Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2012, Article ID 636078, 8 pages
doi:10.1155/2012/636078

Research Article

Estimation of Approximating Rate for Neural Network in L_w^p Spaces

Jian-Jun Wang,[1] Chan-Yun Yang,[2] and Jia Jing[1]

[1] *School of Mathematics & Statistics, Southwest University, Chongqing 400715, China*
[2] *Department of Mechanical Engineering, Taipei Chengshih University of Science and Technology, No.2 Xue-Yuan Rd., Beitou, Taipei 112, Taiwan*

Correspondence should be addressed to Jia Jing, jingj@swu.edu.cn

Received 13 February 2012; Accepted 27 March 2012

Academic Editor: Juan Manuel Peña

A class of Soblove type multivariate function is approximated by feedforward network with one hidden layer of sigmoidal units and a linear output. By adopting a set of orthogonal polynomial basis and under certain assumptions for the governing activation functions of the neural network, the upper bound on the degree of approximation can be obtained for the class of Soblove functions. The results obtained are helpful in understanding the approximation capability and topology construction of the sigmoidal neural networks.

1. Introduction

Artificial neural networks have been extensively applied in various fields of science and engineering. Why is so mainly because the feedforward neural networks (FNNs) have the universal approximation capability [1–13]. A typical example of such universal approximation assertions states that, for any given continuous function defined on a compact set $\mathbf{K}$ of $\mathcal{R}^d$, there exists a three-layer of FNN so that it can approximate the function arbitrarily well. A three-layer of FNN with one hidden layer, d inputs and one output can be mathematically expressed as

$$\mathcal{N}(x) = \sum_{i=1}^{m} c_i \sigma\left(\sum_{j=1}^{d} w_{ij} x_j + \theta_i\right), \quad x \in \mathcal{R}^d, \quad d \geq 1, \tag{1.1}$$

where $1 \leq i \leq m$, $\theta_i \in \mathcal{R}$ are the thresholds, $w_i = (w_{i1}, w_{i2}, \ldots, w_{id})^T \in \mathcal{R}^d$ are connection weights of neuron i in the hidden layer with the input neurons, $c_i \in \mathcal{R}$ are the connection strength of neuron i with the output neuron, and σ is the activation function used in the network. The activation function is normally taken as sigmoid type; that is, it satisfies $\sigma(t) \to 1$ as $t \to +\infty$ and $\sigma(t) \to 0$ as $t \to -\infty$. Equation (1.1) can be further expressed in vector form as

$$\mathcal{N}(x) = \sum_{i=1}^{m} c_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + \theta_i), \quad x \in R^d. \tag{1.2}$$

Universal approximation capabilities for a broad range of neural network topologies have been established by researchers like Cybenko [1], Ito [5], and T. P. Chen and H. Chen [6]. Their work concentrated on the question of denseness. But from the point of application, we are concerned about the degree of approximation by neural networks.

For any approximation problem, the establishment of performance bounds is an inevitable but very difficult issues. As we know, feedforward neural networks (FNNS) have been shown to be capable of approximating general class of functions, including continuous and integrable ones. Recently, several researchers have been derived approximation error bounds for various functional classes (see, e.g., [7–13]) approximated by neural networks. While many open issues remain concerning approximation degree, we stress in this paper on the issue of approximation of functions defined over $[-1, 1]^d$ by FNNS. In [10], the researcher took some basics tools from the theory of weighted polynomial of functions (The weight function is $\omega(x) = \exp(-Q(x))$), under certain assumptions on the smoothness of functions being approximated and on the activation functions in the neural network, the authors present upper bounds on the degree of approximation achieved over the domain R^d.

In this paper, using the Chebyshev Orthogonal series from the approximation theory and moduli of continuity, we obtain upper bounds on the degree of approximation in $[-1, 1]^d$. We take advantage of the properties of the Chebyshev polynomial and the methods of paper [10], we yield the desired results, which can be easily extended to the space R^d.

2. Multivariate Chebyshev Polynomial Approximation

Before introducing the main results, we firstly introduce some basic results on Chebyshev polynomials from the approximation theory. For convenience, we introduce a weighted norm of a function f [14] given by

$$\|f\|_{p,\omega} = \left(\int_{[-1,1]^d} \omega(x) |f(x)|^p dx \right)^{1/p}, \tag{2.1}$$

where $1 \le p < \infty$, $\omega(x) = \prod_{i=1}^{d} \omega(x_i)$ is multivariate weighted function, $\omega(x_i) = (1 - x_i^2)^{-1/2}$, $x = (x_1, x_2, \ldots, x_d) \in R^d$, $m = (m_1, m_2, \ldots, m_d) \in Z^d$, $dx = dx_1 dx_2 \ldots dx_d$. We denote the class of functions for which $\|f\|_{p,\omega}$ is finite by $L_{p,\omega}$.

For function $f : R^d \to R$, the class of functions we wish to approximate in this work is defined as follows:

$$\Psi_{p,\omega}^{r,d} = \left\{ f : \left\| f^{(\lambda)} \right\|_{p,\omega} \le M, |\lambda| \le r \right\}, \tag{2.2}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d)$, $|\lambda| = \lambda_1 + \lambda_2 + \cdots + \lambda_d$, $f^{(\lambda)} = \partial^{|\lambda|} f / \partial_{x_1}{}^{\lambda_1} \ldots \partial_{x_d}{}^{\lambda_d}$, r is a natural number, and $M < \infty$.

2.1. A Chebyshev Polynomial Approximation of Multivariate Functions

As we know, Chebyshev polynomial of a single real variable is a very important polynomial in approximation theory. Using the above notation, we introduce multivariate Chebyshev polynomials: $T_0(x) = 1/\sqrt{\pi}$, $T_n(x) = \prod_{i=1}^{d} T_{n_i}(x_i)$, $T_{n_i}(x_i) = 2/\sqrt{\pi}\cos(n_i \arccos x_i)$. Evidently, for any $m, l \in Z^d$, we have

$$\int_{[-1,1]^d} T_m(x) T_l(x) \omega(x) dx = \begin{cases} 1 & m = l, \\ 0 & m \neq l. \end{cases} \tag{2.3}$$

For $f \in L_{p,\omega}$, $m \in Z^k$, let $\hat{f}(m) = \int_{[-1,1]^d} f(x) T_m(x) \omega(x) dx$, then we have the orthogonal expansion $f(x) \sim \sum_{m=0}^{\infty} \hat{f}(m) T_m(x)$, $x \in [-1,1]^d$.

For one-dimension degree of approximation of a function g by polynomials of degree m, one has the following:

$$E_m(g, P_m, L_{p,\omega}) = \inf_{P \in P_m} \|g - P\|_{p,\omega}, \tag{2.4}$$

where P_m stands for the class of degree-m algebraic polynomials. From [15], we have a simple relationship which we will be used in the following. Let g be differentiable, then we have

$$\begin{gathered} E(g, P_m, L_{p,\omega}) \leq M_1 m^{-1} E(g', P_m, L_{p,\omega}), \\ E(g, P_m, L_{p,\omega}) \leq \|g\|_{p,\omega}. \end{gathered} \tag{2.5}$$

Let $S_n(f, t) = \sum_{k=1}^{n-1} \hat{f}(k) T_k(x)$, and the de la Valle Poussin Operators is defined, that is,

$$V_n(f, t) = \frac{1}{n+1} \sum_{m=n+3/2}^{n+1} S_n(f, t). \tag{2.6}$$

Furthermore, we can simplify $V_n(f, t)$ as follows:

$$V_n(f, t) = \sum_{k=1}^{n} \xi_k \hat{f}(k) T_k(t), \tag{2.7}$$

where

$$\xi_k = \begin{cases} \dfrac{m-1}{2(m+1)}, & \text{if } 0 \leq k \leq \dfrac{m+3}{2}, \\ \dfrac{m-k}{m+1}, & \text{if } \dfrac{m+3}{2} \leq k \leq m. \end{cases} \tag{2.8}$$

A basic result concerning Valle Poussin Operators $V_m(f, t)$ is

$$E_{2m}(f, P_{2m}, L_{p,\omega}) \leq \|f - V_m f\|_{p,\omega} \leq E_m(f, P_m, L_{p,\omega}). \tag{2.9}$$

Now we consider a class of multivariate polynomials defined as follows:

$$P_m = \left\{ P : P(x) = \sum_{0\le|i|\le|m|} b_{i_1,i_2,\ldots,i_d} x_1^{i_1} \cdots x_d^{i_d}, b_{i_1,i_2,\ldots,i_d} \in R, \forall i_1,\ldots,i_d \right\}. \tag{2.10}$$

Hence, we have the following theorem.

Theorem 2.1. *For $1 \le p \le \infty$, let $f \in \Psi_{p,\omega}^{r,d}$. Then for any $m = (m_1, m_2, \ldots, m_d)$, $m_i \le m$, we have*

$$\inf_{P\in P_m} \|f - P\|_{p,\omega} \le Cm^{-r}. \tag{2.11}$$

Proof. We consider the Chebyshev orthogonal polynomials $T_m(x)$, and obtain the following equality from (2.7):

$$V_{i,m_i,}(f) = \sum_{s=1}^{m_i} \xi_s \hat{f}_{s,i} T_s(x_i), \tag{2.12}$$

where $\hat{f}_{s,i} = \int_{[-1,1]^d} f(x) T_s(x_i)\omega(x_i)dx_i$. Hence, we define the following operators:

$$\begin{aligned} V(f) &= V_{1,m_1} V_{2,m_2} \cdots V_{d,m_d} f \\ &= \sum_{s_1=1}^{m_1} \cdots \sum_{s_d=1}^{m_d} \xi_{s_1} \cdots \xi_{s_d} f_{s_1,\ldots,s_d} T_{s_1}(x_1) \cdots T_{s_d}(x_d), \end{aligned} \tag{2.13}$$

where $f_{s_1,\ldots,s_d} = \int_{[-1,1]^d} (\prod_{i=1}^d \omega(x_i) T_{s_i}(x_i)) f(x) dx$. Then we have

$$\begin{aligned} \|f - V(f)\|_{p,\omega} &= \| f - V_{1,m_1}(f) + V_{1,m_1}(f) - V_{1,m_1}V_{2,m_2}(f) \\ &\quad + V_{1,m_1}V_{2,m_2}(f) - \cdots - V(f) \|_{p,\omega} \\ &\le \sum_{i=1}^{d} \| V_0 \cdots V_{i-1,m_i-1} f - V_0 \cdots V_{i,m_i} f \|_{p,\omega}, \end{aligned} \tag{2.14}$$

where V_0 is the identity operator. Let $g = V_0 \cdots V_{i-1,m_i-1} f$, then $V_{i,m_i} g = V_0 \cdots V_{i,m_i} f$, $g^{r_i}(x) = V_0 \cdots V_{i-1,m_i-1} D^{r_i} f(x)$. We view $V_{i,m_i} g$ as a one-dimensional function x_i. Using (2.4), (2.5), and (2.6), we have

$$\begin{aligned} \|g - V_{i,m_i} g\|_{p,\omega} &\le C_1 E_{m_i}(g, P_{m_i}, L_{p,\omega}) \\ &\le C_1 M_1^{r_i} \left(\frac{1}{m_i}\right) \cdots \left(\frac{1}{m_i - r_i + 1}\right) E_{m_i - r_i}(g^{r_i}, P_{m_i - r_i}, L_{p,\omega}) \\ &\le C_1 M_1^{r_i} \left(\frac{1}{m_i}\right) \cdots \left(\frac{1}{m_i - r_i + 1}\right) \|g^{r_i}\|_{p,\omega} \\ &= C_{r_i} \left(\frac{1}{m_i}\right) \cdots \left(\frac{1}{m_i - r_i + 1}\right) \|V_0 \cdots V_{i-1,m_i-1} D^{r_i} f\|_{p,\omega}. \end{aligned} \tag{2.15}$$

Letting $r_i = r$, $m_i = m$, $i = 1, \ldots, d$, if $m > r(r-1)$, we get from (2.15), (2.13), (2.14) and the inequality $\prod_{i=1}^{n}(1+a_i) \geq 1 + \sum_{i=1}^{n} a_i$, $(a_i \geq -1)$,

$$\begin{aligned}
\|f - V(f)\|_{p,\omega} &\leq C_r \sum_{i=1}^{d} \left(\frac{1}{m}\right) \cdots \left(\frac{1}{m-r+1}\right) \|D^r f\|_{p,\omega} \\
&\leq C_r d M \left(\frac{1}{m}\right) \cdots \left(\frac{1}{m-r+1}\right) \\
&= C_r d M m^{-r} \left(1 - \frac{1}{m}\right)^{-1} \left(1 - \frac{2}{m}\right)^{-1} \cdots \left(1 - \frac{r-1}{m}\right) \\
&\leq C_r d M m^{-r} \left(1 - \frac{r(r-1)}{2m}\right)^{-1} \leq 2 d C_r M m^{-r}.
\end{aligned} \tag{2.16}$$

In order to obtain a bound valid for all m, for $m \leq r(r-1)$, we always have the trivial bound $\|f - V(f)\|_{p,\omega} \leq M_2$ since $\|f\|_{p,\omega} \leq M_2$. Letting $C = \max\{2dC_r dM, 2^{-1} M_2 (r(r-1))^r\}$, we conclude an inequality of the desired type for every m. □

This theorem reveals two things: (i) for any multivariate functions $f \in \Psi_{p,\omega}^{r,d}$, there is a polynomial $P \in P_m$ that approximates f arbitrarily well in L_{ω}^{p}, (ii) quantitatively, the approximation accuracy of a polynomial $P \in P_m$ can attain the order of $\bigcirc(m^{-r})$, where m is the dimension of multivariate polynomial, and r is the smoothness of the function to be approximated.

3. Approximation by Feedforward Neural Networks

We consider the approximation of functions by feedforward neural networks with a ridge functions. We define the approximating function class composed of a single hidden layer feedforward neural network with n hidden units. The class of function is

$$F_n = \left\{ f : f(x) = \sum_{k=1}^{n} d_k \phi(a_k \cdot x + b_k);\, a_k \in R^d, b_k, d_k \in R, k = 1, 2, \ldots, n \right\}, \tag{3.1}$$

where $\phi(x)$ satisfy the following assumptions.

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

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

We define the distance from F to G as

$$\operatorname{dist}(F, G, L_{p,\omega}) = \sup_{f \in F} \inf_{g \in G} \|f - g\|_{p,\omega}, \tag{3.2}$$

where F, G are two sets in L_{ω}^{p}. We have the following results.

Theorem 3.1. *Let condition (1) and (2) hold for the activation function $\phi(x)$. Then for every $0 < L < \infty$, $m = (m_1, m_2, \dots, m_d) \in Z_+^d$, $m_i \le m$, $\epsilon > 0$ and $n > (m+1)^d$, we have*

$$\operatorname{dist}\big(BP_m(L), F_n, L_{p,\omega}\big) \le \epsilon, \tag{3.3}$$

where

$$BP_m(L) = \left\{P : p(x) = \sum_{0\le s\le m} a_s x^s; \max_{0\le s\le m} |a_s| \le L\right\}. \tag{3.4}$$

Proof. Firstly, we consider the partial derivative

$$\phi^{(s)}(w \cdot x + b) = \frac{\partial^{(|s|)}}{\partial_{w_1}^{s_1} \cdots \partial_{w_d}^{s_d}} \big(\phi(w \cdot x + b)\big) = \mathbf{x}^{\mathbf{s}} \phi^{|s|}(w \cdot x + b), \tag{3.5}$$

where $|s| = s_1 + \cdots + s_d$, and $\mathbf{x}^{\mathbf{s}} = \prod_{i=1}^{d} x_i^{s_i}$. Thus $\phi^{(s)}(b) = \mathbf{x}^{\mathbf{s}} \phi^{|s|}(b)$.

For any fixed b and $|x| < \infty$ (here $|x| = \sum_{i=1}^{d} x_i$), we consider a finite difference of orders

$$\begin{aligned}
\triangle_{h,x}^{s} \phi(b) &= \sum_{0\le l\le s} (-1)^{|l|} C_s^l \phi(hl \cdot \mathbf{x} + b) \\
&= \mathbf{x}^s \int_0^h \cdots \int_0^h \phi^{(|s|)} [((a_1 + \cdots + a_{s_1}) x_1 + \cdots \\
&\quad + \big(a_{|s|-s_d+1} + \cdots + a_{|s|}\big) x_d \big) + b \big] da_1 \cdots da_{|s|} \\
&\doteq \mathbf{x}^s A_h^{|s|} \phi(\mathbf{x}),
\end{aligned} \tag{3.6}$$

where $C_s^l = \prod_{i=1}^{d} C_{s_i}^{l_i}$, $\triangle_{h,x}^{s} \phi(b) \in F_n$ with $n = \prod_{i=1}^{d}(1 + s_i)$, So

$$\begin{aligned}
\Big|\phi^s(b) - h^{-|s|} \triangle_{h,x}^{s} \phi(b)\Big| &= \Big|\mathbf{x}^s \Big(\phi^{|s|}(b) - h^{-|s|} A_h^s \phi^{|s|}(\mathbf{x})\Big)\Big| \\
&= \Big|\mathbf{x}^s \Big(\phi^{|s|}(b) - \phi^{|s|}(b + \eta\ )\Big)\Big| \\
&\le C_s \omega\Big(\phi^{|s|}, h\Big),
\end{aligned} \tag{3.7}$$

where we derive (3.7) by using (3.6), the mean value theorem of integral, (i.e., there is a $\eta \in [0, h|s \cdot \mathbf{x}|]$, such that $A_h^s \phi^{|s|}(\mathbf{x}) = h^{|s|} \phi^{|s|}(\mathbf{b} + \eta)$) and the moduli of continuity $\omega(g, h) = \sup_{|t|\le h} |f(x+t) - f(x)|$.

From the definition of $\operatorname{dist}(F, G, L_{p,\omega})$ and (3.7), we have

$$\begin{aligned}
\operatorname{dist}\left(BP_m(L), F_n, L_{p,\omega}\right)^p &\leq \left\| \sum_{0\leq s\leq m} a_s x^s - \sum_{0\leq s\leq m} a_s \frac{\triangle_{h,x}^s \phi(b)}{h^s \phi^{|s|}(b)} \right\|_{p,\omega}^p \\
&\leq (m+1)^d \max_{0\leq s\leq m} \left\{ |a_s| \left\| x^s - \frac{\triangle_{h,x}^s \phi(b)}{h^s \phi^{|s|}(b)} \right\|_{p,\omega}^p \right\} \\
&\leq (m+1)^d L \max_{0\leq s\leq m} \left\{ \phi^{|s|}(b) \right\}^{-p} \omega\left(\phi^{|s|}, h\right) \\
&\leq (m+1)^d L C_\phi^p \omega\left(\phi^{|s|}, h\right) < \epsilon.
\end{aligned} \tag{3.8}$$

The last step $\omega(\phi^{|s|}, h)$ can be made arbitrarily small by letting $h \to 0$.

Using the Theorems 2.1 and 3.1, we can easily establish our final result. □

Theorem 3.2. *For $1 \leq p \leq \infty$, we have*

$$\operatorname{dist}\left(\Psi_{p,\omega}^{r,d}, F_n, L_{p,\omega}\right) \leq C n^{-r/d}. \tag{3.9}$$

This theorem reveals two things: (i) for any multivariate functions $f \in \Psi_{p,\omega}^{r,d}$, there is a single hidden layer feedforward neural network $N \in F_n$ with n hidden units that approximates f arbitrarily well in L_ω^p. That is, the feedforward neural networks can be used as the universal approximator of functions in $\Psi_{p,\omega}^{r,d}$; (ii) quantitatively, the approximation accuracy of a mixture network of the form (3.1) can attain the order of $\bigcirc(n^{-r/d})$, where d is the dimension of input space, and r is the smoothness of the function to be approximated.

4. Conclusion

In this work, the approximation order of feedforward neural networks with the form (3.1) has been studied. In terms of smoothness of a function, an upper bound estimation on approximation precision and speed of the neural networks is 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. The results obtained are helpful in understanding the approximation capability and topology construction of the sigmoidal neural networks.

Acknowledgments

This research is supported by Natural Science Foundation of China (no. 11001227), Natural Science Foundation Project of CQ CSTC (no. CSTC,2009BB2306), and the Fundamental Research Funds for the Central Universities (no. XDJK2010B005).

References

[1] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.

[2] J. J. Wang, Z. B. Xu, and W. J. Xu, *Approximation Bounds by Neural Networks in L(w, p)*, vol. 3173 of *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2004.

[3] J. J. Wang, B. Chen, and C. Yang, “Approximation of algebraic and trigonometric polynomials by feedforward neural networks,” *Neural Computing and Applications*, vol. 21, no. 1, pp. 73–80, 2011.

[4] J. J. Wang and Z. B. Xu, “New study of neural networks: the essential order of approximation,” *Neural Networks*, vol. 23, pp. 618–624, 2010.

[5] Y. Ito, “Approximation of continuous functions on R^d by linear combination of shifted rotations of sigmoid function with and without scaling,” *Neural Networks*, vol. 5, no. 1, pp. 105–115, 1992.

[6] T. P. Chen and H. Chen, “Approximation capability to functions of several variables, nonlinear functions, and operators by radial function neural networks,” *IEEE Transactions on Neural Networks*, vol. 6, pp. 904–910, 1995.

[7] A. R. Barron, “Universal approximation bounds for superpositions of a sigmoidal function,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930–945, 1993.

[8] M. Leshno, V. Ya. Lin, A. Pinkus, and S. Schocken, “Multilayer feedforward networks with a non-polynomial activation function and approximate any function,” *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993.

[9] H. N. Mhaskar, “Neural networks for optimal approximation for smooth and analytic functions,” *Neural Computation*, vol. 8, pp. 164–177, 1996.

[10] V. Maiorov and R. S. Meir, “Approximation bounds for smooth functions in $C(R^d)$by neural and mixture networks,” *IEEE Transactions on Neural Networks*, vol. 3, pp. 969–978, 1998.

[11] M. Burger and A. Neubauer, “Error bounds for approximation with neural networks,” *Journal of Approximation Theory*, vol. 112, no. 2, pp. 235–250, 2001.

[12] V. Kurkova and M. Sanguineti, “Comparison of worst case errors in linear and neural network approximation,” *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 264–275, 2002.

[13] J. L. Wang, B. H. Sheng, and S. P. Zhou, “On approximation by non-periodic neural and translation networks in L_w^p spaces,” *ACTA Mathematica Sinica*, vol. 46, pp. 65–74, 2003 (Chinese).

[14] J. J. Wang, C. Yang, and J. Jing, “Approximation order for multivariate Durrmeyer operators with Jacobi weights,” *Abstract and Applied Analysis*, vol. 2011, Article ID 970659, 12 pages, 2011.

[15] A. F. Timan, *Theory of Approximation of Functions of a Real Variable*, Macmillan, New York, NY, USA, 1963.