A Constructive Proof and An Extension of Cybenko's Approximation Theorem

Tianping Chen

Dept. of Mathematics
Fudan University
Shanghai, China

Hong Chen and Ruey-wen Liu

Dept. of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA

Abstract. In this paper, we present a constructive proof of approximation by superposition of sigmoidal functions. We point out a sufficient condition that the set of finite linear combinations of the form $\sum\alpha_j\sigma(y_j\cdot x+\theta_j)$ is dense in $C(\mathbb{I}^n)$, is the boundedness of the sigmoidal function $\sigma(x)$. Moreover, we show that if the set of finite linear combinations of the form $\sum c_j\omega(\xi_j x+\eta_j)$, where ω is a univariate function, is dense in $L^p[a,b]$ ($1\le p<\infty$) (or $C[a,b]$) for any finite a,b, then the set of finite linear combinations of the form $\sum c_j\omega(y_j\cdot x+\theta_j)$ is dense in $L^p(\mathbb{I}^n)$ (or $C(\mathbb{I}^n)$). An extension in another direction is also presented in Theorem 4 of this paper.

Key words. Constructive approximation, Neural networks, Sigmoidal functions.

1. Introduction

Recently, Cybenko[1] stated that if $\sigma(x)$ is a scalar continuous sigmoidal function, then the set of linear combinations of the form

$$\sum_{j=1}^{N}\alpha_j\sigma(y_j\cdot x+\theta_j) \qquad (1)$$

where $x, y_j\in\mathbb{R}^n$, $x\cdot y_j$ is their inner product, and $\alpha_j, \theta_j\in\mathbb{R}$, is dense in the space of all continuous functions on the hypercube $\mathbb{I}^n$, denoted by $C(\mathbb{I}^n)$. Similar problems were also considered in [4],[5],[6] and [7].

This result, not only settles an long-standing problem on the realizability of $f\in C(\mathbb{I}^n)$ by single hidden layer feedforward artificial neural networks, but mathematically is a suitable practical substitution of the well-known Kolmogorov's resolution to Hilbert's 13th Problem. In his work[2], Kolmogorov showed that any continuous function of n variables has an *exact* representation in terms of finite superpositions and compositions of a finite number of univariate functions. In [1], it is shown that any continuous functions on the compact set [0,1] can be *approximated* in terms of finite superpositions of a *single* sigmoidal univariate function as closely as desired.

Due to its theoretical value and its variety of applications in neural networks and other fields, it is important to give a *constructive* proof of this problem, which is not available at the present time. In this paper, we will provide such a constructive proof. Furthermore, we will show that the continuity assumption imposed on the sigmoidal functions, which was, in fact, strictly required in [1], is unnecessary. Instead, the boundedness of the sigmoidal function plays an essential role. In some sense, the boundedness of the sigmoidal function is necessary and sufficient for the validity of the approximation theorem. We will also give various generalizations and extensions of our result.

Definition. $\sigma: \mathbb{R}\to\mathbb{R}$ is called a sigmoidal function, if the limits $\lim_{x\to\infty}\sigma(x)$, $\lim_{x\to-\infty}\sigma(x)$ exist, and $\lim_{x\to\infty}\sigma(x)=1$, $\lim_{x\to-\infty}\sigma(x)=0$.

Note: σ need not be continuous.

Example. $\sigma(x)=\dfrac{1}{1+\exp(-x)}$
is a sigmoidal function, also

$$\sigma(x)=\begin{cases}1 & x\ge 0\\ 0 & x<0\end{cases}$$

is a sigmoidal function.

C. Page et al. (eds.), *Computing Science and Statistics*

2. Main Results

Theorem 1. If $\sigma(x)$ is a bounded sigmoidal function, and $f(x)$ is a continuous function on $(-\infty,\infty)$, for which $\lim_{x\to-\infty} f(x)=A$ and $\lim_{x\to\infty} f(x)=B$, where A, B are constants, then for any $\varepsilon>0$, there exist N, c_i, y_i, θ_i, such that

$$\left| f(x)-\sum_{i=1}^{N} c_i\sigma(y_ix+\theta_i) \right| <\varepsilon \qquad (2)$$

holds for all $x\in(-\infty,\infty)$.

In other words, the set of finite linear combinations of the form $\sum_{i=1}^{N} c_i\sigma(y_ix+\theta_i)$ is dense in the space of continuous functions satisfying the condition that the limits $\lim_{x\to-\infty} f(x)$ and $\lim_{x\to\infty} f(x)$ exist.

Proof. The proof is constructive. From the assumption, for any $\varepsilon>0$, we can find $M>0$, such that $|f(x)-A|<\frac{\varepsilon}{4}$ if $x<-M$; $|f(x)-B|<\frac{\varepsilon}{4}$ if $x>M$; and $|f(x')-f(x'')|<\frac{\varepsilon}{4}$ if $|x'|\le M$, $|x''|\le M$, and $|x'-x''|\le\frac{1}{M}$.

Divide $[-M,M]$ into $2M^2$ equal segments, each has length of $\frac{1}{M}$, and let

$$-M=x_0<x_1<\dots<x_{M^2}=0<x_{M^2+1}<\dots<x_{2M^2}=M \qquad (3)$$

Let $t_i=\frac{1}{2}(x_i+x_{i+1})$ be the center of the interval $[x_i,x_{i+1}]$.

Now, construct

$$g(x) = f(-M) + \sum_{i=1}^{N}[f(x_i)-f(x_{i-1})]\,\sigma(K(x-t_{i-1})) \qquad (4)$$

where $N=2M^2$.

From the assumption, there exists $W>0$, such that if $u>W$ then $|\sigma(u)-1|<\frac{1}{M^2}$, if $u<-W$ then $|\sigma(u)|<\frac{1}{M^2}$. Let $K>0$ such that $K\cdot\frac{1}{2M}>W$.

We claim: $|f(x)-g(x)|<\varepsilon$ holds for all $x\in(-\infty,\infty)$.

(i) If $x<-M$, then $|f(x)-f(-M)|<\frac{\varepsilon}{2}$,

$$|g(x)-f(-M)| \le \sum_{i=1}^{2M^2}|f(x_i)-f(x_{i-1})|\,|\sigma(K(x-t_{i-1}))|$$

$$< \sum_{i=1}^{2M^2} \frac{\varepsilon}{4}\frac{1}{M^2} = \frac{\varepsilon}{2} \qquad (5)$$

Consequently $|f(x)-g(x)|<\varepsilon$.

(ii) If $x>M$, then $|f(x)-f(M)|<\frac{\varepsilon}{2}$,

$$|g(x)-f(M)| \le \sum_{i=1}^{2M^2}|f(x_i)-f(x_{i-1})|\,|\sigma(K(x-t_{i-1}))-1|$$

$$< \sum_{i=1}^{2M^2} \frac{\varepsilon}{4}\frac{1}{M^2} = \frac{\varepsilon}{2} \qquad (6)$$

Consequently $|f(x)-g(x)|<\varepsilon$.

(iii) Consider $x\in[x_{k-1},x_k]$, then $|x-t_{i-1}|\le\frac{1}{2M}$ if $i=k$; $|x-t_{i-1}|>\frac{1}{2M}$ if $i\ne k$. Furthermore, if $i<k$, then $K(x-t_{i-1})>W$, and hence $|\sigma(K(x-t_{i-1}))-1|<\frac{1}{M^2}$; if $i>k$, then $K(x-t_{i-1})<-W$, and hence $|\sigma(K(x-t_{i-1}))|<\frac{1}{M^2}$.

Consequently, we have

$$\Big| g(x) - f(-M) - [f(x_k)-f(x_{k-1})]\,\sigma(K(x-t_{k-1})) - \sum_{i=1}^{k-1}[f(x_i)-f(x_{i-1})] \Big|$$

$$\le \sum_{i=1}^{k-1}|f(x_i)-f(x_{i-1})|\,|\sigma(K(x-t_{k-1}))-1|$$

$$+ \sum_{i=k+1}^{2M^2}|f(x_i)-f(x_{i-1})|\,|\sigma(K(x-t_{i-1}))|$$

$$\leq \sum_{i=1}^{k-1} \frac{\varepsilon}{4}\,\frac{1}{M^2} + \sum_{i=k+1}^{2M^2} \frac{\varepsilon}{4}\,\frac{1}{M^2}$$

$$< \frac{\varepsilon}{2}$$

It is clear that

$$f(-M)+[f(x_k)-f(x_{k-1})]\,\sigma(K(x-t_{k-1})) + \sum_{i=1}^{k-1}[f(x_i)-f(x_{i-1})]$$

$$=f(x_{k-1})+ [f(x_k)-f(x_{k-1})]\,\sigma(K(x-t_{k-1}))$$

which, combined with the previous inequality, yields

$$|g(x)-f(x)| < \frac{\varepsilon}{2} + |f(x)-f(x_{k-1})| + |f(x_k)-f(x_{k-1})||\sigma(K(x-t_{k-1}))|$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon$$

Here, we have assumed , without loss of generality, $|\sigma(x)|\leq 1$ for $x\in(-\infty,\infty)$.

To complete the proof of the theorem, we need only to take care of $f(-M)$, or to show in addition that the constant 1 can be approximated by the linear combinations $\sum c_i\sigma(y_ix+\theta_i)$.

Let N,M,x_i,t_i,K be the same constants as before, then it is easy to verify that

$$\frac{1}{N}\sum_{i=1}^{N}[\sigma(K(x-t_{i-1}))+\sigma(-K(x-t_{i-1}))] \tag{7}$$

is the required sum. The proof is just a repetition of the previous procedure, and the details are omitted.

Q.E.D.

Corollary. If $f(x)$ is a continuous function on some finite interval $[a,b]$, and $\sigma(x)$ is a bounded sigmoidal function, then for any $\varepsilon>0$, there exist N, c_i, y_i, θ_i, such that

$$\Big|f(x)-\sum_{i=1}^{N}c_i\sigma(y_ix+\theta_i)\Big|<\varepsilon \tag{8}$$

holds for all $x\in[a,b]$. In other words, the set of finite linear combinations of the form $\sum_{i=1}^{N}c_i\sigma(y_ix+\theta_i)$ is dense in $C[a,b]$. Furthermore, in this case, we can choose all $y_i>0$. This is because the constant function on $[a,b]$ can be approximated by $\sum_{i=1}^{N}c_i\sigma(y_ix+\theta_i)$ with all y_i positive.

Theorem 2. If $f(\mathbf{x})$ is a continuous function on $\mathbb{I}^n=[0,1]^n$, $\sigma(u)$ is a bounded sigmoidal function. Then for any $\varepsilon>0$, there exist $N,c_i,\theta_i\in\mathbb{R}$, $\mathbf{y}_i\in\mathbb{R}^n$, such that for any $\mathbf{x}\in\mathbb{I}^n$

$$\Big|f(\mathbf{x})-\sum_{i=1}^{N}c_i\sigma(\mathbf{y}_i\cdot\mathbf{x}+\theta_i)\Big|<\varepsilon \tag{9}$$

where $\mathbf{x}\cdot\mathbf{y}$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$. In other words, the set of finite linear combinations of the form $\sum_{i=1}^{N}c_i\sigma(\mathbf{y}_i\cdot\mathbf{x}+\theta_i)$ is dense in $C[0,1]^n$.

Proof. Extend $f(\mathbf{x})$ to be a function $g(\mathbf{x})$ on $\mathbb{J}^n=[-1,1]^n$ according to the following rule: $g(\mathbf{x})=f(\mathbf{x})$ if $\mathbf{x}\in\mathbb{I}^n$, and $g(x_1,\ldots,-x_k,\ldots,x_n)=g(x_1,\ldots,x_k,\ldots,x_n)$. Thus $g(\mathbf{x})$ can be thought of as a 2-periodic even function with respect to every variable x_i, $i=1,\ldots,n$.

Let $c_{m_1,\ldots,m_n}$ be the Fourier coefficients of $g(x_1,\ldots,x_n)$ on $\mathbb{J}^n$. By a well known result on Bochner-Riesz Means ([3], page 256): for any $\varepsilon>0$, there exists R, such that for any $\mathbf{x}=(x_1,\ldots,x_n)\in\mathbb{J}^n$,

$$\Big|\sum_{|\mathbf{m}|\leq R}\Big(1-\frac{|\mathbf{m}|^2}{R^2}\Big)^{\alpha} c_{m_1,\ldots,m_n}\exp\big(i\,\pi(m_1x_1+\ldots+m_nx_n)\big) - g(x_1,\ldots,x_n)\Big| < \frac{\varepsilon}{2} \tag{10}$$

where $\mathbf{m}=(m_1,\ldots,m_n)$, $|\mathbf{m}|^2 = m_1^2+\ldots+m_n^2$, and $\alpha > \frac{n-1}{2}$.

By the definition of the Fourier coefficients and the evenness of $g(\mathbf{x})$, we can rewrite the previous inequality as

$$\left|\sum_{|m|\le R} d_{m_1,\ldots,m_n}\cos(m_1x_1+\ldots+m_nx_n) - g(x_1,\ldots,x_n)\right| < \frac{\varepsilon}{2} \tag{11}$$

where $d_{m_1,\ldots,m_n}$ are real numbers.

It is obvious that for any $x\in\mathbb{I}^n$, there is a unique $u\in[-\sum_{i=1}^{n}|m_i|, \sum_{i=1}^{n}|m_i|]$ such that

$$u=\mathbf{m}\cdot x=m_1x_1+\ldots+m_nx_n$$

Because cos(u) is a continuous function on $[-\sum_{i=1}^{n}|m_i|, \sum_{i=1}^{n}|m_i|]$, by the Corollary of Theorem 1, we can find N^m, S^m, ξ_j^m, η_j^m, such that

$$\left|\sum_{j=1}^{N^m} S_j^m\sigma(\xi_j^m u+\eta_j^m)-\cos(u)\right| < \frac{\varepsilon}{2L}$$

holds uniformly for $u\in[-\sum_{i=1}^{n}|m_i|, \sum_{i=1}^{n}|m_i|]$, where $\mathbf{m}=(m_1,\ldots,m_n)$ and $L=\sum_{|m|\le R}|d_{m_1,\ldots,m_n}|$. Thus

$$\left|\sum_{j=1}^{N^m} S_j^m\sigma(\xi_j^m\mathbf{m}\cdot x+\eta_j^m)-\cos(\mathbf{m}\cdot x)\right| < \frac{\varepsilon}{2L} \tag{12}$$

holds for $x\in\mathbb{I}^n$.

Substituting (12) into the inequality (11), we conclude that there exist $N, c_i, \theta_i\in\mathbb{R}$, $y_i\in\mathbb{R}^n$, such that

$$\left|f(x)-\sum_{i=1}^{N}c_i\sigma(y_i\cdot x+\theta_i)\right|<\varepsilon \tag{13}$$

is true for all $x\in\mathbb{I}^n$.

Q.E.D.

Remark 1. It is clear in the proofs of Theorems 1 and 2 that the continuity of σ(x) is unnecessary for the validity of the the Theorems. The boundedness of σ(x) is sufficient for Theorems 1 and 2 to be true.

Remark 2. It can be shown that if σ(x) is a measurable but unbounded sigmoidal function, then $\sum c_i\sigma(y_i\cdot x+\theta_i)$ may not necessarily be dense in $C(\mathbb{I}^n)$. In fact, we can define

$$\sigma(x)=\begin{cases}\frac{1}{x}+1 & x>0\\ 0 & x=0\\ -\frac{1}{x} & x<0\end{cases} \tag{14}$$

which is clearly a sigmoidal function. But the set $\sum c_i\sigma(y_ix+\theta_i)$ is not dense in C[0,1]. Therefore, we conclude that the boundedness of σ(x) is an essential assumption for the validity of Theorem 2.

Remark 3. If σ(x) is a monotone sigmoidal function, then it is clear that $0\le\sigma(x)\le 1$. As a consequence of Theorem 2, we have that if σ(x) is a monotone sigmoidal function, then the finite linear combinations $\sum c_i\sigma(y_i\cdot x+\theta_i)$ are dense in $C(\mathbb{I}^n)$.

Since $C(\mathbb{I}^n)$ is dense in all $L^p(\mathbb{I}^n)$ $(1\le p\le\infty)$, we have

Theorem 3. If σ(x) is a bounded sigmoidal function, then the set of finite linear combinations $\sum_{i=1}^{N}c_i\sigma(y_i\cdot x+\theta_i)$ is dense in $L^p(\mathbb{I}^n)$ $(1\le p\le\infty)$. (See [1]).

For unbounded sigmoidal functions, we can prove

Theorem 3'. If a sigmoidal function $\sigma(x)\in L^p[a,b]$ $(1\le p\le\infty)$ for every pair of $a,b\in\mathbb{R}$, $a<b$, then the set of finite linear combinations $\sum_{i=1}^{N}c_i\sigma(y_i\cdot x+\theta_i)$ is dense in $L^p(\mathbb{I}^n)$.

In fact, what we need to prove is that, under the assumption of Theorem 3', the Corollary of Theorem 1 is true in $L^p[a,b]$, instead of in C[a,b]. The details are omitted.

Remark 4. If ω(x) is a measurable function and $\omega(x)\in L^p[a,b]$ $(1\le p\le\infty)$, for every $a,b\in\mathbb{R}$, $a<b$, and the set of

finite linear combinations $\sum c_i\omega(\xi_i x+\eta_i)$ is dense in any $L^p[a,b]$, then the set of finite linear combinations $\sum c_i\omega(y_i\cdot x+\theta_i)$ is dense in $L^p(\mathbb{I}^n)$. (Note a special case of this is when $\omega(x)\in C[a,b]$.) This is a mild condition satisfied by many $\omega(x)$. Among those are many functions that frequently occur in approximation theory, such as the well known Schoenberg Cardinal Splines, B-Splines and many others, including those functions satisfying Wiener-Tauberian conditions. For proof, we can follow the same line developed in the proof of Theorem 2, approximating $\exp(i\,(m_1x_1+\cdots+m_nx_n))$ by the finite linear combinations $\sum c_j\omega(\lambda_j(m_1x_1+\cdots+m_nx_n)+\theta_j)$ and using Fourier series development.

We now give another kind of generalization.

First, we will call $[-\infty,\infty]$ the extended real line, with $(-\infty,A)$ as the neighborhood of the point $-\infty$, and $(A,+\infty)$ the neighborhood of $+\infty$. Thus $[-\infty,\infty]$ is a compact space, and $[-\infty,\infty]^n$, being the product of compact spaces, is also compact.

Similar as before, we denote $C[-\infty,\infty]^n$ the set of continuous functions on the compact set $[-\infty,\infty]^n$.

With this notation, we can state Theorem 1 as " the set of finite linear combinations $\sum c_i\sigma(y_ix+\theta_i)$ is dense in $C[-\infty,\infty]$ ".

Theorem 4. If $\sigma(x)$ is a bounded sigmoidal function, then the set of finite linear combinations

$$\sum_{i_1=1}^{N_1}\cdots\cdots\sum_{i_n=1}^{N_n} c_{i_1,\ldots,i_n}\,\sigma(y_{i_1}x_1+\theta_1)\ldots\sigma(y_{i_n}x_n+\theta_n)$$

is dense in $C[-\infty,\infty]^n$.

Proof. Without loss of generality, we assume n=2. Let $f(x,y)\in C[-\infty,\infty]^2$, then $f(x,y)$, fixing y, is uniformly continuous on the compact set $[-\infty,\infty]$. By Theorem 1: for any $\varepsilon>0$, there exist N_1,K_1 such that

$$| f(x,y) - f(-M,y) \\ - \sum_{i=1}^{N_1}[f(x_i,y)-f(x_{i-1},y)]\,\sigma(K_1(x-t_{i-1}))\,| < \frac{\varepsilon}{2} \qquad (15)$$

holds for all $(x,y)\in[-\infty,\infty]^2$.

Writing $C_i(y)=f(x_i,y)-f(x_{i-1},y)$, $i=1,\ldots,N_1$, $C_0(y)=f(-M,y)$. Repeating the argument as in the proof of Theorem 1, there holds

$$| C_i(y) - \sum_{i=1}^{N_2} c_{i,j}\sigma(\xi_jy+\eta_j))\,| < \frac{\varepsilon}{2N_1} \qquad (i=0,1\ldots,N_1) \qquad (16)$$

hence

$$| f(x,y) - \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} c_{i,j}\,\sigma(\omega_iy+\theta_i)\,\sigma(\xi_jy+\eta_j)\,| < \varepsilon \qquad (17)$$

Q.E.D.

Corollary 1. If $\sigma(x)$ is a bounded sigmoidal function, then the set of finite linear combinations

$$\sum_{i_1=1}^{N_1}\cdots\cdots\sum_{i_n=1}^{N_n} c_{i_1,\ldots,i_n}\,\sigma(y_{i_1}x_1+\theta_1)\cdots\sigma(y_{i_n}x_n+\theta_n)$$

is dense in $C[0,1]^n$.

Remark 5. Besides sigmoidal functions, for many systems, there holds similar theorem: If $\omega_i(x)$, $i=0,\pm1,\ldots$, is a system of functions, and if $\sum c_i\omega_i(x)$ are dense in $C[0,1]$, then $\sum c_{i_1,\ldots,i_n}\,\omega_{i_1}(x_1)\cdots\omega_{i_n}(x_n)$ are dense in $C[0,1]^n$, if only $\omega_i(x)$ satisfies very general conditions. For example, $\{\exp(ik\pi x)\}_{k=-\infty}^{+\infty}$ is such a system. In this case $\omega_{i_1}(x_1)\cdots\omega_{i_n}(x_n) = \omega(i_1x_1+\cdots+i_nx_n)$. This is just the reason why we used $e^{i\,m\cdot x}$ as the medium in the proof of Theorem 2. Of course we can also use $e^{m\cdot x}$ as the medium instead.

References

[1] G.Cybenko, "Approximation by Superpositions of a Sigmoidal Function", *Mathematics of Control, Signals and Systems*, V.2, No.4 (1989), P.303-314.

[2] A.N.Komogorov, “On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of One Variable and Addition”, *Dokl. Akad. Nawk. SSSR*, 114, 1957 (pp.953-956) English translation, *American Math. Soc. Transl.* (2), 28 (1963), pp.55-59, MR.22#2669; 27#3760.

[3] E.M.Stein and Guido Weiss, "Introduction to Fourier Analysis on Euclidean Spaces", Princeton University Press (1971).

[4] L.K.Jones, "Constructive Approximation for Neural Networks by Sigmoidal Functions", Technical Report Series, No.7, Dept. of Mathematics, University of Lowell (1988).

[5] S.M.Carrol and B.W.Dickinson, "Construction of Neural Nets Using Radon Transform", Preprint 1989.

[6] K.Funahashi, "On the Approximate Realization of Continuous Mappings by Neural Networks", *Journal of International Neural Networks* (to appear).

[7] K.Hornik, M.Stinchcombe and H.White, "Multilayer Feedforward Networks are Universal Approximators", Preprint 1988.