Accepted for Publication by IEEE Transactions on Neural Networks.

Approximation Capability to Functions of Several Variables, Nonlinear Functionals and Operators by Radial Basis Function Neural Networks

Tianping Chen[1] and Robert Chen[2]

Abstract

The purpose of this paper is to explore the representation capability of radial basis function (RBF) neural networks. The main results are: (1) The necessary and sufficient condition for a function of one variable to be qualified as an activation function in RBF network is that the function is not an even polynomial. (2) The capability of approximation to nonlinear functionals and operators by RBF networks is revealed, using sample data either in frequency domain or in time domain, which can be used in system identification by neural networks.

Key words and phrases: Approximation theory, neural networks, system identification, radial basis networks, functionals, operators, wavelets.

[1]The author is with the Department of Mathematics, Fudan University, Shanghai, P.R.China. His work has been supported in part by NSF of China, Shanghai NSF, the "Climbing" Project in China, Grant NSC 92092, and the Doctoral Program Foundation of Educational Commission in China.

[2]The author is with VLSI Libraries, Inc., 1836 Cabrillo Ave., Santa Clara, CA 95050, USA.

1 Introduction

There have been several recent works concerning the representation capabilities of multilayer feedforward neural networks. For example, in the past few years, several papers ([1]-[7] and many others) related to this topic have 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 $\mathbf{R}^n$. In many of those papers, sigmoidal functions need to be assumed to be continuous or monotone. Recently [9] [10], we pointed out that the *boundedness* of the sigmoidal function plays an essential role for its being an activation function in the hidden layer, *i.e.*, instead of continuity or monotonity, the boundedness of sigmoidal functions ensures the network capability.

In addition to sigmoidal functions, many other functions can be used as activation functions of neural networks. For example, in [12], we proved that for a function $g \in C(R^1) \cap S'(R^1)$ to be an activation function in feedforward neural networks, the necessary and sufficient condition is that the function is not a polynomial (also see [5]).

The above papers are all significant advances towards solving the problem of whether a function is qualified as an activation function in neural networks. However, they only dealt with affine-basis-function neural networks (ABF), also called multilayer perceptrons (MLP) and the goal there is to approximate continuous functions by the family

$$\sum_{i=1}^{N} c_i\, g(y_i \cdot x + \theta_i)$$

where $y_i \in \mathbf{R}^n$, $c_i, \theta_i \in \mathbf{R}^1$ and $y_i \cdot x$ denotes the inner product of y and x.

Among the various kinds of promising neural networks currently under active research, there is another type called *radial-basis-function (RBF) networks* [13] (also called *localized receptive field network* [16]), where the activation functions are radially symmetric and produce a localized response to input stimulus. For a survey, see, for example, [15]. A block diagram of an RBF network is shown in Fig. 1. One of the special basis functions that are commonly used is a Gaussian kernel function. Using the Gaussian basis function, RBF networks is capable of forming an arbitrarily close approximation to any continuous functions, as shown by [17] [18] [19].

More generally, the goal here is to approximate functions of a finite number of real variables by

$$\sum_{i=1}^{N} c_i\, g(\lambda_i \|x - x_i\|_{\mathbf{R}^n})$$

where $x \in \mathbf{R}^n$ and $\|x - x_i\|_{\mathbf{R}^n}$ is the distance between x and x_i in $\mathbf{R}^n$. Here, the activation function g is not necessarily Gaussian. In this direction, several results concerning RBF neural networks were obtained [13] [14]. In [14], Park and Sandberg proved the following important theorem:

Let $K : \mathbf{R}^n \to \mathbf{R}$ be a radial symmetric, integrable, bounded function such that K is continuous almost everywhere and $\int_{\mathbf{R}^n} K(x)dx \neq 0$, then the family

$$\sum_{i=1}^{N} c_i\, g\left(\frac{\|x - x_i\|_{\mathbf{R}^n}}{\sigma}\right)$$

is dense in $L^p(\mathbf{R}^n)$, where $g(\|x\|_{\mathbf{R}^n}) = K(x)$.

In [23], Park and Sandberg discussed several related results on the L^1, L^2 approximation. For example, they proved the following very interesting theorem:

Theorem A.

Assuming that K: $\mathbf{R}^n \to \mathbf{R}^1$ is a square-integrable function, then the family

$$\sum_{i=1}^{N} c_i\, g(\frac{\|x - x_i\|_{\mathbf{R}^n}}{\sigma})$$

is dense in $L^2(\mathbf{R}^n)$ if and only if K is pointable.

In [24], we generalized this result and proved

Theorem B.

Suppose that g: $\mathbf{R}^+ \to \mathbf{R}^1$, such that $g(\|x\|_{\mathbf{R}^n}) \in L^2(\mathbf{R}^n)$, then the family of functions

$$\sum_{i=1}^{N} c_i\, g(\frac{\|x - x_i\|_{\mathbf{R}^n}}{\sigma_i})$$

is dense in $L^2(\mathbf{R}^n)$.

It is now natural to ask the following questions: (1) What is the *necessary* and *sufficient* condition for a function to be qualified as an activation function in RBF neural networks? (2) How to approximate nonlinear *functionals* by RBF neural networks? (3) How to approximate nonlinear *operators* (eg. the output of a system) by RBF neural networks, using sample data in frequency (phase) domain or in time (state) domain?

The purpose of this paper is to give strong results in answering those questions.

This paper is organized as follows. In Section 2, we list our symbols, notations and review some definitions. In Section 3, we show that the necessary and sufficient condition for a continuous function in $S'(\mathbf{R}^1)$ to be qualified as an activation function in RBF networks is that it is not an even polynomial. In Section 4, we show the capability of RBF neural networks to approximate nonlinear functionals and operators on some Banach space as well as on some compact set in $C(K)$, where K is a compact set in any Banach space. Furthermore, we establish the capability of neural networks to approximate nonlinear *operators* from $C(K_1)$ to $C(K_2)$. Approximations using samples in both frequency domain and time domain are discussed. Examples are given, which includes the use of wavelet coefficients to the approximation. It is also pointed out that the main results in Section 4 can be used in computing the outputs of nonlinear dynamical systems, thus identifying the system. We conclude this paper with Section 5.

2 Notations and Definitions

We list here the main symbols and notations that will be used throughout this paper.

X: some Banach space with norm $\|\cdot\|_X$.

$\mathbf{R}^n$: Euclidean space of dimension n with norm $\|\cdot\|_{\mathbf{R}^n}$.

K: some compact set in a Banach space.

$C(K)$: Banach space of all continuous functions defined on K, with norm

$$\|f\|_{C(K)} = \max_{x \in K} |f(x)|$$

.

V: some compact set in $C(K)$.

$S(R^n)$: All Schwartz functions in distribution theory, i.e., all the infinitely differentiable functions, which are rapidly decreasing at infinity.

$S'(R^n)$: All the distributions defined on $S(R^n)$, i.e., all the linear continuous functionals defined on $S(R^n)$.

$C^\infty(R^n)$: All infinitely differentiable functions defined on R^n.

C_C^∞: all infinitely differentiable functions with compact support in R^n.

We review the following definitions.

Definition 1. A function $\sigma : \mathbf{R}^1 \to \mathbf{R}^1$ is called a (generalized) sigmoidal function, if it satisfies

$$\begin{cases} \lim_{x \to -\infty} \sigma(x) = 0 \ , \\ \lim_{x \to \infty} \sigma(x) = 1 \ . \end{cases} \tag{1}$$

Definition 2. Let X be a Banach space with norm $\|\cdot\|_X$. If there are elements $x_n \in X$, $n = 1, 2, \ldots$, such that for every $x \in X$ there is a unique real number sequence $a_n(x)$, such that

$$x = \sum_{n=1}^{\infty} a_n(x) x_n \ ,$$

where the series converges in X, then $\{x_n\}_{n=1}^{\infty}$ is called a Schauder basis in X and X is called a Banach space with Schauder basis.

Definition 3. Suppose that X is a Banach space, $V \subseteq X$ is called a compact set in X, if for every sequence $\{x_n\}_{n=1}^{\infty}$ with all $x_n \in V$, there is a subsequence $\{x_{n_k}\}$ which converges to some element $x \in V$. It is well known that if $V \subseteq X$ is a compact set in X, then for any $\delta > 0$, there is a δ-net $N(\delta) = \{x_1, \ldots, x_{n(\delta)}\}$, with all $x_i \in V$, $i = 1, \ldots, n(\delta)$, *i.e.* for every $x \in X$, there is some $x_i \in N(\delta)$ such that $\|x_i - x\|_X < \delta$.

3 Characteristics of Continuous Functions as Activation in RBF Networks

In this section, we show that for a continuous function to be qualified as an activation function in RBF networks, the *necessary* and *sufficient* condition is that it is not an even polynomial and we prove two approximation theorems by RBF networks.

More precisely, we prove

Theorem 1 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$, i.e., all those continuous functions such that $\int_{\mathbf{R}^1} g(x)s(x)\,dx$ makes sense for all $s \in S(\mathbf{R}^1)$, then the family*

$$\sum_{i=1}^{N} c_i g(\lambda_i \|x - y_i\|_{\mathbf{R}^n})$$

is dense in $C(K)$, if and only if g is not an even polynomial, where K is a compact set in $\mathbf{R}^n$, $y_i \in \mathbf{R}^n$, c_i, $\lambda_i \in \mathbf{R}^1$, $i = 1, \ldots, N$.

Theorem 2 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ and is not an even polynomial, K is a compact set in $\mathbf{R}^n$, V is a compact set in $C(K)$, then for any $\epsilon > 0$, there are* *a positive integer N, $\lambda_i \in \mathbf{R}^1$, $y_i \in \mathbf{R}^n$, $i = 1, \ldots, N$, which are all independent of $f \in V$ and constants $c_i(f)$ depending on f, $i = 1, \ldots, N$, such that*

$$|f(x) - \sum_{i=1}^{N} c_i(f)\, g(\lambda_i \|x - y_i\|_{\mathbf{R}^n})| < \epsilon \tag{2}$$

holds for all $x \in K$, $f \in V$. Moreover, every $c_i(f)$ is a continuous functional defined on V.

Remark 1. It is worth noting that λ_i and y_i are all independent of f in V and $c_i(f)$ are continuous functionals. This fact will play an important role in the approximation to nonlinear operators by RBF networks.

To prove Theorem 1, we need to prove the following Lemma which is of great significance by itself.

Lemma 1 *Suppose that $h(x) \in C(\mathbf{R}^n) \cap S'(\mathbf{R}^n)$, then the family*

$$\sum_{i=1}^{N} c_i\, h(\lambda_i \rho_i(x) + y_i)$$

is dense in $C(K)$ if an only if h is not a polynomial, where ρ_i are rotations in $\mathbf{R}^n$, $y_i \in \mathbf{R}^n$, $\lambda_i \in \mathbf{R}^1$, $i = 1, \ldots, N$.

Proof. Sufficiency. Suppose that the linear combinations $\sum_{i=1}^{n} c_i\ h(\lambda_i \rho_i(x) + \theta_i)$ are not dense in $C[a, b]$, then by Hahn-Banach extension theorem of linear functionals and Riesz representation theorem (see [6]), we conclude that there is a bounded signed Borel measure $d\mu$ with $\mathrm{supp}(d\mu) \subseteq K$ and

$$\int_a^b h(\lambda\, \rho(x) + y)\, d\mu(x) = 0 \tag{3}$$

for all $\lambda \in \mathbf{R}^1$, $y \in \mathbf{R}^n$ and all rotation ρ. Pick any $w \in S(\mathbf{R}^n)$, then

$$\int_{\mathbf{R}^n} w(y)\, dy \int_{\mathbf{R}^n} h(\lambda\rho(x) + y)\, d\mu(x) = 0 \tag{4}$$

Changing the order of integral, we have

$$\int_{\mathbf{R}^n} h(u) du \int_{\mathbf{R}^n} w(y)\, d\mu(\rho^{-1}(\frac{u-y}{\lambda})) = 0 \tag{5}$$

which is equivalent to

$$\langle \hat{h}(t), \hat{w}(t)\hat{d\mu}(\lambda\rho^{-1}(t))\rangle = 0 \tag{6}$$

for all $\lambda \in \mathbf{R}^1$, $\lambda \neq 0$, rotations ρ, where $\hat{h}$ represents the Fourier transform of h in the sense of distribution.

In order for (6) to make sense, we have to show that

$$\hat{w}(t)\hat{d\mu}(\lambda\rho^{-1}(t)) \in S'(\mathbf{R}^n)\ .$$

In fact, $\hat{w}(t) \in S(\mathbf{R}^n)$. Moreover, since $\mathrm{supp}(d\mu) \subseteq K$, it is easy to verify that $\hat{d\mu}(t) = \int e^{-it\cdot x}\, d\mu(x) \in C^\infty(\mathbf{R}^n)$ and there are constants c_k, $k = 1, 2, \ldots$, such that

$$|\partial^k \hat{d\mu}(t)| \leq c_k\ . \tag{7}$$

Thus, $\hat{w}(t)\hat{d\mu}(t) \in S(\mathbf{R}^n)$.

Since $d\mu \neq 0$ and $\hat{d\mu}(t) \in C^\infty(\mathbf{R}^n)$, there is $t_0 \in \mathbf{R}^n$, $t_0 \neq 0$ and a neighborhood $O(t_0, \delta) = \{x : \|x - t_0\|_{\mathbf{R}^n} < \delta\}$ such that $|\hat{d\mu}(t)| > c > 0$ for all $t \in O(t_0, \delta)$.

Pick $t_1 \in \mathbf{R}^n$, $t_1 \neq 0$ arbitrarily. Let $t_0 = \lambda\rho(t_1)$, where ρ is a rotation in $\mathbf{R}^n$. Then $|\hat{d\mu}(\lambda\rho(t))| > c$ for all $t \in O(t_1, \delta/\lambda)$.

Let $\hat{w} \in C_c^\infty(O(t_1, \delta/\lambda))$, then $\hat{w}(t)/\hat{d\mu}(\lambda\rho(t)) \in C_c^\infty(O(t_1, \frac{\delta}{\lambda}))$, because $|\hat{d\mu}(\lambda\rho(t))| > \eta$ and $\hat{d\mu}(\lambda\rho(t)) \in C^\infty(\mathbf{R}^n)$. Therefore,

$$\langle \hat{h}(t), \hat{w}(t) \rangle = \langle \hat{h}(t), \ \frac{\hat{w}(t)}{\hat{d\mu}(\lambda\rho(t))} \hat{d\mu}(\lambda\rho(t)) \rangle = 0 \ . \tag{8}$$

Previous argument shows that for any $t^* \in \mathbf{R}^n$, $t^* \neq 0$, there is a neighborhood of t^*: $O(t^*, \eta)$ such that

$$\langle \hat{h}(t), \hat{w}(t) \rangle = 0 \tag{9}$$

holds for all $\hat{w}$ with support $\text{supp}(\hat{w}) \subseteq O(t^*, \eta)$, which means that $\text{supp}(\hat{g}) \subseteq \{0\}$. It is well known that a distribution is the Fourier transform of a polynomial if and only if its support is a subset of $\{0\}$. (see [24, Section 7.16]). Thus g is a polynomial.

Necessity. If g is a polynomial of degree m, then all the functions

$$\sum_{i=1}^{N} c_i g(\lambda_i \rho_i(x) + y_i)$$

are polynomials of $x_1, \ldots, x_n$ with total degree m, which, of course, are not dense in $C(K)$. Lemma 1 is proved. $\square$

Proof of Theorem 1.

Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$, then $h(x) = g(\|x\|_{\mathbf{R}^n}) \in S'(\mathbf{R}^n) \cap C(\mathbf{R}^n)$ and

$$\begin{aligned} & \sum_{i=1}^{N} c_i \, g(\lambda_i \|x - y_i\|_{\mathbf{R}^n}) \\ = & \sum_{i=1}^{N} c_i \, g(\lambda_i \|\rho(x) - \rho(y_i)\|_{\mathbf{R}^n}) \\ = & \sum_{i=1}^{N} c_i \, g(\|\lambda_i \rho(x) - z_i\|_{\mathbf{R}^n}) \end{aligned} \tag{10}$$

where $z_i = \lambda_i \rho(y_i)$.

From Lemma 1, we see that the family $\sum_{i=1}^{N} c_i\, g(\lambda_i \|x - y_i\|_{\mathbf{R}^n})$ is dense in $C(K)$, if and only if $g(\|x\|_{\mathbf{R}^n})$ is not a polynomial in $\mathbf{R}^n$, which is equivalent to that g is not an even polynomial. Theorem 1 is proved. $\square$

Lemma 2 *[20]* *Let $V \subseteq C(K)$, then V is a compact set in $C(K)$, if and only if (i) V is a closed set in $C(K)$; (ii) there is a constant M such that $\|f\|_{C(K)} \leq M$ for all $f \in V$; (iii) for any $\epsilon > 0$, there is $\delta > 0$, such that $|f(x') - f(x'')| < \epsilon$ for all $f \in V$, provided that x', $x'' \in K$ and $\|x' - x''\|_X < \delta$.*

We are now ready to prove Theorem 2.

Proof of Theorem 2.

Let $h(x) = e^{-\|x\|_{\mathbf{R}^n}}$ and $h_\delta(x) = \delta^{-n} h(\frac{x}{\delta})$ and

$$(f * h_\delta)(x) = \int_K f(t)\, h_\delta(x - t)\, dt\ . \tag{11}$$

It is easy to prove that for any $\epsilon > 0$, there is $\delta > 0$ such that

$$|(f * h_\delta)(x) - f(x)| < \epsilon/3 \tag{12}$$

holds for all $x \in K$ and $f \in V$.

Writing $(f * h_\delta)(x)$ in Riemann sum approximately, we see that there are positive integer N, $x_j \in \mathbf{R}^n$, constants $c_j^*(f)$ depending on f continuously, $i = 1, \cdots, N$ such that

$$|\sum_{j=1}^{N} c_j^*(f) e^{-\frac{1}{\delta}\|x - x_j\|^2_{\mathbf{R}^n}} - (f * h_\delta)(x)| < \epsilon/3 \tag{13}$$

for all $x \in K$ and $f \in V$.

Now, for every $j = 1, \ldots, N$, by Theorem 1 we can find another integer N_j, constants d_{ij}, λ_i, $y_i^{(j)} \in \mathbf{R}^n$, $i = 1, \ldots, N_j$ such that

$$|\sum_{i=1}^{N_j} d_{ij} g(\lambda_i \|x - y_i^{(j)}\|_{\mathbf{R}^n}) - e^{-\frac{1}{\delta}\|x - x_j\|^2_{\mathbf{R}^n}}| < \frac{\epsilon}{3L} \tag{14}$$

for all $x \in K$, where $L = \sup_{f \in V} \sum_{j=1}^{N} |c_j^*(f)|$.

Combining inequalities (12) (13) and (14), we conclude that there are integer N, vectors $y_i \in \mathbf{R}^n$, $c_i, \lambda_i \in \mathbf{R}^1$, $i = 1, \ldots, N$ such that

$$|f(x) - \sum_{i=1}^{N} c_i(f) g(\lambda_i \|x - y_i\|_{\mathbf{R}^n})| < \epsilon \tag{15}$$

for all $x \in K$ and $f \in V$.

Moreover, for every i, $c_i(f)$ is a continuous functional defined on V. Theorem 2 is proved. $\square$

4 Approximation to Nonlinear Functionals and Operators by RBF Neural Networks

In this section, we show some results concerning capability of RBF neural networks in approximating nonlinear functionals and operators defined on some Banach space, which can be used to approximate outputs of dynamical systems using sample data in either frequency (phase) domain or time (state) domain.

We first introduce one of our results.

Theorem 3 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ and is not an even polynomial, X is a Banach space with Schauder basis $\{x_n\}_{n=1}^{\infty}$, $K \subseteq X$ is a compact set in X, f is* *a continuous functional defined on K. Then for any $\epsilon > 0$, there are positive integers M, N, $y_i^M \in \mathbf{R}^M$, constants $c_i, \lambda_i \neq 0$, $i = 1, \ldots, N$, such that*

$$|f(x) - \sum_{i=1}^{N} c_i\, g(\lambda_i \|x^M - y_i^M\|_{\mathbf{R}^M})| < \epsilon \tag{16}$$

holds for all $x \in K$, where $x^M = (a_1(x), \ldots, a_M(x)) \in \mathbf{R}^M$, $x = \sum_{n=1}^{\infty} a_n(x) x_n$.

To prove Theorem 3, we need the following two lemmas.

Lemma 3 *[21] Suppose that X be a Banach space with Schauder basis $\{x_n\}_{n=1}^{\infty}$, then $K \subseteq X$ is a compact set in X if and only if the following two conditions are satisfied simultaneously: (1) K is a closed set in X; (2) for any $\delta > 0$, there is a positive integer M such that*

$$\| \sum_{n=M+1}^{\infty} a_n(x) x_n \|_X < \delta$$

holds for all $x \in K$.

Lemma 4 *Suppose that K is a compact set in a Banach space X with Schauder basis $\{x_n\}_{n=1}^{\infty}$. Define $K^n = \{\sum_{i=1}^{n} a_i(x) x_i, x \in K\}$ and $K^* = K \cup \bigcup_{n=1}^{\infty} K^n$, then K^n is a compact set in $\mathbf{R}^n$ and K^* is a compact set in X.*

Proof. It is easy to verify that K^n is a compact set in X_n (also in X), provided that K is a compact set in X.

Now, suppose $\{u^n\}_{n=1}^{\infty}$ is a sequence in K^*, then one of the following two cases occurs: (i) There is a subsequence $\{u^{n_k}\}_{k=1}^{\infty}$ of $\{u^n\}_{n=1}^{\infty}$ with all elements being in K or in some fixed K^n; (ii) There is no such subsequence.

In case (i), it is obvious there is another subsequence of $\{u^{n_k}\}_{k=1}^{\infty}$, which converges to some element u in K or in K^n, because K and K^n are compact sets in X.

In case (ii), there is a sequence $v^n = \sum_{i=1}^{\infty} a_i(v^n)x_i$ and integers M_n tending to infinity as $n \to \infty$, such that $u^n = \sum_{i=1}^{M_n} a_i(v^n)x_i$. By taking a suitable subsequence, without loss of generality, we can assume that v^n converges to some $v \in K$ as $n \to \infty$. By Lemma 3, $u^n - v^n \to 0$ as $n \to \infty$. Thus u^n converges to v.

Combining the two cases, we conclude that K^* is a compact set in X. Lemma 4 is proved. $\square$

Proof of Theorem 3. By Tietze extension theorem, we can extend f to a continuous functional f^* defined on K^*.

Since f^* is a continuous functional defined on the compact set K^*, then for any $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x') - f(x'')| < \epsilon/2$ provided that $x', x'' \in K^*$ and $\|x' - x''\| < \delta$.

By Lemma 3, there is an integer M such that

$$\|x' - x^M\|_X < \delta$$

for all $x \in K$.

Therefore

$$|f(x) - f^*(x^M)| < \epsilon/2 \tag{17}$$

for all $x \in K$. K^M is homomorphic to some compact set $K_{\mathbf{R}}^M$ in $\mathbf{R}^M$ by the map:

$$\{\sum_{n=1}^{M} a_n(x)x_n : x \in K\} \to \{x^M = (a_1(x), \ldots, a_M(x)) \in \mathbf{R}^M : x \in K\}.$$

Now, f^* is a continuous functional defined on every K_R^M. Applying Theorem 1, we can find an integer N, $y_i^M \in \mathbf{R}^M$, $c_i, \lambda_i \in \mathbf{R}^1$, $i = 1, \ldots, N$, such that

$$|f^*(x^M) - \sum_{i=1}^{N} c_i\, g(\lambda_i \|x^M - y_i^M\|_{\mathbf{R}^M})| < \epsilon/2 \tag{18}$$

for all $x^M \in K^M$.

Combining (17) and (18), we conclude that

$$|f(x) - \sum_{i=1}^{N} c_i\, g(\lambda_i \|x^M - y_i^M\|_{\mathbf{R}^M})| < \epsilon \tag{19}$$

for all $x \in K$. Theorem 3 is proved. □

To illustrate the applications of Theorem 3, we now give some examples.

Example 1. Let H be a Hilbert space, $K \subseteq H$ be a compact set in H. As we showed in [22], K can be considered as a compact set in a Hilbert space H_1 with countable basis $\{x_n\}_{n=1}^{\infty}$. Thus, Theorem 3 can be applied and we conclude that every continuous functional can be arbitrarily well approximated by the sum

$$\sum_{i=1}^{N} c_i\, g(\lambda_i \|x^M - y_i^M\|_{\mathbf{R}^M})$$

Example 2. Let $X = L^2[0, 2\pi]$, then $\{1, \{\cos nx\}_{n=1}^{\infty}, \{\sin nx\}_{n=1}^{\infty}\}$ is a Schauder basis and $a_n(x)$ are just the Fourier coefficients correspondingly. Thus, we can approximate every nonlinear functional defined on some compact set in $L^2[0, 2\pi]$ by RBF neural networks using sample data in *frequency domain.*

Example 3. Let $X = L^2(\mathbf{R}^1)$ and $\{\psi_{j,k}\}_{j,k=1}^{\infty}$ be wavelets in $L^2(\mathbf{R}^1)$. Then we can approximate continuous functionals defined on a compact set in $L^2(\mathbf{R}^1)$ by RBF neural network using *wavelet* coefficients as sample data (also in frequency domain).

Remark 2. Since in Theorem 3 we only require that $\{x_n\}_{n=1}^{\infty}$ is a Schauder basis (no orthogonal requirement is imposed), therefore we do not require that $\{\psi_{j,k}\}_{j,k=1}^{\infty}$ be orthogonal wavelets. This is a property of significant advantage, for non-orthogonal wavelets are much easier to construct than orthogonal wavelets.

The following theorem shows the possibility of approximating functionals by RBF neural networks using sample data in *time* (or state) domain.

Theorem 4 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ is not an even polynomial, X is a Banach space, $K \subseteq X$ is a compact set, V is a compact set in $C(K)$, f is a continuous functional defined on V, then for any $\epsilon > 0$, there are positive integers N, M, $x_1, \ldots, x_M \in K$ and $\lambda_i, c_i \in \mathbf{R}^1$, $\xi_i = (\xi_{i1}, \ldots, \xi_{iM}) \in \mathbf{R}^M$, $i = 1, \ldots, N$ such that*

$$|f(u) - \sum_{i=1}^{N} c_i\, g(\lambda_i \|u^M - \xi_i^M\|_{\mathbf{R}^M})| < \epsilon \tag{20}$$

for all $u \in V$, where $u^M = (u(x_1), \ldots, u(x_M)) \in \mathbf{R}^M$.

Proof of Theorem 4. We follow a similar line as used in [12]. Pick a sequence $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_n \to 0$, then we can find another sequence $\delta_1 > \delta_2 > \cdots > \delta_n \to 0$, such that $|f(u) - f(v)| < \epsilon_k$ for all $f \in V$, whenever $u, v \in V$ and $\|u - v\|_{C(K)} < \delta_k$. Moreover, we can find $\eta_1 > \eta_2 > \cdots \eta_n \to 0$, such that $|u(x') - u(x'')| < \delta_k$ for all $u \in V$, whenever $x', x'' \in K$ and $\|x' - x''\|_X < \eta_k$.

By induction and rearrangement, we can find a sequence $\{x_i\}_{i=1}^{\infty}$ with each $x_i \in K$ and a sequence of positive integers $n(\eta_1) < n(\eta_2) < \ldots < n(\eta_k) \to \infty$, such that the first $n(\eta_k)$ elements $N(\eta_k) = \{x_1, \ldots, x_{n(\eta_k)}\}$ is an η_k-net in K.

For each η_k-net, define functions

$$T^*_{\eta_k,j}(x) = \begin{cases} 1 - \frac{\|x-x_j\|_k}{\eta_k} & \text{if } \|x - x_j\|_X < \eta_k \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

and

$$T_{\eta_k,j}(x) = \frac{T^*_{\eta_k,j}(x)}{\sum_{j=1}^{n(\eta_k)} T^*_{\eta_k,j}(x)} \tag{22}$$

for $j = 1, \ldots, n(\eta_k)$. It is easy to verify that $\{T_{\eta_k,j}(x)\}$ is a partition of unity, *i.e.*

$$\begin{cases} 0 \le T_{\eta_k,j}(x) \le 1 \\ \sum_{j=1}^{n(\eta_k)} T_{\eta_k,j}(x) \equiv 1 \qquad \text{for } x \in K \\ T_{\eta_k,j}(x) = 0 \qquad \text{if } \|x - x_j\|_X > \eta_k \ . \end{cases} \tag{23}$$

For each $u \in V$, define a function

$$u_{\eta_k}(x) = \sum_{j=1}^{n(\eta_k)} u(x_j) T_{\eta_k,j}(x) \tag{24}$$

Moreover, let $V_{\eta_k} = \{u_{\eta_k} : u \in V\}$ and $V^* = V \cup (\cup_{k=1}^{\infty} V_{\eta_k})$. We then have the following conclusion:[3]

1. For any fixed k, V_{η_k} is a compact set in a space of dimension $n(\eta_k)$ in $C(K)$.

2. For every $u \in V$, there holds

$$\|u - u_{\eta_k}\|_{C(K)} < \delta_k \tag{25}$$

3. V^* is a compact set in $C(K)$.

Now, similar to the proof of Theorem 3, we can extend f to a continuous functional f^* on V^*, such that

$$f^*(x) = f(x) \qquad \text{if } x \in V \ . \tag{26}$$

[3] For a proof of these three propositions, see Lemma 7 [12] or the Appendix of this paper.

Now, for any $\epsilon > 0$, we can find a $\delta > 0$ such that $|f^*(u) - f^*(v)| < \epsilon/2$ provided that $u, v \in V^*$ and $\|u - v\|_{C(K)} < \delta$.

Let k be fixed such that $\delta_k < \delta$, then by (24), for every $u \in V$,

$$\|u - u_{\eta_k}\|_X < \delta_k \tag{27}$$

which implies

$$|f^*(u) - f^*(u_{\eta_k})| < \epsilon/2 \tag{28}$$

for all $u \in V$.

By the argument used in the proof of Theorem 3, letting $M = n(\eta_k)$, there is an integer N, $\lambda_i, c_i \in \mathbf{R}^1$, $\xi_i^M = (\xi_{i1}^M, \ldots, \xi_{iM}^M) \in \mathbf{R}^M$, $i = 1, \ldots, N$ and M points $x_1, \ldots, x_M \in K$ such that

$$|f^*(u_{\eta_k}) - \sum_{i=1}^{N} c_i \, g(\lambda_i \|u^M - \xi_i^M\|_{\mathbf{R}^M})| < \epsilon/2 \tag{29}$$

which, combined with (28), leads to

$$|f(u) - \sum_{i=1}^{N} c_i \, g(\lambda_i \|u^M - \xi_i^M\|_{\mathbf{R}^M})\| < \epsilon \tag{30}$$

for all $u \in V$. This completes the proof of Theorem 4. $\square$

To conclude this section, we construct an RBF neural network to approximate nonlinear operators. More precisely, we prove

Theorem 5 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ and is not an even polynomial, X is a Banach space, $K_1 \subseteq X$, $K_2 \subseteq \mathbf{R}^n$ are two compact sets in X and $\mathbf{R}^n$ respectively, V is a compact set in $C(K_1)$, G is a nonlinear continuous operator, which maps V into*

$C(K_2)$, then for any $\epsilon > 0$, there are positive integers M, N, m, constants $c_i^k, \lambda_k, \mu_i \in \mathbf{R}^1$, $k = 1, \ldots, N$, $i = 1, \ldots, M$, m points $x_1, \ldots, x_m \in K_1$, $\omega_1, \ldots, \omega_N \in \mathbf{R}^n$, such that

$$|G(u)(y) - \sum_{i=1}^{M} \sum_{k=1}^{N} c_i^k \, g(\mu_i \|u^m - \xi_{ik}^m\|_{\mathbf{R}^m}) \, g(\lambda_k \|y - \omega_k\|_{\mathbf{R}^n})| < \epsilon \tag{31}$$

for all $u \in V$ and $y \in K_2$, where $u^m = (u(x_1), \ldots, u(x_m))$, $\xi_k^m = (\xi_{k1}^m, \ldots, \xi_{k,m}^m)$, $k = 1, \ldots, N$.

Proof of Theorem 5. Since G is a continuous operator which maps a compact set V in $C(K_1)$ into $C(K_2)$, then the range $G(V) = \{G(u) : u \in V\}$ is also a compact set in $C(K_2)$. By Theorem 2, for any $\epsilon > 0$, there are a positive integer N, $\omega_k \in \mathbf{R}^n$, $\lambda_k \in \mathbf{R}^1$, $c_k(G(u)) \in \mathbf{R}$, $k = 1, \ldots, N$, such that

$$|G(u)(y) - \sum_{k=1}^{N} c_k(G(u)) \, g(\lambda_k \|y - \omega_k\|_{\mathbf{R}^n})| < \epsilon/2 \tag{32}$$

for all $u \in V$ and $y \in K_2$, and all $c_k(G(u))$, $k = 1, \ldots, N$ are continuous functional defined on V. Carefully examing the proof of Theorem 5, we can find a common m, integer N_k, constants c_i^k, μ_i, $\xi_{ik}^m \in \mathbf{R}^m$, $i = 1, \ldots, N_k$, m points $x_1, \ldots, x_m \in K_1$, $u^m = (u(x_1), \ldots, u(x_m)) \in \mathbf{R}^m$, such that

$$|c_k(G(u)) - \sum_{i=1}^{N_k} c_i^k g(\mu_i \|u^m - \xi_{ik}^m\|_{\mathbf{R}^m})| < \frac{\epsilon}{2L} \tag{33}$$

holds for all $k = 1, \ldots, N$, $u \in V$, where

$$L = \sum_{k=1}^{N} \sup_{y \in K_2} |g(\lambda_k \|y - \omega_k\|_{\mathbf{R}^n})| \, . \tag{34}$$

Substituting (33) into (32), we obtain

$$|G(u)(y) - \sum_{k=1}^{N} \sum_{i=1}^{N_k} c_i^k \, g(\mu_i \|u^m - \xi_{ik}^m\|_{\mathbf{R}^m}) \, g(\lambda_k \|y - \omega_k\|_{\mathbf{R}^n})| < \epsilon \tag{35}$$

for all $u \in V$ and $y \in K_2$.

Let $M = \max_k\{N_k\}$ and let $c_i^k = 0$ for $N_k < i \leq M$. Thus (35) can be rewritten as

$$|G(u)(y) - \sum_{k=1}^{N}\sum_{i=1}^{M} c_i^k\, g(\mu_i\|u^m - \xi_{ik}^m\|_{\mathbf{R}^m})\, g(\lambda_k\|y - \omega_k\|_{\mathbf{R}^n})| < \epsilon$$

for all $u \in V$ and $y \in K_2$. Theorem 5 is proved. $\square$

Remark 3. Theorem 5 shows the capability of RBF networks in approximating nonlinear operators using sample data in *time* (or state) domain. Likewise, by Theorem 3, we can construct RBF networks using sample data in *frequency* (or phase) domain.

Remark 4. We can also construct neural networks, where affine basis functions are mixed with radial basis functions. For example, we can approximate $G(u)(y)$ by

$$\sum_{k=1}^{N}\sum_{i=1}^{M} c_i^k\, g(\mu_i\|u^m - \xi_{ik}^m\|_{\mathbf{R}^m})\, g(\omega_k \cdot y + \zeta_k)\ .$$

The details are omitted here.

Remark 5. In engineering, a system can be viewed as a operator (linear or nonlinear). Theorem 5 shows the capability of RBF neural networks in identifying systems, as compared with the Cybenko's theorem, which shows the capability of ABF neural networks in pattern recognition.

5 Conclusion

In this paper, the problem of approximating functions of several variables, functionals and nonlinear operators by radial basis function neural networks are studied. The necessary and sufficient condition for a continuous function to be qualified as an activation function in RBF networks is given. Results on using RBF neural networks for computing the output of dynamical systems by sample data in frequency domain or time domain are given.

Acknowledgements. The authors wish to thank Prof. R.-W. Liu of University of Notre Dame and Prof. I.W. Sandberg of University of Texas-Austin 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," in *IEEE ICNN 1*, pp. 641-648 (1988).

[3] S. M. Carroll and B. W. Dickinson, "Construction of Neural Nets using Radon Transform," in *IJCNN Proc. I*, pp. 607-611 (1989).

[4] K. Funahashi, "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, pp. 183-192, Vol. 2, (1989).

[5] H.N. Mhaskar, C.A. Micchelli, "Approximation by Superposition of Sigmoidal and Radial Functions," *Advances on Applied Mathematics*, pp. 350-373, vol. 13, (1992).

[6] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function," in *Math. of Control, Signals and Systems*, Vol. 2, No. 4, pp. 303-314 (1989).

[7] Y. 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).

[8] K. Hornik, "Approximation Capabilities of Multilayer Feedforward Networks," *Neural Networks*, Vol. 4, pp. 251-257 (1991).

[9] T. Chen, H. Chen and R.-W. Liu, "Approximation Capability in $C(\mathbf{R}^n)$ by Multilayer Feedforward Networks and Related Problems," *IEEE Trans. on Neural Networks*, accepted.

[10] T. Chen, H. Chen and R.-W. Liu, "A Constructive Proof of Cybenko's Approximation Theorem and Its Extensions," pp. 163 - 168 in *Computing Science and Statistics* (editors LePage and Page), Proc. of the 22nd Symposium on the Interface (East Lansing, Michigan), May 1990.

[11] T. Chen and H. Chen, "Approximation to Continuous Functionals by Neural Networks with Application to Dynamic Systems," *IEEE Trans. on Neural Networks* (in press).

[12] T. Chen and H. Chen, "Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamic Systems," submitted for publication.

[13] R.P. Lippmann, "Pattern Classification using Neural Networks," *IEEE Commun. Magazine*, pp. 47-64, Vol. 27 (1989).

[14] J. Park and I.W. Sandberg, "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, pp. 246-257, Vol. 3 (1991).

[15] D. Hush and B. Horne, "Progress in Supervised Neural Networks," *IEEE SP Magazine*, Jan. (1993).

[16] J.E. Moody and C.J. Darken, "Fast Learning in Networks of Locally-tuned Processing Units," *Neural Computation*, 1:281-293 (1989).

[17] F. Girosi and T. Poggio, "Networks and the Best Approximation Property," Artificial Intelligence Lab, Memo 1164, MIT (1989).

[18] E.J. Hartman, J.D. Keeler and J.M. Kowalski, "Layered Neural networks with Gaussian Hidden Units as Universal Approximators," *Neural Computation*, pp. 210-215, Vol. 2(2) (1990).

[19] S. Lee and R.M. Kil, "A Gaussian Potential Function Network with Hierarchically self-organizing learning," *Neural Networks*, pp. 207-224, Vol. 4 (1991).

[20] J. Diedonne, *Foundation of Modern Analysis*, Academic Press : New York and London (1969), p. 142.

[21] L.A. Liusternik and V.J. Sobolev, *Elements of Functional Analysis* (3rd ed., translated from Russian 2nd. ed.), Wiley : New York (1974).

[22] T. Chen, "Approximation to Nonlinear Functionals in Hilbert Space by Superposition of Sigmoidal Functions," *Kexue Tongbao* (1992).

[23] J. Park and I.W. Sandberg, "Approximation and Radial-Basis-function Networks," *Neural Computation*, Vol 5 (1993).

[24] T. Chen and H. Chen, "$L^2(\mathbf{R}^n)$ Approximation by RBF neural Networks," *Chinese Annals of Mathematics*, (in press).

[25] W.Rudin, *Functional Analysis*, McGraw-Hill, New York, 1973.

Appendix

Proof of The Three Propositions in the Proof of Theorem 4:

We will prove the three propositions individually as follows.

1. For a fixed k, let $u^{(j)}_{\eta_k}$, $j = 1, 2, \ldots$, be a sequence in V_{η_k} and $u^{(j)}$ be the sequence in V, such that
$$u^{(j)}_{\eta_k} = \sum_{j=1}^{n(\eta_k)} u^{(j)}(x_j) T_{\eta_k, j}(x) \ . \tag{36}$$
Since V is compact, there is a subsequence $u^{(j_l)}(x)$ which converges to some $u \in V$, it then follows that $u^{(j_l)}_{\eta_k}(x)$ converges to $u_{\eta_j}(x) \in V_{\eta_k}$, which means that V_{η_k} is a compact set in $C(K)$.

2. By the definition and the property of unity partition, we have
$$\begin{aligned} u(x) - u_{n(\eta_k)}(x) &= \sum_{j=1}^{n(\eta_k)} [u(x) - u(x_j)] T_{\eta_k, j}(x) \\ &= \sum_{\|x - x_j\|_X \le \eta_k} [u(x) - u(x_j)] T_{\eta_k, j}(x) \ . \end{aligned} \tag{37}$$
Consequently,
$$\|u(x) - u_{n(\eta_k)}(x)\|_X \le \delta_k \sum_{j=1}^{n(\eta_k)} T_{\eta_k, j}(x) = \delta_k \tag{38}$$
for all $u \in V$.

3. Suppose $\{u^j\}_{j=1}^\infty$ is a sequence in V^*. If there is a subsequence $\{u^{j_l}\}_{l=1}^\infty$ of $\{u^j\}_{j=1}^\infty$ with all $u^{j_l} \in V$, $l = 1, \ldots$, then by the fact that V is compact, there is a subsequence of $\{u^{j_l}\}_{l=1}^\infty$ which converges to some $u \in V$. Otherwise, to each u^j, there corresponds a positive integer $k(j)$ such that $u^j = v^j_{n(\eta_{k(j)})}$. There are two possibilities: (i) We can find infinite j_l such that $u^{j_l} \in V_{\eta_{k_0}}$ for some fixed k_0. By proposition 1. proved in this Appendix, $V_{n(\eta_{k_0})}$ is a compact set, hence there is a subsequence of $\{v^j_{n(\eta_{k(j)})}\}$ which converges to some $v \in V_{\eta_{k_0}}$, *i.e.* a subsequence of $\{u^j\}$ converging to v. (ii) There are sequences $j_1 < j_2 < \ldots \to \infty$ and $k(j_1) < k(j_2) < \ldots \to \infty$ such that $u^{j_l} \in V_{\eta_{k(j_l)}}$. Let $v^{j_l} \in V$ be such that
$$u^{j_l}(x) = \sum_{i=1}^{n(\eta_{k(j_l)})} v^{j_l}(x_i) T_{\eta_{k(j_l)}, i}(x) \ . \tag{39}$$
Since $V^{j_l} \in V$ and V is compact, we observe that there is a subsequence of $\{v^{j_l}\}_{l=1}^\infty$, which converges to some $v \in V$. By the proposition 2. proved in this Appendix, the corresponding subsequence of $\{u^{j_l}\}_{l=1}^\infty$ also converges to v. Thus the compactness of V^* is proved. $\square$

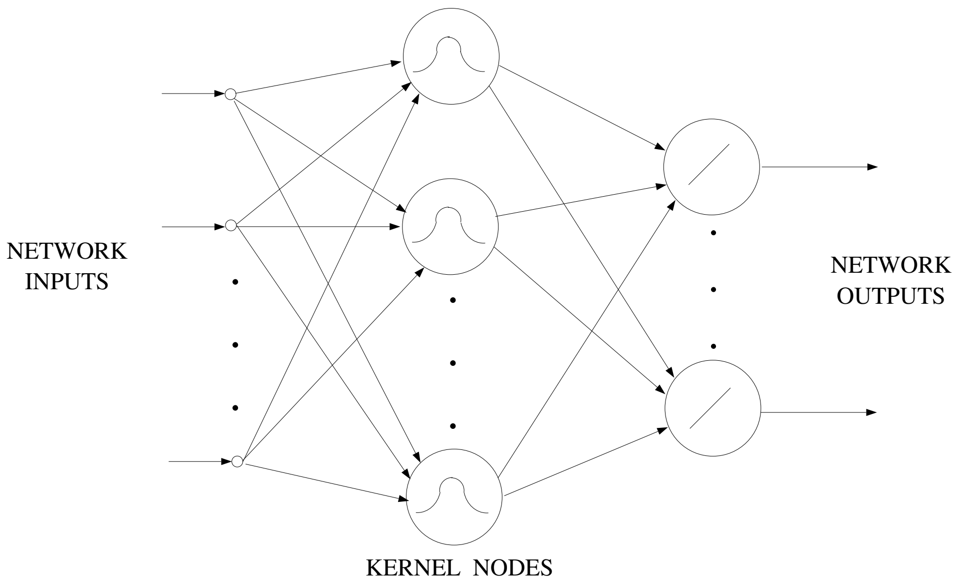

Figure 1: A Radial Basis Function Network