Appeared in IEEE Transactions on Neural Networks, Nov. issue, 1993.

Approximations of Continuous Functionals by Neural Networks with Application to Dynamical Systems

Tianping Chen[1] and Hong Chen[2]

Abstract

The main concern of this paper is to give several strong results on neural network representation in an *explicit* form. Under very mild conditions, a functional defined on a compact set in $C[a,b]$ or $L^p[a,b]$, spaces of *infinite* dimensions, can be approximated arbitrarily well by a neural network with one hidden layer. In particular, if U is a compact set in $C[a,b]$, σ is a bounded sigmoidal function, and f is a continuous functional defined on U, then for all $u \in U$, $f(u)$ can be approximated by $\sum_{i=1}^{N} c_i\, \sigma(\sum_{j=0}^{m} \xi_{i,j}\, u(x_j) + \theta_i)$, where c_i, ξ_{ij}, θ_i are real numbers. $u(x_j)$ is the value of u evaluated at point x_j. These results are a significant development beyond the original works (see [1-9]), where theorems of approximating continuous functions defined on $\mathbf{R}^n$, a space of *finite* dimension by neural networks with one hidden layer were given. Finally, all the results are shown applicable to the approximation of the output of dynamical systems at any particular time.

Key words: Approximation theory, neural networks, dynamical systems, compact set, convex set, functional.

[1]The author is with the Department of Mathematics, Fudan University, Shanghai, P.R.China.
[2]The author is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA.

1 Introduction

The problem of approximating a function of several variables by neural network has been studied by many authors. In 1987, Wieland and Leighton dealt with the capability of networks consisting of one or two hidden layers. Miyake (1988) obtained an integral representation formula with an integrable kernel fixed beforehand. This representation formula is a kind which would be realized by a three-layered neural network. In 1989, several papers related to this topic appeared. They all claimed that a three-layered neural network with sigmoid units on the hidden layer can approximate continuous or other kinds of functions defined on compact sets in $\mathbf{R^n}$. They used different methods. Carrol and Dickinson used the inverse Radon transformation. Cybenko used the functional analysis method, combining the Hahn-Banach theorem and Riesz representation theorem. However, his proof is existential. Funahashi approximated Irie and Miyake's integral representation by a finite sum, using a kernel which can be expressed as a difference of two sigmoid functions. Hornik et al. applied the Stone-Weierstrass theorem, using trigonometric functions, where their approximations were not only in the uniform topology on a compact set, but also in the ρ_μ-topology. However, the latter can be attained if the uniform approximation can be attained in any compact set, because the uniform convergence topology is stronger than ρ_μ-topology.

Recently [9], we gave a *constructive* approach of proving the above result, and proved that instead of the continuity of $\sigma(x)$, a sufficient condition for Cybenko's theorem to be true is the *boundedness* of $\sigma(x)$. Moreover, if $(f_1(x), \ldots, f_q(x))$ is a continuous map from $[0,1]^n$ to $\mathbf{R}^q$, then for any $\epsilon > 0$, there exist N, c_j, $\theta_i \in \mathbf{R}$, $y_j \in \mathbf{R}^n$, $c_{j,k} = c_j(f_k) \in \mathbf{R}$, $j = 1, \ldots, N$, $k = 1, \ldots, q$, such that

$$|f_k(x) - \sum_{j=1}^{N} c_{j,k}\, \sigma(y_j \cdot x + \theta_j)| < \epsilon \tag{1}$$

for all $x \in [0,1]^n$, and $f_1, \ldots, f_q$, where $x \cdot y$ is the inner-product of x and y.

This type of approximation theorem is useful in the theory and application of artificial neural networks, since many types of neural networks are formed from compositions and superpositions of one simple nonlinear activation function. A nontrivial but simple class of neural networks are those with one hidden layer and they exactly implement the set of functions given by

$$\sum_{j=1}^{N} c_j\, \sigma(y_j \cdot x + \theta_j) \tag{2}$$

Approximation of functions by neural networks is not only interesting and meaningful in pure and applied mathematics, but also useful in engineering and physical sciences, where such approximations have found wide applications in areas such as system identification, modeling and realization, signal decomposition and generation, pattern classification, adaptive filering, etc. Theoretically, the aforementioned result not only settles a long-standing question on the realizability of $C([0,1]^n)$ by a single hidden layer feedforward neural network, but also is an alternative substitution to Kolmogorov's well-known resolution to Hilbert's 13th Problem. In a famous paper, Kolmogorov proved that for any continuous function defined on $[0,1]^n$, there is the following representation

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} g(\lambda_1 \phi_q(x_1) + \ldots + \lambda_n \phi_q(x_n)) \tag{3}$$

where g, $\phi_i(x)$, $i = 1, \ldots, n$ are functions of a single variable, $0 \leq \lambda_i \leq 1$, $i = 1, \ldots, n$, and $0 \leq \phi_q(t) \leq 1$, $t \in [0,1]$, $q = 1, \ldots, n$. However, the construction of g and ϕ_q is very complicated. Cybenko's theorem shows that every continuous function defined on $[0,1]^n$ can be approximated within any prescribed error by finite linear combination as in (2) where σ is a very simple univariate function.

All these works are concerned with approximation to continuous functions defined on a compact set in $\mathbf{R}^n$ (a space of *finite* dimensions). However, in practice we often encounter situations where we need to compute functionals defined on some set of functions (a space of *infinite* dimensions). For example, the output of a dynamical system at any particular time can be viewed as a functional (see Example 1 in Section 3). Thus it is of great importance to discuss the problem of approximation to nonlinear functionals by neural networks. This is the main motivation and concern of our paper.

Recently, Sandberg did important work [11] and obtained interesting approximation theorems for discrete-time dynamical systems. Despite the restriction to the discrete-time case, his work began to reveal the possibility and effectiveness of using neural networks (with so-called sigmoidal nonlinearity, as will be introduced shortly, or more general nonlinearity) in approximating dynamical systems. The further treatment (including the more general continuous-time systems) of this topic, however, remains still unclear. Especially, we ask: *can we give any result with a form as explicit as in (2) ?*

This paper is organized as follows, we first concentrate on approximating continuous functionals by (single-hidden layer feedforward) neural networks. We will obtain several strong results, which are of high interest in the research of neural network representation capability. Then, we study approximation of dynamical systems and will provide a uniform viewpoint and treatment to both continuous-time and discrete-time systems. All the results presented in the first part of this paper can be readily applied to the approximation of the outputs of dynamical systems (at any particular time). As one example, some of the results in [11] can be obtained from our results in this paper. Thus, ours are a significant generalization of those in [11].

2 Approximations of Continuous Functionals

Let $C[a,b]$ and $L^p[a,b]$ denote the space of all continuous functions and p-th integrable functions on $[a,b]$, with norms $\|f\|_{C[a,b]} = \sup_{x\in[a,b]} |f(x)|$, and $\|f\|_{L^p[a,b]} = (\int_a^b |f(x)|^p\, dx)^{1/p}$ respectively. Throughout this paper, unless otherwise specified, we shall always let $1 < p < \infty$. A set U in $C[a,b]$ is called *compact* if for any sequence $f_n \in U$, there exists a function f in U and a subsequence f_{n_k} of f_n, such that $\|f - f_{n_k}\|_{C[a,b]} \to 0$. Similarly, we can define a compact set in $L^p[a,b]$. If $\sigma(\cdot)$: $\mathbf{R} \to \mathbf{R}$ satisfies

$$\sigma(x) \to \begin{cases} 1 & \text{as} \quad x \to +\infty \ , \\ 0 & \text{as} \quad x \to -\infty \ . \end{cases} \tag{4}$$

then we call $\sigma(\cdot)$ a *generalized sigmoidal function.* It is worth noting that all monotone increasing sigmoidal functions belong to this class. Moreover, continuity of $\sigma(x)$ is not required in this definition or in later theorems.

The main results of this paper are as follows.

2.1 Main Results

Theorem 1 *Suppose that U is a compact set in $L^p[a,b]$ $(1 < p < \infty)$, f is a continuous functional defined on U, and $\sigma(x)$ is a bounded generalized sigmoidal function, then for any $\epsilon > 0$, there exist $h > 0$, a positive integer m, $m+1$ points $a = x_0 < x_1 < \cdots < x_m = b$, $x_j = a + j(b-a)/m$, $j = 0, 1, \ldots, m$, a positive integer N and constants c_i, θ_i, $\xi_{i,j}$, $i = 1, \ldots, N$, $j = 0, 1, \ldots, m$, such that*

$$|f(u) - \sum_{i=1}^{N} c_i\, \sigma(\sum_{j=0}^{m} \xi_{i,j}\, \frac{1}{2h} \int_{x_j-h}^{x_j+h} u(t)\, dt + \theta_i)| < \epsilon \tag{5}$$

holds for all $u \in U$. Here it is assumed that $u(x) = 0$, if $x \notin [a,b]$. □

Theorem 2 *Suppose that U is a compact set in $C[a,b]$, f is a continuous functional defined on U, and $\sigma(x)$ is a bounded generalized sigmoidal function, then for any $\epsilon > 0$, there exist $m+1$ points $a = x_0 < \cdots < x_m = b$, a positive integer N and constants c_i, θ_i, $\xi_{i,j}$, $i = 1, \ldots, N$, $j = 0, 1, \ldots, m$, such that*

$$|f(u) - \sum_{i=1}^{N} c_i\, \sigma(\sum_{j=0}^{m} \xi_{i,j}\, u(x_j) + \theta_i)| < \epsilon\ , \qquad \forall u \in U\ . \tag{6}$$

□

Suppose that $V = \mathbf{R}^q$, $\prod_{k=1}^{n}[a_k, b_k]$ is a rectangle in $\mathbf{R}^n$, and $C_V(\prod_{k=1}^{n}[a_k, b_k])$ stands for the set of all continuous maps $G(x_1, \ldots, x_n) = (g_1(x_1, \ldots, x_n), \ldots, g_q(x_1, \ldots, x_n))$ defined on $\prod_{k=1}^{n}[a_k, b_k]$, taking values in V, that is, each $g_l(x_1, \ldots, x_n)$ is continuous in $\prod_{k=1}^{n}[a_k, b_k]$, $l = 1, \ldots, q$. Moreover, if G, $F \in C_V(\prod_{k=1}^{n}[a_k, b_k])$, define

$$\rho(G, F)_C = \sqrt{\sum_{l=1}^{q} \|g_l(x_1, \ldots, x_n) - f_l(x_1, \ldots, x_n)\|_C^2}\ . \tag{7}$$

Similarly, let $L^p_V(\prod_{k=1}^n [a_k, b_k])$ denote all the mappings $(g_1, \ldots, g_q)$, where each g_l $(x_1,$ ..., $x_n)$ is p-th integrable over $\prod_{k=1}^n [a_k, b_k]$. If G, $F \in L^p_V(\prod_{k=1}^n [a_k, b_k])$, define

$$\rho(G,F)_{L^p} = \sqrt{\sum_{l=1}^{q} \|g_l(x_1,\ldots,x_n) - f_l(x_1,\ldots,x_n)\|^2_{L^p}} \ . \tag{8}$$

Theorem 3 *Suppose that U is a compact set in $L^p_V(\prod_{k=1}^n [a_k, b_k])$ $(1 < p < \infty)$, f is a continuous functional on U, and $\sigma(x)$ is a bounded generalized sigmoidal function, then for any $\epsilon > 0$, there exist $(m+1)^n$ points $(x_1^{j_1}, \ldots, x_n^{j_n})$, $x_k^{j_k} = a_k + j_k(b-a)/m$, $j_k = 0, 1, \ldots, m$, $k = 1, \ldots, n$, a positive integer N and constants c_i, θ_i and $q \times (m+1)^n$-vectors $\bar{\xi}_i$, such that*

$$|f(u) - \sum_{i=1}^{N} c_i\, \sigma(\bar{\xi}_i \cdot \bar{u}^*_{q,n,m} + \theta_i)| < \epsilon \ , \qquad \forall u \in U \ , \tag{9}$$

where $\bar{u}^*_{q,n,m}$ are $q \times (m+1)^n$-vectors obtained by replacing $u_l(x_1^{j_1}, \ldots, x_n^{j_n})$ in $\bar{u}_{q,n,m}$ by $(\frac{1}{2h})^n \int \cdots \int_{x_k^{j_k} - h \le x_k \le x_k^{j_k} + h} u_l(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$. □

Theorem 4 *Suppose that U is a compact set in $C_V(\prod_{k=1}^n [a_k, b_k])$, f is a continuous functional defined on U, and $\sigma(x)$ is a bounded generalized sigmoidal function, then for any $\epsilon > 0$, there exist $(m+1)^n$ points $(x_1^{j_1}, \ldots, x_n^{j_n})$, $x_k^{j_k} = a_k + j_k(b-a)/m$, $j_k = 0, 1, \ldots, m$, $k = 1, \ldots, n$, a positive integer N and constants c_i, θ_i and $q \times (m+1)^n$-vectors $\bar{\xi}_i$, such that*

$$|f(u) - \sum_{i=1}^{N} c_i\, \sigma(\bar{\xi}_i \cdot \bar{u}_{q,n,m} + \theta_i)| < \epsilon \ , \qquad \forall u \in U \ , \tag{10}$$

where $\bar{u}_{q,n,m} = (u_l(x_1^{j_1}, \ldots, x_n^{j_n}))$, $l = 1, \ldots, q$, $j_k = 0, 1, \ldots, m$, $k = 1, \ldots, n$ are $q \times (m+1)^n$-vectors. □

2.2 Remarks

We first explain the significance of these theorems by the following remarks.

Remark 1: Most of the papers published (see [1-9]) discuss the problem of approximating a continuous function defined on some compact subset in $\mathbf{R}^n$ (a space of finite dimensions). Instead, this paper discusses approximation to continuous *functionals* defined on some compact subset in some space of functions (a space of *infinite* dimensions), thus making the problem much more difficult and complicated than that in the *finite* dimensional case.

Remark 2: Our above theorems not only solve the representation capability by neural networks, but also give an *explicit* form of the approximant. These results can not be obtained by the Stone-Weierstrass theorem, which forms the basis of several papers. It is known that the Stone-Weierstrass theorem is existential and gives no explicit form for the approximant.

Remark 3: Let $a = t_1 < t_2 < \ldots < t_n = b$, $t_i = a + (i-1)\frac{b-a}{n-1}$, then for any point $x = (x_1, \ldots, x_n) \in [0,1]^n$, define a function $u_x(t)$ as follows

$$u_x(t) = x_i + \frac{x_{i+1} - x_i}{t_{i+1} - t_i}(t - t_i), \qquad t_i \leq t \leq t_{i+1}, \qquad i = 1, \ldots, n-1 \tag{11}$$

which is a piecewise linear function taking values x_j at point t_j.

Let U be the set of all these functions defined above. It is easy to verify that there is a one-to-one mapping between U and $[0,1]^n$. Moreover, for any continuous function $f(x_1, \ldots, x_n)$ defined on $[0,1]^n$, there corresponds a unique functional $f(u)$ defined on U, which is a compact set in $C[a,b]$. By Theorem 2, $f(x_1,\ldots,x_n) = f(u_x)$ can be approximated by $\sum c_i\sigma(\sum_{j=1}^m \xi_{ij}u(t_j)+\theta_j)$=$\sum c_i\sigma(\sum_{j=1}^m \xi_{ij}x_j+\theta_j)$, for $u_x(t_i) = x_i$, which can be re-written as $\sum c_i\sigma(\xi\cdot x+\theta_i)$, where $\xi_i = (\xi_{i1},\ldots,\xi_{in})$, $x = (x_1,\ldots,x_n)$.

The previous argument shows that the results of approximation by neural networks in $\mathbf{R}^n$ can be viewed as a special case of our Theorem 2, when all functions in U are piecewise linear functions with n knots. However, in general U can be an arbitrary compact set in $C[a,b]$, which is much more complicated than those piecewise linear functions. It is dealing with these more general situations that constitutes the main contribution of this paper.

2.3 Proofs of Main Results

Prior to the proof, we first give a sketch as a road map.

For the case of L^p space,

Step 1: Find an $h > 0$ small enough, such that $\|\frac{1}{2h}\int_{-h}^{h} u(x+t)\,dt - u(x)\|_{L^p[a,b]}$ is uniformly small for all $u\in U$. If U is a convex compact set in $L^p[a,b]$, then for the fixed h, $U_h = \{u_h(x) : u_h(x) = \frac{1}{2h}\int_{-h}^{h} u(x+t)\,dt, u\in U\}$ is a convex compact set in both $L^p[a,b]$ and $C[a,b]$.

Step 2: On U_h, define a functional $\tilde{f}$, such that $|f(u)-\tilde{f}(u_h)|$ is uniformly small for all $u\in U$.

Step 3: Find an integer m and points $a = x_0 < x_1 < \ldots < x_m = b$, such that the piecewise linear interpolation $u_{h,m}$ of u_h at points x_i, $i = 0,1,\ldots,m$ satisfies

$\|u_{h,m} - u_h\|_{C[a,b]}$ being uniformly small for all $u_h \in U_h$.

Step 4: Define $\tilde{\tilde{f}}$ on $U_{h,m} = \{u_{h,m} : u_{h,m}$ is piecewise linear and $u_{h,m}(x_i) = u_h(x_i)\}$, such that $|\tilde{\tilde{f}}(u_{h,m}) - \tilde{f}(u_h)|$ is uniformly small for all $u_h \in U_h$.

Step 5: At this point, $\tilde{\tilde{f}}$ is a functional defined on $U_{h,m}$, which can be viewed as a space of dimension m, and the proof will be completed by using the result in finite dimensional case.

For the case of $C[a,b]$, we just modify the previous procedure.

The following lemmas will work out the details.

Lemma 1 *U is a compact set in $C[a,b]$, if and only if U is a closed set and*

(1) *All functions in U are* uniformly bounded, *i.e. for all $x \in [a,b]$, $|u(x)| \leq M$, for any $u \in U$.*

(2) *These functions are* equicontinuous, *i.e. for any $\epsilon > 0$, there exists $\delta > 0$, such that for every pair of points $x', x'' \in [a,b]$ and $|x' - x''| < \delta$, then $|u(x') - u(x'')| < \epsilon$ holds for every $u \in U$.*

U is a compact set in $L^p[a,b]$ $(1 < p < \infty)$, if and only if U is a closed set and

(1) *There is a constant M such that $\|u\|_{L^p[a,b]} \leq M$, for all $u \in U$;*

(2) *For any $\epsilon > 0$, there exists an $h_0 > 0$ such that if $h < h_0$, then*

$$\|\frac{1}{2h}\int_{-h}^{h} u(x+t)\,dt - u(x)\|_{L^p[a,b]} < \epsilon \tag{12}$$

holds for all $u \in U$, where $u(x) = 0$ if $x \notin [a,b]$. $\square$

Proof. See [12]. $\square$

Let U^c be the *convex hull* of U, that is, $U^c = \text{closure}\{u : u = \sum_{i=1}^k \lambda_i u_i,\ u_i \in U,\ 0 \le \lambda_i \le 1,\ \sum_{i=1}^k \lambda_i = 1\}$. Obviously, U^c is a compact set, whenever U is compact. On U^c we can define a continuous functional, which is an extension of f. Therefore, from the very beginning, we will assume that U is a compact *convex* set.

Because f is a continuous functional on a compact set U in $L^p[a,b]$, there exists a $\delta > 0$ such that if $\|u_1 - u_2\|_{L^p[a,b]} < \delta$, then $|f(u_1) - f(u_2)| < \epsilon/6$. By taking $\delta = \min(\epsilon/6, \delta)$, it can be assumed that $\delta \le \epsilon/6$.

For the fixed δ mentioned above, by Lemma 1, there exists an $h > 0$ such that

$$\|\frac{1}{2h}\int_{-h}^{h} u(x+t)\,dt - u(x)\|_{L^p[a,b]} < \delta/2 , \qquad \forall u \in U . \tag{13}$$

Define $U_h = \{u_h(x) : u_h(x) = \frac{1}{2h}\int_{-h}^{h} u(x+t)\,dt, u \in U\}$. It is easy to verify that U_h is also a convex compact set in $L^p[a,b]$, whenever U is convex and compact.

On U_h, define a new functional $\tilde{f}$ as follows: for all $u_h \in U_h$,

$$\tilde{f}(u_h) = f(v) + \|u_h - v\|_{L^p[a,b]} , \tag{14}$$

where v is the unique function in U such that

$$\|u_h - v\|_{L^p[a,b]} = \min_{w \in U} \|u_h - w\|_{L^p[a,b]} . \tag{15}$$

Lemma 2 *Functional $\tilde{f}$ defined by equation (14) makes sense for every $u_h \in U_h$, and is uniformly continuous on U_h.* $\square$

Proof. It is well known that $L^p[a,b]$ $(1 < p < \infty)$ is a strictly convex space, i.e. for any $g_1 \neq kg_2$, $0 \leq \lambda_1, \lambda_2 \leq 1$, $\lambda_1 + \lambda_2 = 1$, then $\|\lambda_1 g_1 + \lambda_2 g_2\|_{L^p[a,b]} < \lambda_1 \|g_1\|_{L^p[a,b]} + \lambda_2 \|g_2\|_{L^p[a,b]}$. Moreover, if U is a convex compact set in $L^p[a,b]$, then for $u_h \in U_h$, there is a unique v such that

$$\|u_h - v\|_{L^p[a,b]} = \min_{w \in U} \|u_h - w\|_{L^p[a,b]} \ , \tag{16}$$

due to the fact $L^p[a,b]$ $(1 < p < \infty)$ is strictly convex, which indicates that $\tilde{f}$ makes sense on U_h.

Now, we will prove the continuity of $\tilde{f}$. Suppose $u_{h,n}$ and $u \in U_h$, $\lim_{n\to\infty} \|u_{h,n} - u\|_{L^p[a,b]} = 0$, then we claim that the corresponding v_n and v defined in (15) satisfy $\lim_{n\to\infty} \|v_n - v\| = 0$. Otherwise, there is a subsequence of v_n, say $\{v_{n_k}\}$, which converges to some $v_1 \in L^p[a,b]$, and $v_1 \neq v$. Then, by the definition of v in (15), $\|u_h - v\|_{L^p[a,b]} < \|u_h - v_1\|_{L^p[a,b]}$. Therefore, there exists $\epsilon > 0$, such that

$$\|u_h - v\|_{L_p[a,b]} + 2\epsilon < \|u_h - v_1\|_{L^p[a,b]} - \epsilon \ . \tag{17}$$

However, $\|u_{h,n} - u_h\|_{L^p[a,b]} \to 0$, $\|v_{n_k} - v_1\|_{L^p[a,b]} \to 0$. Therefore, for sufficiently large k, we have

$$\begin{aligned} \|u_{h,n_k} - v\|_{L^p[a,b]} + \epsilon &< \|u_h - v\|_{L^p[a,b]} + 2\epsilon \\ &< \|u_h - v_1\|_{L^p[a,b]} - \epsilon \\ &< \|u_{h,n_k} - v_{n_k}\|_{L^p[a,b]} + \epsilon \ , \end{aligned} \tag{18}$$

which means $\|u_{h,n_k} - v\|_{L^p[a,b]} < \|u_{h,n_k} - v_{n_k}\|_{L^p[a,b]}$, contradicting the definition of v_{n_k} by (15). Therefore, the map $u_h \to v$ defined by (15) is a continuous map from U_h

to U. Furthermore, if $u_{h,n} \to u_h$, then $v_n \to v$, $f(v_n) \to f(v)$, $\|u_{h,n} - v_n\|_{L^p[a,b]} \to \|u_h - v\|_{L^p[a,b]}$, i.e. $\tilde{f}$ is a continuous functional defined on U_h. Since U_h is compact, $\tilde{f}$ is uniformly continuous. Lemma 2 is proved. $\square$

We now estimate $\tilde{f}(u_h) - f(u) = f(v) - f(u) + \|u_h - v\|_{L^p[a,b]}$. According to the definition of u_h and v, $\|u_h - u\|_{L^p[a,b]} < \delta/2$, $\|u_h - v\|_{L^p[a,b]} \le \|u_h - u\|_{L^p[a,b]} < \delta/2$. Consequently, $\|u - v\|_{L^p[a,b]} \le \|u - u_h\|_{L^p[a,b]} + \|u_h - v\|_{L^p[a,b]} < \delta \le \epsilon/6$, which implies $|f(u) - f(v)| < \epsilon/6$. Thus

$$|f(u) - \tilde{f}(u_h)| \le |f(u) - f(v)| + \|u_h - v\|_{L^p[a,b]} < \epsilon/3 \ . \tag{19}$$

From now on, instead of f, we will discuss the functional $\tilde{f}$ defined on U_h.

For fixed h, we claim that U_h is a *uniformly bounded* and *equicontinuous* set in $C[a,b]$. In fact,

$$\begin{aligned} |u_h(x)| &\le \frac{1}{2h}\int_{-h}^{h} |u(x+t)|\,dt \le (\frac{1}{2h}\int_{-h}^{h} |u(x+t)|^p\,dt)^{1/p} \\ &\le (\frac{1}{2h})^{1/p}(\int_a^b |u(x)|^p\,dt)^{1/p} \le (\frac{1}{2h})^{1/p}\ M \end{aligned} \tag{20}$$

The second inequality comes from the Jensen Inequality, and the last inequality comes from Lemma 1. Moreover,

$$\begin{aligned} &|u_h(x') - u_h(x'')| = |\frac{1}{2h}\int_{x'-h}^{x'+h} u(t)\,dt - \frac{1}{2h}\int_{x''-h}^{x''+h} u(t)\,dt| \\ \le\ & |\frac{1}{2h}\int_{x'-h}^{x''-h} u(t)\,dt| + |\frac{1}{2h}\int_{x'+h}^{x''+h} u(t)\,dt| \\ \le\ & \frac{1}{2h}(\int_{x'-h}^{x''-h} |u(t)|^p\,dt)^{1/p}|x''-x'|^{1/q} + \frac{1}{2h}(\int_{x'+h}^{x''+h} |u(t)|^p\,dt)^{1/p}\,|x''-x'|^{1/q} \\ \le\ & \frac{M}{h}|x''-x'|^{1/q}\ , \end{aligned} \tag{21}$$

where $(1/p) + (1/q) = 1$, and the second inequality comes from Hölder's Inequality. Therefore, U_h is *uniformly bounded* and *equicontinuous*; consequently, U_h is a *compact* set in $C[a,b]$ as well as in $L^p[a,b]$. Now, for any $\epsilon > 0$, there exist an integer m, and $m+1$ points $a = x_0 < x_1 < \cdots < x_m = b$, $x_j = a + j(b-a)/m$, $j = 0, 1, \ldots, m$, such that for all $u_h \in U_h$, if $|x' - x''| < \frac{b-a}{m}$, then $|u_h(x') - u_h(x'')| < \epsilon$.

For the fixed h and m, associated with all $u_h \in U_h$, we define a function

$$u_{h,m}(x) = u_h(x_j) + \frac{u_h(x_{j+1}) - u_h(x_j)}{x_{j+1} - x_j}(x - x_j), \quad x_j \leq x \leq x_{j+1}, \quad j = 0, 1, \ldots, m-1, \tag{22}$$

that is, $u_{h,m}(x)$ is piecewise linear and interpolates $u_h(x)$ at points x_j, $j = 0, 1, \ldots, m$. (It is possible that there might be several $u_h \in U_h$ corresponding to one $u_{h,m}(x)$).

For fixed h and m, let $U_{h,m} = \{u_{h,m}, u \in U\}$, then it is clear that $U_{h,m}$ is also a convex compact set in $L^p[a,b]$ as well as in $C[a,b]$.

Similar to the arguments made previously, on $U_{h,m}$, we define a new functional

$$\tilde{\tilde{f}}(u_{h,m}) = f(v_h) + \|u_{h,m} - v_h\|_{L^p[a,b]} \ , \tag{23}$$

where v_h is such that for all $u_{h,m} \in U_{h,m}$,

$$\|u_{h,m} - v_h\|_{L^p[a,b]} = \min_{w_h \in U_h} \|u_{h,m} - w_h\|_{L^p[a,b]} \ . \tag{24}$$

Similar to the proof of Lemma 2, we can prove

Lemma 3 *The functional $\tilde{\tilde{f}}$ defined on $U_{h,m}$ by (23) is continuous. Moreover, $|\tilde{\tilde{f}}(u_{h,m}) - \tilde{f}(u_h)| < \epsilon/3$ holds for all $u_h \in U_h$.* □

We need one more lemma.

Lemma 4 *For fixed h and m, the set $S_{h,m} = \{u_h(x_0), \ldots, u_h(x_m), u \in U\}$ is a compact set in $\mathbf{R}^{m+1}$.* $\square$

Proof. The boundedness of $S_{h,m}$ is a direct consequence of the boundedness of U_h in $L^p[a,b]$. Next, if $(u_{k,h}(x_0), \ldots, u_{k,h}(x_m))$, $u_k \in U$, $k = 1, 2, \ldots$, converges to $(u_0, \ldots, u_m)$, then there exists a subsequence u_{k_j} which converges to some $u \in U$, since U is compact. Therefore, $(u_{k_{j,h}}(x_0), \ldots, u_{k_{j,h}}(x_m))$ converges to $(u_h(x_0), \ldots, u_h(x_m))$, which implies that $(u_0, \ldots, u_m) = (u_h(x_0), \ldots, u_h(x_m))$. Thus $S_{h,m}$ is compact. $\square$

Having established these Lemmas, we now proceed to prove Theorem 1.

Proof of Theorem 1.

According to Lemmas 1-4 and previous arguments, for any $\epsilon > 0$, there exist $h > 0$, and a functional $\tilde{f}$ on U_h, such that $|f(u) - \tilde{f}(u_h)| < \epsilon/3$ for all $u \in U$. For the fixed h, there exist an integer m and a functional $\tilde{\tilde{f}}$, such that $|\tilde{f}(u_h) - \tilde{\tilde{f}}(u_{h,m})| < \epsilon/3$ for all $u \in U$.

Now, define a function g on $S_{h,m}$, by

$$g(u_{h,m}(x_0), \ldots, u_{h,m}(x_m)) = \tilde{\tilde{f}}(u_{h,m}) \ . \tag{25}$$

Because $u_{h,m}$ is piecewise linear, the fact that $(u^k_{h,m}(x_0), \ldots, u^k_{h,m}(x_m))$ converges to $(u_{h,m}(x_0), \ldots, u_{h,m}(x_m))$ in $S_{h,m}$ implies $\|u^k_{h,m} - u_{h,m}\|_{L^p[a,b]} \to 0$, thus $\tilde{\tilde{f}}(u^k_{h,m}) \to f(u_{h,m})$, as $k \to \infty$, which means g is continuous on $S_{h,m}$.

By the well-known approximation theorem (see [3], [9]), there exist a positive integer N and constants c_i, θ_i, $\xi_{i,j}$, $i = 1, \ldots, N$, $j = 0, 1, \ldots, m$, such that

$$|g(u_{h,m}(x_0), \ldots, u_{h,m}(x_m)) - \sum_{i=1}^{N} c_i \, \sigma(\sum_{j=0}^{m} \xi_{i,j} \, u_{h,m}(x_j) + \theta_i)| < \epsilon/3 \; . \tag{26}$$

Summing up, we obtain

$$|f(u) - \sum_{i=1}^{N} c_i \, \sigma(\sum_{j=0}^{m} \xi_{i,j} \, \frac{1}{2h} \int_{x_j-h}^{x_j+h} u(t) \, dt + \theta_i)| < \epsilon \; , \tag{27}$$

or

$$|f(u) - \sum_{i=1}^{N} c_i \, \sigma(\bar{\xi}_i \cdot \bar{u}_h + \theta_i)| < \epsilon \; , \tag{28}$$

where $\bar{\xi}_i$ is the vector $(\xi_{i,0}, \ldots, \xi_{i,m})$, $\bar{u}_h = (\frac{1}{2h} \int_{x_0-h}^{x_0+h} u(t) \, dt, \ldots, \frac{1}{2h} \int_{x_m-h}^{x_m+h} u(t) \, dt)$. The proof of Theorem 1 is thus complete. $\square$

In order to prove Theorem 2, we need to modify the *definition process* of the functional $\tilde{f}$, because $C[a,b]$ is *not* strictly convex.

Since U is a compact set in $C[a,b]$, for all $\delta > 0$, there is a positive integer m, such that for all $u \in U$, for all $x', x'' \in [a,b]$, if $|x' - x''| < \frac{b-a}{m}$, then $|u(x') - u(x'')| < \delta/2$. Let $x_j = a + j(b-a)/m$, $j = 0, 1, \ldots, m$, for this fixed m, we define the function

$$u_m(x) = u(x_j) + \frac{u(x_{j+1}) - u(x_j)}{x_{j+1} - x_j} (x - x_j) \; , \qquad x_j \leq x \leq x_{j+1} \; , \quad j = 0, 1, \ldots, m-1 \; . \tag{29}$$

It is clear that $\|u(x) - u_m(x)\|_{C[a,b]} < \delta$.

Let $U_m = \{u_m, u \in U\}$, and on U_m we define a functional $\tilde{f}$ by

$$\tilde{f}(u_m) = f(v) + \|u_m - v\|_{C[a,b]} \; , \tag{30}$$

where v is determined by

$$\|u_m - v\|_{L^p[a,b]} = \min_{w \in U} \|u_m - w\|_{L^p[a,b]} \; , \tag{31}$$

(here v is the nearest function in U from u_m, in $L^p[a,b]$, not in $C[a,b]$).

Lemma 5 *Suppose that U is a compact set in $C[a,b]$, $u \in U$, $u_n \in U$, $n = 1, 2, \ldots$, then $\|u - u_n\|_{C[a,b]} \to 0$ if and only if $\|u_n - u\|_{L^p[a,b]} \to 0$.* □

Proof. It is obvious that $\|u - u_n\|_{C[a,b]} \to 0$ implies $\|u - u_n\|_{L^p[a,b]} \to 0$. Now, assuming $\|u - u_n\|_{L^p[a,b]} \to 0$, we will prove $\|u - u_n\|_{C[a,b]} \to 0$. Suppose that $\|u - u_n\|_{C[a,b]} \not\to 0$, then there exist two subsequences $u_{n_{k_1}}$ and $u_{n_{k_2}}$ which converge to v_1, v_2 in $C[a,b]$, respectively, since U is compact in $C[a,b]$. Thus, $\|u_{n_{k_1}} - v_1\|_{L^p[a,b]} \to 0$, $\|u_{n_{k_2}} - v_2\|_{L^p[a,b]} \to 0$, which, combined with the assumption that $\|u - u_n\|_{L^p[a,b]} \to 0$, leads to the fact $u = v_1 = v_2$ a.e., a contradiction. Thus, the proof of Lemma 5 is complete. □

Lemma 6 *Functional $\tilde{f}$ defined by (30) is continuous on U_m.* □

Proof. If $u_m \in U_m$ and $u_{k,m} \in U_m$, $k = 1, 2, \ldots$ and $\|u_{k,m} - u_m\|_{C[a,b]} \to 0$, then $\|u_{k,m} - u_m\|_{L^p[a,b]} \to 0$. Consequently, v_k, $v \in U$ determined by (31) corresponding to $u_{k,m}$, u_m respectively, satisfy $\|v_k - v\|_{L^p[a,b]} \to 0$. By Lemma 5, $\|v_k - v\|_{C[a,b]} \to 0$. Because U_m is a compact set in $C[a,b]$, we have for all $\delta > 0$, there is an $\eta > 0$ such that for all $u_m', u_m'' \in U_m$, $\|u_m' - u_m''\|_{C[a,b]} < \eta$ implies $\|v' - v''\|_{C[a,b]} < \delta$. By the arguments similar to those used in Lemma 5, we conclude that $|\tilde{f}(u_m) - f(v)| < \epsilon/2$, for all $u_m \in U_m$. □

Lemma 7 *The set $S_m = \{u_m(x_0), \ldots, u_m(x_m), u \in U\}$ is a compact set in $\mathbf{R}^{m+1}$.* □

Proof. The proof is the same as that of Lemma 4. □

Having established these lemmas, we proceed to prove Theorem 2.

Proof of Theorem 2.

Similar to the proof of Theorem 1, we define a function on S_m by

$$g(u(x_0), \ldots, u(x_m)) = \tilde{f}(u_m) \ . \tag{32}$$

Because $u_m(x)$ is piecewise linear, the fact that $(u_k(x_0), \ldots, u_k(x_m))$ converges to $(u(x_0), \ldots, u(x_m))$ in $\mathbf{R}^{m+1}$ implies $\|u_{k,m}(x) - u_m(x)\|_{C[a,b]} \to 0$, which means g is a continuous function on S_m. Therefore, for any $\epsilon > 0$, there exist N, c_i, θ_i, $\xi_{i,j}$, $i = 1, \ldots, N$, $j = 0, 1, \ldots, m$, such that

$$|g(u_m(x_0), \ldots, u_m(x_m)) - \sum_{i=1}^{N} c_i\, \sigma(\sum_{j=0}^{m} \xi_{i,j}\, u(x_j) + \theta_i)| < \epsilon/2 \ . \tag{33}$$

Combining it with $|f(u) - \tilde{f}(u_m)| < \epsilon/2$, we obtain

$$|f(u) - \sum_{i=1}^{N} c_i\, \sigma(\sum_{j=0}^{m} \xi_{i,j}\, u(x_j) + \theta_i)| < \epsilon \ , \tag{34}$$

or

$$|f(u) - \sum_{i=1}^{N} c_i\, \sigma(\bar{\xi}_i \cdot \bar{u}_m + \theta_i)| < \epsilon \ , \tag{35}$$

where $\bar{\xi}_i = (\xi_{i,0}, \ldots, \xi_{i,m})$, $\bar{u}_m = (u(x_0), \ldots, u(x_m))$. Therefore, the proof of Theorem 2 is complete. □

The proofs of Theorems 3 and 4 proceed in the similar line to those in the proofs of Theorems 1 and 2. The only places that need changing are:

(1) Instead of the linear interpolant appearing in (22) and (29), we use multilinear interpolant;

(2) Instead of directly using Cybenko's Theorem, we use (1), which is obtained from [9].

Remark 4. In [9], we pointed out that for any function (continuous or discontinuous) $\omega(x)$, if the linear combinations $\sum c_i\,\omega(\eta_i\,x + \phi_i)$, where $c_i,\ \eta_i,\ \phi_i \in \mathbf{R}$, are dense in any $C[a,b]$, then the linear combinations $\sum c_i\,\omega(\xi_i \cdot x + \theta_i)$ are dense in $C(\mathbf{K})$, where $c_i,\ \theta_i, \in \mathbf{R},\ \xi_i,\ x \in \mathbf{R}^n$ and $\mathbf{K}$ is a compact set in $\mathbf{R}^n$. Combining this fact with the proof of our theorems in this paper, we conclude that if $\sum c_i\omega(\eta_i\,x + \phi_i)$ are dense in any $C[a,b]$, then all our theorems remain valid, if the sigmoidal function σ is replaced by ω. Therefore, *in order to approximate nonlinear functionals using neural networks with a nonlinear activation function ω, what we need to do is to prove the denseness of $\sum c_i\omega(\eta_i\,x + \phi_i)$ in any $C[a,b]$.*

Among them there exist many functions that occur in approximation theory. For example, Schoenberg Cardinal Splines, B-Splines, wavelets, etc., all satisfy the conditions imposed on ω. Consequently any of them can be used to replace the generalized sigmoidal functions in our previous theorems. The details are not elaborated here.

3 Application to Dynamical Systems

As an application, we discuss approximation of dynamical systems. Sandberg [11] made important contribution to approximation of discrete-time dynamical systems. Here, we consider approximation of dynamical systems as approximation of continuous functionals defined on a compact set (of input functions). By doing so, we are able to significantly generalize Sandberg's result and provide a uniform viewpoint and treatment to both continuous-time and discrete-time systems. The main results obtained earlier in this paper can be readily applied to the approximation of the output of dynamical systems at any particular time[3].

First, we introduce some notations and definitions, which come basically from [11] and [10] with slight variation.

Suppose that X_1 (or X_2) stands for the set of $\mathbf{R}^{q_1}$-valued (or $\mathbf{R}^{q_2}$-valued) functions defined in $\mathbf{R}^n$.

A dynamical system G can be viewed as a map from X_1 to X_2, that is, for all $u \in X_1$, $Gu = v \in X_2$.

Let $x \in X$, define a "windowing" operator W by

$$(W_{\alpha,a}x)(\beta) = \begin{cases} x(\beta) & \text{if } \beta \in \Gamma_{\alpha,a} \ , \\ 0 & \text{if } \beta \notin \Gamma_{\alpha,a} \ , \end{cases} \tag{36}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{R}^n$, $\Gamma_{\alpha,a} = \{r = (r_1, \ldots, r_n) \in \mathbf{R}^n, |r_j - \alpha_j| \leq a, \forall j = 1, \ldots, n\}$, that is, $W_{\alpha,a}$ is a "windowed version" of x with the (n-dimensional) window centered at α and its width $2a$.

If U is a nonempty set in X_1, define $U_{\alpha,a} = \{u|_{\Gamma_{\alpha,a}}, u \in U\}$, where $u|_{\Gamma_{\alpha,a}}$ is the restriction of u to $\Gamma_{\alpha,a}$, that is, $u|_{\Gamma_{\alpha,a}} = W_{\alpha,a}u$.

A map G from X_1 to X_2 is said to be of *approximately finite memory*, if for all

[3]We are not yet approximating the whole system transfer function at this time.

$\epsilon > 0$, there is an $a > 0$ such that

$$|(Gu)_j(\alpha) - (GW_{\alpha,a}u)_j(\alpha)| < \epsilon\ , \qquad j = 1, \ldots, q_2 \tag{37}$$

holds for any $\alpha \in \mathbf{R}^n$, $u \in U$.

For each $\beta \in \mathbf{R}^n$, define T_β: $X_1 \to X_1$ to be the (shift) operator given by $(T_\beta x)(\alpha) = x(\alpha - \beta)$ for all $\alpha \in \mathbf{R}^n$. A map from $X_1 \to X_2$ is *shift invariant* if $(GT_\beta u)(\alpha) = (Gu)(\alpha - \beta)$ for any pair (α, β), $\alpha \in \mathbf{R}^n$, $\beta \in \mathbf{R}^n$, $u \in X_1$.

We assume that the entire set U (the domain of G, in which we deal with the approximation problem) satisfies

(1) If $u \in U$, then $u|_{\Gamma_{\alpha,a}} \in U$ for any $\alpha \in \mathbf{R}^n, a > 0$;

(2) for all $\alpha \in \mathbf{R}^n$, $a > 0$, $U_{\alpha,a}$ is a compact set in $C_V(\prod_{k=1}^n [\alpha_k - a_k, \alpha_k + a_k])$ or a compact set in $L_V^p(\prod_{k=1}^n [\alpha_k - a_k, \alpha_k + a_k])$, where V stands for $\mathbf{R}^{q_1}$;

(3) Let $(Gu)(\alpha) = ((Gu)_1(\alpha), \ldots, (Gu)_{q_2}(\alpha))$, then each $(Gu)_j(\alpha)$ is a continuous functional defined over $U_{\alpha,a}$, with the corresponding topology in $C_V(\prod_{k=1}^n [\alpha_k - a_k, \alpha_k + a_k])$ or $L_V^p(\prod_{k=1}^n [\alpha_k - a_k, \alpha_k + a_k])$.

Theorem 5 *If U and G satisfy all the assumptions* (1)-(3) *made above, and G is of approximately finite memory, then for any $\epsilon > 0$, there exist $a > 0$, a positive integer m, $(m+1)^n$ points in $\prod_{k=1}^n [\alpha_k - a_k, \alpha_k + a_k]$, a positive integer N, constants $c_i(G, \alpha, a)$ depending on G, α, a only, and $q_2 \times (m+1)^n$-vectors $\bar{\xi}_i$, $i = 1, \ldots, N$, such that*

$$|(Gu)_j(\alpha) - \sum_{i=1}^{N} c_i(G, \alpha, a)\, \sigma(\bar{\xi}_i \cdot \bar{u}_{q_1,n,m} + \theta_i)| < \epsilon\ , \qquad j = 1, 2, \ldots, q_2\ , \tag{38}$$

or

$$|(Gu)_j(\alpha) - \sum_{i=1}^{N} c_i(G,\alpha,a)\,\sigma(\bar{\xi}_i \cdot \bar{u}^*_{q_1,n,m} + \theta_i)| < \epsilon\ , \qquad j = 1,2,\ldots,q_2\ , \tag{39}$$

where $\bar{u}_{q_1,n,m}$ *and* $\bar{u}^*_{q_1,n,m}$ *are the same vectors as defined in Theorem* 3 *and Theorem* 4*, and* $\sigma(x)$ *is any bounded generalized sigmoidal function.* □

Proof. Because G is of approximately finite memory, for all $u \in U$ we can find $W_{\alpha,a}u \in U_{\alpha,a}$ such that

$$|(Gu)_j(\alpha) - (GW_{\alpha,a}u)_j(\alpha)| < \epsilon\ , \qquad j = 1,2,\ldots,q_2\ . \tag{40}$$

Applying Theorems 3 and 4 to $GW_{\alpha,a}$ yields the desired result. □

Remark 5. Let N and N^+ denote $\{0,1,\ldots\}$ and $\{1,2,\ldots\}$ respectively. Let S denote the metric space of all maps from N (or N^+) to the compact set E of $\mathbf{R}^p$ with the metric given by

$$\rho(s_a, s_b) = \sup_{t\in N} \|s_a(t) - s_b(t)\|\ , \tag{41}$$

and R denotes the collection of all $\mathbf{R}$-valued maps defined on N. In [11], under the assumption (denoted as *A.1*) that $G : S \to R$ is causal, time-invariant and of approximately-finite memory, $\sigma(x)$ is a continuous generalized sigmoidal function, and $G(\cdot)(t) : S \to R$ is continuous for each $t \in N^+$, the following interesting theorem was obtained.

Theorem A (Sandberg[11]) *For* $a \in N^+$*, let* $T_a : S \to \mathbf{R}^{p(a+1)}$ *be defined by*

$$(T_a s)(t) = [s(t), s(t-1), \ldots, s(t-a)]^{\mathrm{tr}}\ , \qquad s \in S\ , \tag{42}$$

where ${}^{\mathrm{tr}}$ *denotes transpose, and if the condition* A.1 *above is satisfied, then for any* $\epsilon > 0$*, there exist* m *and* $a \in N^+$*, real numbers* $\kappa_1, \ldots, \kappa_m$*,* $\rho_1, \ldots, \rho_m$ *and real row vectors* $\eta_1, \ldots, \eta_m$ *of order* $p(a+1)$*, such that*

$$|(Gs(t)) - \sum_{l=1}^{m} \kappa_l \, \sigma[\eta_l (T_a s)(t) + \rho_l]| < \epsilon \ , \qquad t \in N \tag{43}$$

for all $s \in S$. □

We point out that Theorem A is the discrete case of our Theorem 2. In fact, let

$$(W_{t,a}s)(\tau) = \begin{cases} s(\tau) & \text{if} \quad t - a \leq \tau \leq t \ , \\ 0 & \quad \text{otherwise,} \end{cases} \tag{44}$$

then $G(W_{t,a}s)(t)$ can be viewed as a continuous functional defined on the compact set $[s(t), \ldots, s(t-a)]^{\mathrm{tr}}$, hence Theorem A can be obtained by the approximately-finite memory property and Theorem 2. That is, our result generalizes Sandberg's Theorem A.

A graphical representation [11] of the approximation to $(Gs)(t)$ is shown in Fig. 1.

To illustrate the effectiveness of our theorems, we give some more examples.

Example 1:

Suppose that the input $u(x)$ and output $s(x) = G(u(x))$ of a nonlinear system G, are subject to the following differential equation

$$\frac{d}{dx} s(x) = g(s(x), u(x), x), \qquad s(a) = s_0 \tag{45}$$

where $g(v, w, x)$ satisfies Lipschitz condition with respect to variables v and w, i.e. there is a constant $c > 0$ such that

$$|g(v, w, x) - g(v', w, x)| \leq c|v - v'| \tag{46}$$

$$|g(v,w,x) - g(v,w',x)| \le c|w - w'| \tag{47}$$

Moreover, we assume that the differential equation has a unique solution for any $u(x) \in C[a,b]$.

Under these assumptions, we have

$$(Gu)(x) = s_0 + \int_a^x g((Gu)(t)), u(t), t)\, dt \tag{48}$$

If we are given two inputs, $u_1(x)$ and $u_2(x)$, then for any fixed $d \in [a,b]$, we have

$$\begin{aligned}
&|(Gu_1)(d) - (Gu_2)(d)| \\
&\quad \le \int_a^d |g((Gu_1)(t), u_1(t), t) - g((Gu_2)(t), u_2(t), t)|\, dt \\
&\quad \le c\int_a^d |(Gu_1)(t) - (Gu_2)(t)|\, dt + c\int_a^d |u_1(t) - u_2(t)|\, dt \\
&\quad \le c\int_a^d |u_1(t) - u_2(t)|\, dt + c\int_a^d \int_a^v |u_1(t) - u_2(t)| e^{t-v}\, dt\, dv
\end{aligned} \tag{49}$$

The last inequality comes from generalized Gronwall Inequality. From Inequality (49), we conclude that $(Gu)(d)$ is a continuous functional defined on $C[a,b]$. If the input set U is a compact set in $C[a,b]$ (also in $C[a,d]$), then our previous Theorem 2 shows that the output of the nonlinear system G at a specified time d can be approximated by

$$\sum_{i=1}^{N} c_i \sigma(\sum_{j=1}^{m} \xi_{ij} u(x_i) + \theta_i) \tag{50}$$

where $a = x_0 < x_1 < \cdots < x_m = d$.

Example 2:

For a Boolean function $\{-1,1\}^n \to \{-1,1\}$, we define

$$U = \{u : u(x) \;=\; u(j) + [u(j+1) - u(j)](x - j), \text{ if } j \le x \le j+1,$$

$$u(j) = 1 \text{ or } -1, \ j = 1, \ldots, n-1; \quad u(n) = 1 \text{ or } -1\} \tag{51}$$

which is a compact set in $C[1,n]$. Moreover, there is a one-to-one correspondence between $\{-1,1\}^n$ and U, and every Boolean function $b \in B$ corresponds to a (continuous) functional defined on U, thus b can be approximated by

$$\sum_{i=1}^{N} c_i \sigma(\sum_{j=1}^{n} \xi_{ij} u(j) + \theta_i) \ . \tag{52}$$

In this case, we can take $N = 2^n$ neurons with their 2^n weights being $(\xi_{i1}, \ldots, \xi_{in})$ $= (k\eta_{i1}, \ldots, k\eta_{in})$, where k is a positive real number depending on the threshold value, $(\eta_{i1}, \ldots, \eta_{in})$, $i = 1, 2, \ldots, 2^n$ are the 2^n distinct elements in $\{-1,1\}^n$, which means that we can use 2^n neurons to identify a Boolean function $\{-1,1\}^n \to \{-1,1\}$, as illustrated in Fig. 2.

4 Conclusion

In this paper, several strong approximation theorems concerning neural network representation capability have been obtained. The "generalized sigmoidal" basis functions that exist in many (nonlinear) neural networks are used to approximate continuous functionals defined on spaces $C[a,b]$ or $L^p[a,b]$ $(1 < p < \infty)$. As discussed, many other functions, instead of sigmoidal functions, can also be used without the loss of the validity of the results in this paper. Some application to dynamical systems has been reported, including showing that an earlier result in [11] can be obtained from our theorems. Applying the method used in this paper, we can also discuss approximations of continuous functionals in more general topological spaces, which will be reported later.

Acknowledgements. The authors wish to express their gratefulness to the anonymous reviewers for their valuable comments and suggestions on revising this paper. They also wish to thank Prof. R.-W. Liu of University of Notre Dame, for bringing some of the papers in this area to their attention.

References

[1] A. Wieland and R. Leighten, "Geometric analysis of neural network capacity", *IEEE First ICNN. 1*, pp. 385-392 (1987).

[2] B. Irie and S. Miyake, "Capacity of three-layered perceptrons", *IEEE ICNN 1*, pp. 641-648, (1988).

[3] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of Control, Signals and Systems*, Vol. 2, No. 4, pp. 303-314 (1989).

[4] S. M. Carroll and B. W. Dickinson, "Construction of neural nets using Radon transform", *IJCNN Proc. I*, pp. 607-611 (1989).

[5] K. Hornik, M. Stichcombe and H. White, "Multi-layer feedforward networks are universal approximators", *Neural Networks*, Vol. 2, pp. 359-366 (1989).

[6] K. Hornik, "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, Vol. 4, pp. 251-257 (1991).

[7] V. Y. Kreinovich, "Arbitrary nonlinearity is sufficient to represent all functions by neural networks: a theorem", *Neural Networks*, Vol. 4, pp. 381-383 (1991).

[8] Yoshifusa Ito, "Representation of functions by superpositions of a step or sigmoidal function and their applications to neural network theory", *Neural Networks*, Vol. 4, pp. 385-394 (1991).

[9] Tianping Chen, Hong Chen and Ruey-wen Liu, "A constructive proof of Cybenko's approximation theorem and its extensions", pp. 163 - 168 in *Computing Science and Statistics* (editors LePage and Page), Proceedings of the 22nd Symposium on the Interface (East Lansing, Michigan, May 1990), Springer-Verlag ISBN 0-387-97719-8.

[10] I. W. Sandberg, "Representation theory and nonlinear systems", *Proc. Int. Conf. Integral Methods in Science and Engineering*, Arlington, TX, May (1990).

[11] I. W. Sandberg, "Approximation theorems for discrete-time systems", *IEEE Trans. on Circuits and Systems*, Vol. 38, No. 5, pp. 564-566, May (1991).

[12] I. P. Natanson, "Theory of functions of a real variable" (= Teoria functsiy veshchestvennoy peremennoy), translated from the Russian by Edwin Hewitt, Rev. ed., New York, (1961).

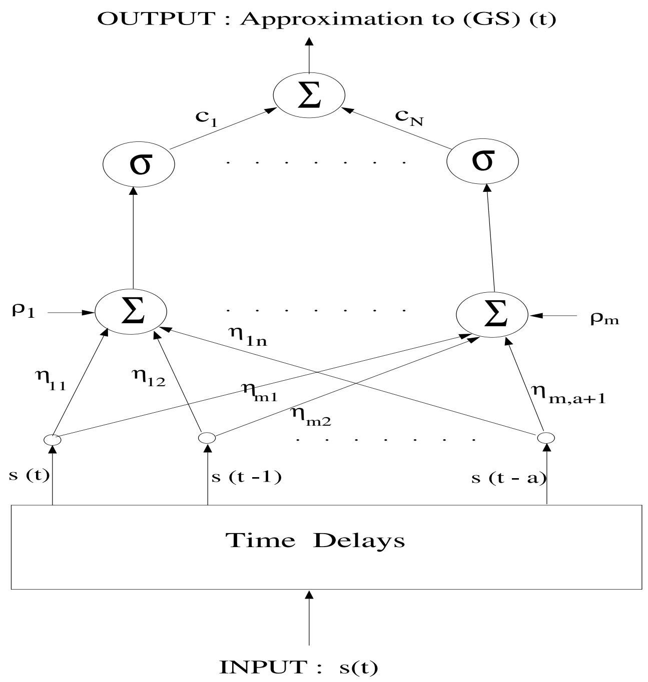

Figure 1: Graphical representation of approximation to $(Gs)(t)$

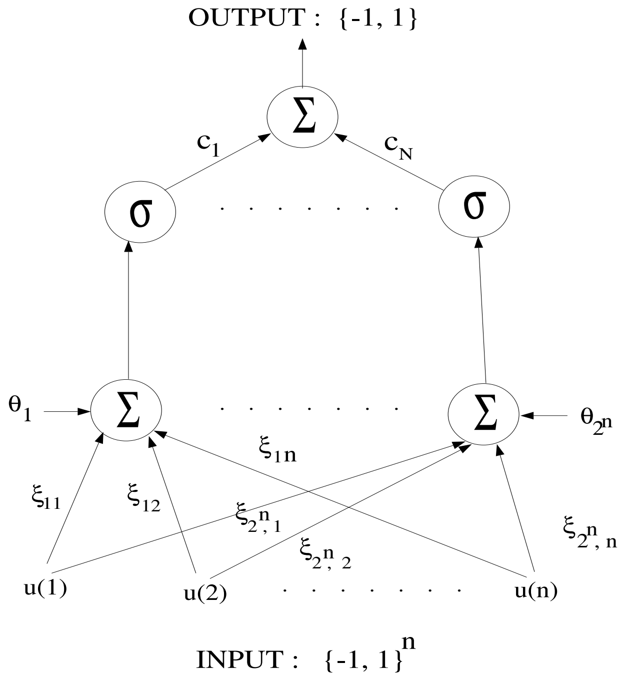

Figure 2: Identification of Boolean functions by a neural network