\Big)\right)dy\right).\end{aligned}$$

If $y_k>T_1$, $k=1,\dots,d$, then $\sum_{k=1}^d\sigma(y_k)\ge d-\frac{1}{4}$ because of the definition of T_1. So $\sigma(n^s(\frac{1}{2}-d+\sum_{k=1}^d\sigma(y_k)))\le\sigma(\frac{n^s}{4})$ if $y=(y_1,\dots,y_d)\in[T_1,n^{r-1}]^d$. Because $\sigma\ge 0$, by the above consideration, we can continue the estimation for S_2:

$$\begin{aligned}S_2 &\le (n^d-(n-1)^d)n^{-d(r-1)}\left(\int_{[0,n^{r-1}]^d\setminus[T_1,n^{r-1}]^d}dy+\int_{[T_1,n^{r-1}]^d}\left(1-\sigma\big(\frac{n^s}{4}\big)\right)dy\right)=\\ &=(n^d-(n-1)^d)\left(1-\left(1-\frac{T_1}{n^{r-1}}\right)^d+\left(1-\frac{T_1}{n^{r-1}}\right)^d\left(1-\sigma\big(\frac{n^s}{4}\big)\right)\right)=:\xi_n^{12}(\sigma).\end{aligned}$$

Now we estimate

$$\begin{aligned}n^{r-d}\xi_n^{12}(\sigma) &= n^{r-d}\left(n^d-(n-1)^d\right)\left(1-\left(1-\frac{T_1}{n^{r-1}}\right)^d+\left(1-\frac{T_1}{n^{r-1}}\right)^d\left(1-\sigma\big(\frac{n^s}{4}\big)\right)\right)\sim\\ &\sim n^{r-d}n^{d-1}\left(\frac{T_1}{n^{r-1}}+\left(1-\sigma\Big(\frac{n^s}{4}\Big)\right)\right)=T_1+\frac{1}{n}n^r\left(1-\sigma\Big(\frac{n^s}{4}\Big)\right).\end{aligned}$$

Again, by the assumption of Lemma 4.2, we obtain $n^{r-d}\xi_n^{12}(\sigma)\sim T_1$, which means that $\xi_n^{12}(\sigma)=O(\frac{1}{n^{r-d}})$.

Now we set $\xi_n^1(\sigma):=\xi_n^{11}(\sigma)+\xi_n^{12}(\sigma)$, which completes the proof of Lemma 4.2. □

Lemma 4.3. *Let σ be a sigmoidal function such that $\sigma(-n^s)=O(\frac{1}{n^r})$. Then there exists a sequence $\xi_n^2(\sigma)\ge 0$ such that*

$$\int_{[0,1]^d}\sum_{\alpha\in J_n;\neg(\alpha\le nx)}\sigma_\alpha^n(x)dx\le\xi_n^2(\sigma)$$

and $\lim_{n\to\infty}\xi_n^2(\sigma)=0$. Moreover, $\xi_n^2(\sigma)=O(\frac{1}{n^{r-d}})$.

Proof. There is

$$\int_{[0,1]^d} \sum_{\alpha\in J_n;\neg(\alpha\le nx)} \sigma_\alpha^n(x)dx = \int_{[0,1]^d} \sum_{\alpha\in J_n;\neg(\alpha\le nx)} \sigma\Bigg(n^s\Big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma\Big(n^r\big(x_k-\frac{\alpha_k}{n}\big)\Big)\Big)\Bigg)dx =$$

$$= \sum_{\beta\in J_n} \int_{[\frac{\beta}{n},\frac{\beta+1}{n}]^d} \sum_{\alpha\in J_n;\neg(\alpha\le nx)} \sigma\Bigg(n^s\Big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma\Big(n^r\big(x_k-\frac{\alpha_k}{n}\big)\Big)\Big)\Bigg)dx =$$

$$= \sum_{\beta\in J_n} \sum_{\alpha\in J_n;\neg(\alpha\le \beta)} \int_{[\frac{\beta}{n},\frac{\beta+1}{n}]^d} \sigma\Bigg(n^s\Big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma\Big(n^r\big(x_k-\frac{\alpha_k}{n}\big)\Big)\Big)\Bigg)dx =$$

$$= \sum_{\beta\in J_n} \sum_{\alpha\in J_n;\neg(\alpha\le \beta)} \int_{[\frac{\beta-\alpha}{n},\frac{\beta-\alpha+1}{n}]^d} \sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma(n^r t_k)\big)\Big)dt =$$

$$= \sum_{\beta\in J_n} \sum_{\beta-\gamma\in J_n;\neg(0\le \gamma)} \int_{[\frac{\gamma}{n},\frac{\gamma+1}{n}]^d} \sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma(n^r t_k)\big)\Big)dt = \dots$$

Let $i = 1,\dots,2^d-1$. Then i can be represented in the binary system. $i = (i_d,\dots,i_1)_2$, where $i_j = [\frac{i}{2^{j-1}}](mod 2)$; i.e., we can write $i = \sum_{j=1}^{d} 2^{j-1} i_j$, where $i_j = 0,1$ for every $j = 1,\dots,d$. Now denote $\gamma \in J_n(i)$ if and only if $\gamma_j \in \{0,\dots,n-1\}$ for $i_j = 0$ and $\gamma_j \in \{-(n-1),\dots,-1\}$ for $i_j = 1$ by $j = 1,\dots,d$. Additionally, $\gamma \in J_n^*(i)$ if and only if $\gamma_j \in \{0,\dots,n-1\}$ for $i_j = 0$ and $\gamma_j \in \{-(n-1),\dots,-2\}$ for $i_j = 1$ by $j = 1,\dots,d$. We now may continue the estimating process:

$$\dots = \sum_{i=1}^{2^d-1} \sum_{\beta\in J_n} \sum_{\beta-\gamma\in J_n;\gamma\in J_n(i)} \int_{[\frac{\gamma}{n},\frac{\gamma+1}{n}]^d} \sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma(n^r t_k)\big)\Big)dt \le$$

$$\le \sum_{i=1}^{2^d-1} n^d \sum_{\gamma\in J_n(i)} \int_{[\frac{\gamma}{n},\frac{\gamma+1}{n}]^d} \sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma(n^r t_k)\big)\Big)dt =$$

$$= \sum_{i=1}^{2^d-1} \Bigg(\underbrace{n^d \sum_{\gamma\in J_n^*(i)} \int_{[\frac{\gamma}{n},\frac{\gamma+1}{n}]^d} \sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma(n^r t_k)\big)\Big)dt}_{S_1(i)} + \underbrace{n^d \sum_{\gamma\in J_n(i)\setminus J_n^*(i)} \int_{[\frac{\gamma}{n},\frac{\gamma+1}{n}]^d} \sigma\Big(n^s\big(\frac{1}{2}-d+\sum_{k=1}^{d}\sigma(n^r t_k)\big)\Big)dt}_{S_2(i)} \Bigg).$$

Now for a given $i = 1,\dots,2^d-1$, denote:

$$|i| := i_1 + \dots + i_d,$$

$$T_2(i) := \min\Big\{0, \inf\Big\{x : \sigma(x) > 1 - \frac{3}{4|i|}\Big\}\Big\}.$$

Therefore, if $x < T_2(i)$, then $\sigma(x) \leq 1 - \frac{3}{4|i|}$.

Let $\gamma = (\gamma_1, \ldots, \gamma_d) \in J_n^*(i)$; then $\gamma_j \in \{0, 1, \ldots, n-1\}$ for $i_j = 0$ and $\gamma_j \in \{-n+1, \ldots, -3, -2\}$ for $i_j = 1$. Assume that $t \in [\frac{\gamma}{n}, \frac{\gamma+1}{n}]$ then $n^r t_j \in [\gamma_j n^{r-1}, (\gamma_j+1)n^{r-1}]$. If $\gamma_j \geq 0$, then $n^r t_j \leq (\gamma_j+1)n^{r-1}$ and $\sigma(n^r t_j) \leq \sigma((\gamma_j + 1)n^{r-1}) \leq 1$, because σ is non-decreasing and bounded by 1. Moreover, there is $\#\{j : \gamma_j \geq 0\} = d - |i|$. If $\gamma_j \leq -2$, then $n^r t_j \leq -n^{r-1}$ and $\sigma(n^r t_j) \leq \sigma(-n^{r-1}) \leq 1 - \frac{3}{4|i|}$ for $n > (-T_2(i))^{\frac{1}{r-1}}$. Therefore, we can estimate $\sum_{k=1}^{d} \sigma(n^r t_j) \leq d - |i| + |i|(1 - \frac{3}{4|i|}) = d - \frac{3}{4}$ and we come to

$$\sigma\Big(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^{d} \sigma\big(n^r t_j\big)\Big)\Big) \leq \sigma\Big(n^s\Big(\frac{1}{2} - d + d - \frac{3}{4}\Big)\Big) = \sigma\Big(-\frac{n^s}{4}\Big), \tag{5}$$

when $n > (-T_2(i))^{\frac{1}{r-1}}$. Then by (5), for every $n > (-T_2(i))^{\frac{1}{r-1}}$, there holds

$$\begin{aligned} S_1(i) &= n^d \sum_{\gamma \in J_n^*(i)} \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \sigma\Big(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^{d} \sigma\big(n^r t_k\big)\Big)\Big) dt \leq \\ &\leq n^d \sum_{\gamma \in J_n^*(i)} \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \sigma\left(\frac{-n^s}{4}\right) dt \leq \\ &\leq n^d \sum_{\gamma \in J_n^*(i)} n^{-d} \sigma\left(\frac{-n^s}{4}\right) \leq \\ &\leq n^{d-|i|}(n-1)^{|i|} \sigma\left(\frac{-n^s}{4}\right) =: \xi_n^{21}(i)(\sigma). \end{aligned}$$

Thus $\lim_{x\to\infty} \xi_n^{21}(i)(\sigma) = 0$.

Moreover, there is

$$n^{r-d} \xi_n^{21}(i)(\sigma) = n^{r-d} n^{d-|i|}(n-1)^{|i|} \sigma\left(\frac{-n^s}{4}\right) \leq n^r \sigma\left(\frac{-n^s}{4}\right).$$

By the assumption of Lemma 4.3, we obtain $\lim_{n\to\infty} n^r \sigma\left(\frac{-n^s}{4}\right) = 0$, so $\lim_{n\to\infty} n^{r-d} \xi_n^{21}(i)(\sigma) = 0$ and $\xi_n^{21}(i)(\sigma) = O\left(\frac{1}{n^{r-d}}\right)$.

Next, observe that

$$\begin{aligned} S_2(i) &= n^d \sum_{\gamma \in J_n(i) \setminus J_n^*(i)} \int_{[\frac{\gamma}{n}, \frac{\gamma+1}{n}]^d} \sigma\Big(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^{d} \sigma\big(n^r t_k\big)\Big)\Big) dt \leq \\ &\leq n^d \sum_{\gamma \in J_n(i) \setminus J_n^*(i)} \int_{[-\frac{1}{n}, 0]^d} \sigma\Big(n^s\Big(\frac{1}{2} - d + d - |i| + \sum_{k=1,\ldots,d; i_k=1} \sigma\big(n^r t_k\big)\Big)\Big) dt. \end{aligned}$$

Since the function σ is non-decreasing, the latter inequality holds true. Note that $\#(J_n(i) \setminus J_n^*(i)) = n^d - n^{d-|i|}(n-1)^{|i|}$. Therefore, we may proceed as follows:

$$S_2(i) \le (n^d - n^{d-|i|}(n-1)^{|i|})n^d \int_{[-\frac{1}{n},0]^d} \sigma\Big(n^s\Big(\frac{1}{2} - d + d - |i| + \sum_{k=1,\ldots,d;i_k=1} \sigma(n^r t_k)\Big)\Big) dt =$$
$$= (n^d - n^{d-|i|}(n-1)^{|i|})n^d n^{-dr} \int_{[-n^{r-1},0]^d} \sigma\Big(n^s\Big(\frac{1}{2} - |i| + \sum_{k=1,\ldots,d;i_k=1} \sigma(y_k)\Big)\Big) dy =$$
$$= (n^d - n^{d-|i|}(n-1)^{|i|})n^{-d(r-1)}\cdot$$
$$\cdot \Bigg(\int_{[-n^{r-1},0]^d \setminus [-n^{r-1},T_2(i)]^d} \sigma\Big(n^s\Big(\frac{1}{2} - |i| + \sum_{k=1,\ldots,d;i_k=1} \sigma(y_k)\Big)\Big) dy +$$
$$+ \int_{[-n^{r-1},T_2(i)]^d} \sigma\Big(n^s\Big(\frac{1}{2} - |i| + \sum_{k=1,\ldots,d;i_k=1} \sigma(y_k)\Big)\Big) dy\Bigg) \le$$
$$\le (n^d - n^{d-|i|}(n-1)^{|i|})n^{-d(r-1)}\Bigg(\int_{[-n^{r-1},0]^d \setminus [-n^{r-1},T_2(i)]^d} dy +$$
$$+ \int_{[-n^{r-1},T_2(i)]^d} \sigma\Big(n^s\Big(\frac{1}{2} - |i| + |i|\big(1 - \frac{3}{4|i|}\big)\Big)\Big) dy\Bigg),$$

because $\sigma \le 1$ and $\sigma(y_k) \le 1 - \frac{3}{4|i|}$ if $y_k < T_2(i)$, which is the case in the last integral. We can continue the estimation for $S_2(i)$:

$$S_2(i) \le \Big(n^d - n^{d-|i|}(n-1)^{|i|}\Big)n^{-d(r-1)}\Bigg(n^{d(r-1)} - \big(n^{r-1} + T_2(i)\big)^d +$$
$$+ \big(n^{r-1} + T_2(i)\big)^d \sigma\Big(-\frac{n^s}{4}\Big)\Bigg) =$$
$$= \Big(n^d - n^{d-|i|}(n-1)^{|i|}\Big)\Bigg(1 - \Big(1 + \frac{T_2(i)}{n^{r-1}}\Big)^d + \Big(1 + \frac{T_2(i)}{n^{r-1}}\Big)^d \sigma\Big(-\frac{n^s}{4}\Big)\Bigg) =:$$
$$=: \xi_n^{22}(i)(\sigma).$$

Now we estimate:

$$n^{r-d}\xi_n^{22}(i)(\sigma) = n^{r-d}(n^d - n^{d-|i|}(n-1)^{|i|})\Bigg(1 - \Big(1 + \frac{T_2(i)}{n^{r-1}}\Big)^d + \Big(1 + \frac{T_2(i)}{n^{r-1}}\Big)^d \sigma\Big(-\frac{n^s}{4}\Big)\Bigg) \sim$$
$$\sim n^{r-d}n^{d-1}\Bigg(\frac{-T_2(i)}{n^{r-1}} + \sigma\Big(-\frac{n^s}{4}\Big)\Bigg) = -T_2(i) + \frac{1}{n}n^r\sigma(-\frac{n^s}{4}).$$

Once more, by the assumption of Lemma 4.3, there is $n^{r-d}\xi_n^{22}(i)(\sigma) \sim -T_2(i)$, which means that $\xi_n^{12}(\sigma) = O(\frac{1}{n^{r-d}})$.

Denote $\xi_n^2(\sigma) := \sum_{i=1}^{2^d-1}(\xi_n^{21}(i)(\sigma) + \xi_n^{22}(i)(\sigma))$, which completes the proof of Lemma 4.3. □

Proof of Theorem 3.3. Notice that

$$\begin{aligned}
B_nf(x) &= \sum_{\alpha\in J_n}\sigma_\alpha^n(x)\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big)=\\
&=\sum_{\alpha\in J_n;\alpha\le nx}\sigma_\alpha^n(x)\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big)+\\
&\quad+\sum_{\alpha\in J_n;\neg(\alpha\le nx)}\sigma_\alpha^n(x)\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big)=\\
&=\sum_{\alpha\in J_n;\alpha\le nx}\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big)-\\
&\quad-\sum_{\alpha\in J_n;\alpha\le nx}(1-\sigma_\alpha^n(x))\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big)+\\
&\quad+\sum_{\alpha\in J_n;\neg(\alpha\le nx)}\sigma_\alpha^n(x)\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big).
\end{aligned}$$

Now using Remark 4.1, we can write

$$\begin{aligned}
B_nf(x) &= f(x)-\Big(f(x)-f\Big(\frac{[nx]}{n}\Big)\Big)-\\
&\quad-\sum_{\alpha\in J_n;\alpha\le nx}(1-\sigma_\alpha^n(x))\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big)+\\
&\quad+\sum_{\alpha\in J_n;\neg(\alpha\le nx)}\sigma_\alpha^n(x)\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\Big(\frac{\beta}{n}\Big).
\end{aligned}\tag{6}$$

Next we estimate

$$\begin{aligned}
\|B_nf-f\|_{L^1} &\le \int_{[0,1]^d}\left|f(x)-f\left(\frac{[nx]}{n}\right)\right|dx+\\
&\quad+\int_{[0,1]^d}\sum_{\alpha\in J_n;\alpha\le nx}(1-\sigma_\alpha^n(x))\left|\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\left(\frac{\beta}{n}\right)\right|dx+\\
&\quad+\int_{[0,1]^d}\sum_{\alpha\in J_n;\neg(\alpha\le nx)}\sigma_\alpha^n(x)\left|\sum_{\beta\in J_n;\alpha-1\le\beta\le\alpha}(-1)^{|\alpha-\beta|}f\left(\frac{\beta}{n}\right)\right|dx\\
&\le \omega_f\left(\frac{\sqrt{d}}{n}\right)+2^d\|f\|_\infty\left(\int_{[0,1]^d}\sum_{\alpha\in J_n;\alpha\le nx}(1-\sigma_\alpha^n(x))dx+\int_{[0,1]^d}\sum_{\alpha\in J_n;\neg(\alpha\le nx)}\sigma_\alpha^n(x)dx\right).
\end{aligned}$$

The latter inequality calls for some explanation. Namely, since $\|x-\frac{[nx]}{n}\|_2\le\frac{\sqrt{d}}{n}$ for every $x\in[0,1]^d$, we obtain $\|f(x)-f(\frac{[nx]}{n})\|\le\omega_f(\frac{\sqrt{d}}{n})$. This, when combined with the fact that $\#\{\beta \in J_n : \alpha - 1 \le \beta \le \alpha\} \le 2^d$ for fixed $\alpha \in J_n$, implies that the desired inequality holds true.

In order to finish the proof of Theorem 3.3, it is sufficient to use Lemma 4.2 and Lemma 4.3.

Finally we set $\xi_n(\sigma) := 2^d(\xi_n^1(\sigma)+\xi_n^2(\sigma))$ and the proof of Theorem 3.3 is complete. □

5. EXAMPLE

We deal with a special case of a function σ of "signum"-type:

$$\sigma(x) := \begin{cases} 0, & x < 0, \\ 1, & x \ge 0. \end{cases} \tag{7}$$

Then $T_1 = 0$, $\xi_n^{11} = \xi_n^{12} = 0$ and $T_2(i) = 0$, $\xi_n^{21}(i) = \xi_n^{22}(i) = 0$ for $i = 1, \dots, 2^d - 1$. Consequently, $\xi_n(\sigma) = 0$. This implies that inequality (3.3) has the form:

$$\|B_n f - f\|_{L^1} \le \omega_f\Big(\frac{\sqrt{d}}{n}\Big).$$

Indeed, for σ as in (7), σ_α^n takes values 0 or 1. We proceed as follows

$$\sigma_\alpha^n(x) = 1 \Leftrightarrow \sigma\Bigg(n^s\Big(\frac{1}{2} - d + \sum_{k=1}^{d} \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big)\Big)\Bigg) = 1 \Leftrightarrow$$
$$\Leftrightarrow n^s\Big(\frac{1}{2} - d + \sum_{k=1}^{d} \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big)\Big) \ge 0 \Leftrightarrow \sum_{k=1}^{d} \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big) \ge d - \frac{1}{2} \Leftrightarrow$$
$$\Leftrightarrow \sum_{k=1}^{d} \sigma\Big(n^r\big(x_k - \frac{\alpha_k}{n}\big)\Big) = d \Leftrightarrow x_k \ge \frac{\alpha_k}{n} \qquad \forall k = 1, \dots, n \Leftrightarrow \alpha \le nx.$$

Therefore, we can write

$$\sigma_\alpha^n(x) := \begin{cases} 0, & \neg(\alpha \le nx), \\ 1, & \alpha \le nx. \end{cases}$$

From (6), we get

$$B_n f(x) = f\Big(\frac{[nx]}{n}\Big) - \sum_{\alpha \in J_n; \alpha \le nx} (1 - \sigma_\alpha^n(x)) \sum_{\beta \in J_n; \alpha - 1 \le \beta \le \alpha} (-1)^{|\alpha - \beta|} f\Big(\frac{\beta}{n}\Big) +$$
$$+ \sum_{\alpha \in J_n; \neg(\alpha \le nx)} \sigma_\alpha^n(x) \sum_{\beta \in J_n; \alpha - 1 \le \beta \le \alpha} (-1)^{|\alpha - \beta|} f\Big(\frac{\beta}{n}\Big).$$

Notice that in the case of sigmoidal function given by (7), the second and third summands of the above sum are equal to 0. Therefore, the values of B_n are step functions:

$$B_n f(x) = f\Big(\frac{[nx]}{n}\Big).$$

Indeed, for these functions the rate of convergence is less then or equal to $\omega_f\left(\frac{\sqrt{d}}{n}\right)$.

Acknowledgments
The author would like to extend his gratitude to Maria Malejki, Zygmunt Wronicz and Witold Majdak for their valuable remarks.

REFERENCES

[1] G. Anastassiou, *Quantitative Approximations*, Chapman & Hall/CRC, Boca Raton, 2001.

[2] G. Anastassiou, *Rate of convergence of some neural network operators to the unit-univariate case*, Journal of Mathematical Analysis and Applications **212** (1997), 237–262.

[3] A.R. Barron, *Universal approximation bounds for superpositions of sigmoid function*, IEEE Transactions on Information Theory **3** (1993), 930–945.

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

[5] G. Cybenko, *Approximations by superpositions of a sigmoidal function*, Mathematics of Control, Signals and Systems **2** (1989), 303–314.

[6] K.I. Funahashi, *On the approximate realization of continuos mappings by neural networks*, Neural Networks **2** (1989), 183–192.

[7] K. Hornik, M. Stinchcombe, H. White, *Multilayer feedforward networks are universal approximators*, Neural Networks **2** (1989), 359–366.

[8] K. Hornik, M. Stinchcombe, H. White, *Universal approximation of an unknown mapping and its derivatives*, Neural Networks **3** (1989), 551–560.

[9] M.A. Kon, L. Plaskota, *Information complexity of neural networks*, Neural Networks **13** (2000), 365–375.

[10] B. Lenze, *How to make sigma-pi neural networks perform perfectly on regular training sets*, Neural Networks **7** (1994), 1285–1293.

[11] B. Lenze, *One-sided approximation and interpolation operators generating hyperbolic sigma-pi neural networks*, In: Multivariate Approximation and Splines, G. Nürnberger, J.W. Schmidt, G. Walz (eds.), Birkhäuser Verlag, Basel 1997, 99–112.

[12] Leshno M., Lin V., Pinkus A., Schocken S., *Multilayer feedforward networks with non-polynomial activation functions can approximate any continuos function*, Neural Networks **6** (1993), 861–867.

[13] Scarselli F., Ah Chung Tsoi, *Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results*, Neural Networks **11** (1998), 15–37.

[14] Xi Min Zhanga, Yan Qiu Chenb, Nirwan Ansaria, Yun Q. Shia, *Mini-max initialization for function approximation*, Neurocomputing **57** (2004), 389–409.

Michał A. Nowak
manowak@wms.mat.agh.edu.pl

AGH University of Science and Technology
Faculty of Applied Mathematics
al. Mickiewicza 30, 30-059 Cracow, Poland

Received: May 17, 2004.