Neural Networks 24 (2011) 752–758

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

# Essential rate for approximation by spherical neural networks

Shaobo Lin [a,*], Feilong Cao [b], Zongben Xu [a]

[a] *Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an 710049, Shannxi Province, PR China*

[b] *Institute of Metrology and Computational Science, China Jiliang University, Hangzhou 310018, Zhejiang Province, PR China*

ARTICLE INFO

*Article history:*
Received 31 December 2010
Received in revised form 18 April 2011
Accepted 22 April 2011

*Keywords:*
Neural networks
Spherical polynomials
Approximation
Essential rate

ABSTRACT

We consider the optimal rate of approximation by single hidden feed-forward neural networks on the unit sphere. It is proved that there exists a neural network with $n$ neurons, and an analytic, strictly increasing, sigmoidal activation function such that the deviation of a Sobolev class $W_{2r}^2(\mathbf{S}^d)$ from the class of neural networks $\Phi_n^\phi$, behaves asymptotically as $n^{-\frac{2r}{d-1}}$. Namely, we prove that the essential rate of approximation by spherical neural networks is $n^{-\frac{2r}{d-1}}$.



## 1. Introduction

In the $(d+1)$-dimensional Euclidean space $\mathbf{R}^{d+1}$, feed-forward neural networks (FNNs) have attracted the attention of large number of scholars for their universal approximation property. There are two main problems concerning the research of FNN approximation. The first one is called density, which deals with deciding whether it is possible to approximate the target function arbitrarily well by choosing suitable network models. The typical results can be found in Chen and Chen (1995), Chui and Li (1992), Cybenko (1989), Funahashi (1989), Hornik, Stinchcombe, and White (1990), Leshno, Lin, Pinks, and Schocken (1993) and Park and Sandberg (1991, 1993) and so on.

The other problem of such approximation called complexity is to determine how many neurons are necessary to yield a prescribed degree of approximation, which mainly describes the relationship among the topology structure of hidden layers, the approximation ability and the approximation rate. There have been many studies for this problem. We refer the readers to Barron (1993), Bulsari (1993), Ferrari and Stengel (2005), Korain (1993), Maiorov and Meir (1998), Makovoz (1998), Mhaskar and Micchelli (1995), Suzuki (1998) and Xu and Cao (2004).

Rates of approximation describe the trade-off between the accuracy of approximation and the complexity of approximating functions. When such functions belong to a parameterized family, their complexity can be measured by the lengths of parameter vectors (depending on the number of variables on the degree of polynomials, or on the number of hidden units in neural networks, etc.). The comparison of rates of approximation between polynomials and FNNs have been studied by several authors. For example, in the previous paper (Cao, Lin, & Xu, 2010), we proved that if the activation function of FNNs is analytic and non-polynomial, then the approximation rate of FNNs is not lower than that of the polynomial. On the other hand, Konovalov, Leviatan, and Maiorov (2008) proved that if the target function is radial, then the approximation rate of algebraic polynomials is not slower than that of FNNs in the square integrable function space (indeed, Konovalov et al., 2008, proved this property for any ridge function manifolds). Similar results can be found in Maiorov and Pinkus (1999), Mhaskar (1996), Petrushev (1999) and Xie and Cao (2010) and references therein.

In order to reflect the approximation capability of FNNs more precisely, it is natural to raise the question: what about the lower bound of approximation? As regards to this question, there have been some papers such as Konovalov et al. (2008); Konovalov, Leviatan, and Maiorov (2009), Maiorov (1999, 2003) and Xu and Cao (2004) etc. dealing with the lower bound for approximation by FNNs with various activation functions and target functions. If the upper and lower bounds are asymptotically identical, then we call the degree of the bounds as the essential rate of approximation.

On the other hand, many applications such as geophysics, metrology, graph rendering and so on, the data are usually collected over a sphere or sphere-like area. One then seeks to find a functional model for the mechanism that generates the data. For example, the mathematical models of some satellite missions such as GOCE and CHAMP, studying the gravity potential of the earth, need to solve spherical Fredholm integral equations of the first kind. Hence, find a tool which can deal with spherical data by using some special properties of the sphere becomes more and more important.

* Corresponding author.
*E-mail address:* ssxvkihc@yahoo.com.cn (S. Lin).

0893-6080/$ – see front matter 
doi:10.1016/j.neunet.2011.04.005

A feasible tool for dealing with spherical data is the spherical polynomials (SPs). The direct and inverse approximation theorem of SPs have been studied by several scholars by using some well-known spherical polynomial operators: Lizorkin and Nikol'skiĭ (1983) for spherical Jackson operator; Mhaskar, Narcowich, and Ward (1999), for spherical delay means operator; Wang and Li (2000) for spherical de la Vallée Poussin operator; Dai and Ditzian (2008), for the best approximation operator etc.

A major problem of approximation by SPs is the so-called curse of dimensionality, whereby performance degrades rapidly as the dimensionality of the problem increases. Several procedures have been suggested in order to circumvent this problem. A typical approach on the sphere is the zonal function networks (ZFNs) formed as

$$x \to \sum_{k=1}^{n} a_k \phi(\langle \xi_k, x\rangle), \tag{1.1}$$

where the weights $\xi_k$ are the site of scattered spherical data, and $\langle x, y\rangle$ denotes the inner product of $(d+1)$ dimensional vectors $x$ and $y$. In the seminal paper (Sun & Cheney, 1997), the sufficient and necessary conditions for the density of ZFNs have been deduced. Two years later, Mhaskar et al. (1999) established the complexity of approximation by ZFNs. They compared the rate of approximation of ZFNs with that of SPs. They proved that if the activation functions of the ZFNs satisfy some conditions (such as Gaussian function), then the upper rate of approximation by ZFNs and SPs are identical when the neurons of ZFNs $n$ and the degree of SPs $s$ satisfy $n \sim s^d$, i.e. they proved that the rate of approximation by ZFNs with Gaussian activation function in a Sobolev $W_{2r}^2$ (which will be defined in Section 2) is $\mathcal{O}\left(n^{-\frac{2r}{d}}\right)$. For general target functions and activation functions, Mhaskar et al. (1999) used the summation of the best approximation of SPs and a redundancy depending on the smoothness of the activation functions to bound the best approximation of ZFNs. Some studies for approximation by ZFNs on the sphere can also be found in Mhaskar (2006), Mhaskar, Narcowich, and Ward (2003) and Narcowich, Sun, Ward, and Wendland (2007).

In this paper, by using the traditional idea of neural networks, we introduce a new approximant on the sphere called spherical neural networks (SNNs) formed as

$$N_{\phi,n}(x) := \sum_{i=1}^{n} c_i \phi(\langle w_i, x\rangle + \theta_i), \quad x \in \mathbf{S}^d, \tag{1.2}$$

where $w_i \in \mathbf{R}^{d+1}, \theta_i, c_i \in \mathbf{R}$. We denote by $\Phi_{\phi,n}$ the collection of all functions formed as (1.2). It is obvious that ZFN is a special type of SNN (by setting the thresholds to 0 and restricting the inner weight to the sphere). Thus results about ZFNs are automatically results about SNNs. Our main idea of introducing SNNs is that by adding thresholds to the ZFNs, we can essentially improve the rate of approximation. More precisely, by using SNNs, we can deduce a similar result as that of ZFNs by using much less neurons. Indeed, it will be shown in Section 3 that if $n \sim s^{d-1}$, then there exists an SNN with analytic, strictly increasing and sigmoidal activation function such that the upper bound of approximation is not larger than that of SPs. Therefore, the upper bound of approximation by SPs can deduce the upper bound of approximation by SNNs. For example, if $f \in W_{2r}^2$, then the approximation rate of SNNs is $\mathcal{O}\left(n^{-\frac{2r}{d-1}}\right)$, which is better than that of ZFNs.

The other work of this paper is to study the lower bound of approximation by SNNs. By help of a lemma proved by Maiorov (1999) and the Funk–Hecke formula, we will prove that for arbitrary $f \in W_{2r}^2$, the lower rate of approximation by SNNs also asymptotically behaves as $n^{-\frac{2r}{d-1}}$.

The rest of this paper is organized as follows. In the next section, we will give some preliminaries about the classical spherical polynomials. The upper bound of approximation by SNNs will be proved in Section 3, where the relation between approximation by SNNs and SPs will be also given. The lower bound of approximation by SNNs will be shown in Section 4, while in the last section, we will give some remarks.

To aid our description, we adopt the following convention regarding symbols. Let $C, C_1, C_2, \ldots$ be constants depending only on $d$, whose values will be different at different occurrences, even within the same formula. The symbol $A \sim B$ means $CA \le B \le C_1 A$. The volume of $\mathbf{S}^d$ is denoted by $\Omega_d$, and it is easy to deduce that

$$\Omega_d := \int_{\mathbf{S}^d} \mathrm{d}\omega = \frac{2\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)}.$$

## 2. Notations and preliminaries

At first, we introduce a Sobolev space on the sphere.

Consider the Hilbert space $L^2(\mathbf{S}^d)$ with norm

$$\|f\|_2 := \|f\|_{L^2(\mathbf{S}^d)} := \left(\int_{\mathbf{S}^d} |f(x)|^2 \mathrm{d}\omega(x)\right)^{1/2}$$

and inner product

$$\langle f, g\rangle_2 := \int_{\mathbf{S}^d} f(x)\overline{g(x)}\mathrm{d}\omega(x),$$

where $\mathrm{d}\omega(x)$ is the elementary surface piece on $\mathbf{S}^d$, The Laplace–Beltrami operator $\Delta$ is defined by (see Freeden, Gervens, & Schreiner, 1998; Müller, 1966; Wang & Li, 2000)

$$\Delta f := \sum_{i=1}^{d+1} \frac{\partial^2 g(x)}{\partial x_i^2}\Bigg|_{|x|:=(x_1^2+x_2^2+\cdots+x_{d+1}^2)^{1/2}=1}, \quad g(x) = f\left(\frac{x}{|x|}\right).$$

For every positive integer $r$, we denote by $H_{2r}^2(\mathbf{S}^d)$ the class of functions $f$ for which $\Delta^r f \in L^2(\mathbf{S}^d)$, where $\Delta^r f := \Delta^{r-1}(\Delta f)$, $r = 2, 3, \ldots$. And we let $W_{2r}^2$ be the subset of $H_{2r}^2(\mathbf{S}^d)$ with $\|\Delta^r f\|_2 \le 1$, $f \in H_{2r}^2(\mathbf{S}^d)$. The Sobolev space has been widely used to describe the smoothness of functions. We refer the readers to Chapter 5 of Freeden et al. (1998) for more details about $W_{2r}^2$.

Let $\mathcal{V}$ be a measurable set. For any two sets $W, U \subset \mathcal{V}$, we define the deviation of $W$ from $U$ as

$$\mathrm{dist}(W, U, \mathcal{V}) := \sup_{f\in W} \mathrm{dist}(f, U, \mathcal{V}) := \sup_{f\in W} \inf_{g\in U} \|f-g\|_{\mathcal{V}}.$$

We denoted by $\mathbf{H}_k^d$ and $\Pi_s^d$ the class of all spherical harmonics with degree $k$, and the class of all SPs with degree $k \le s$, respectively. The space $\mathbf{H}_k^d$ can be characterized intrinsically as the eigenspace corresponding to $\lambda_k := k(k+d-1)$, $k = 0, 1, \ldots$, i.e.

$$\Delta H_k = -\lambda_k H_k, \quad H_k \in \mathbf{H}_k^d. \tag{2.1}$$

Since the $\lambda_k$'s are distinct, and the operator is self-adjoint, the spaces $\mathbf{H}_k^d$ are mutually orthogonal. So we have $\Pi_s^d = \bigoplus_{k=0}^{s} \mathbf{H}_k^d$ and $L^2(\mathbf{S}^d) = \mathrm{closure}\{\bigoplus_k \mathbf{H}_k^d\}$. Hence, if we choose an orthonormal basis $\{Y_{k,l} : l = 1, \ldots, d_k^d\}$ for each $\mathbf{H}_k^d$, then the set $\{Y_{k,l} : k = 0, 1, \ldots, l = 1, \ldots, d_k^d\}$ is an orthonormal basis for $L^2(\mathbf{S}^d)$. The dimension of $\mathbf{H}_k^d$ is given by

$$d_k^d := \dim \mathbf{H}_k^d = \begin{cases} \dfrac{2k+d-1}{k+d-1}\dbinom{k+d-1}{k}, & k \ge 1; \\ 1, & k = 0, \end{cases}$$

and that of $\Pi_s^d$ is $\sum_{k=0}^{s} d_k^d = d_s^{d+1} \sim s^d$.

The addition formula establishes a connection between spherical harmonics of degree $k$ and the Legendre polynomial $P_k^{d+1}$ (see Müller, 1966; Wang & Li, 2000):

$$\sum_{l=1}^{d_k^d} Y_{k,l}(x)Y_{k,l}(y) = \frac{d_k^d}{\Omega_d} P_k^{d+1}(\langle x, y\rangle), \tag{2.2}$$

where $P_k^{d+1}$ is the Legendre polynomial with degree $k$ and dimension $d+1$. The Legendre polynomial $P_k^{d+1}$ can be normalized such that $P_k^{d+1}(1) = 1$, and satisfies the orthogonality relations

$$\int_{-1}^{1} P_k^{d+1}(t)P_j^{d+1}(t)(1-t^2)^{\frac{d-2}{2}}\,\mathrm{d}t = \frac{\Omega_d}{\Omega_{d-1}d_k^d}\delta_{k,j}, \tag{2.3}$$

where $\delta_{k,j}$ is the usual Kronecker symbol.

The following Funk–Hecke formula plays an important role in computing the eigenvalues of kernel $\phi \in L^1([-1, 1])$ (see Müller, 1966; Wang & Li, 2000)

$$\int_{\mathbf{S}^d} \phi(\langle x, z\rangle)P_k^{d+1}(\langle y, z\rangle)\mathrm{d}\omega(z) = B(\phi, k)P_k^{d+1}(\langle x, y\rangle), \tag{2.4}$$

where

$$B(\phi, k) = \Omega_{d-1}\int_{-1}^{1} P_k^{d+1}(t)\phi(t)(1-t^2)^{\frac{d-2}{2}}\,\mathrm{d}t.$$

From (2.4) it is easy to deduce the following general Funk–Hecke formula (see Müller, 1966; Wang & Li, 2000), i.e. for arbitrary $H_k \in \mathbf{H}_k^d$, we have

$$\int_{\mathbf{S}^d} \phi(\langle x, y\rangle)H_k(y)\mathrm{d}\omega(y) = B(\phi, k)H_k(x). \tag{2.5}$$

In order to give the upper bound of approximation by SNNs, we need the following representation theorem, which was proven by Cao et al. (2010). For the sake of completeness, we sketch its proof.

**Lemma 2.1.** *Let $s \in \mathbf{N}$. Then for any $P_s \in \Pi_s^d$, there exists a set of points $\{a_d\}_{k=1}^{d_s^d} \in \mathbf{S}^d$ and a set of univariate polynomials $\{g_j\}_{k=1}^{d_s^d}$ defined on $[-1, 1]$ with degrees not larger than $s$ such that*

$$P_s(x) = \sum_{k=1}^{d_s^d} g_k(\langle a_k, x\rangle), \quad x \in \mathbf{S}^d. \tag{2.6}$$

**Sketch of proof.** Set $\{Y_{j,i} : j = 0, 1, \ldots, s, i = 1, 2, \ldots, d_j^d\}$ be an orthonormal basis of $\Pi_n^d$. Since for any univariate polynomials $g_k$ $(k = 1, \ldots, N)$ defined on $[-1, 1]$ with degrees not larger than $n$ and any points $a_k \in \mathbf{S}^d$ $(k = 1, \ldots, d_s^d)$, $\sum_{k=1}^N g_k(\langle a_k, x\rangle) \in \Pi_n^d$. In order to prove (2.6), it is sufficient to prove that there exist a set of univariate polynomials $\{g_k\}_{k=1}^{d_s^d}$ defined on $[-1, 1]$ with degrees not larger than $s$ and a set of points $\{a_k\}_{k=1}^{d_s^d} \subset \mathbf{S}^d$ such that the Fourier coefficients related to the orthonormal basis $\{Y_{j,i} : j = 0, 1, \ldots, n, i = 1, 2, \ldots, d_j^d\}$ of both sides of (2.6) coincide. If we set $\int_{\mathbf{S}^d} P_s(x)Y_{j,i}(x)\mathrm{d}\omega(x) = Q_{j,i}$, then we only need to prove that there exist a set of univariate polynomials $\{g_k\}_{k=1}^{d_s^d}$ defined on $[-1, 1]$ with degrees not larger than $s$ and a set of points $\{a_k\}_{k=1}^{d_s^d} \subset \mathbf{S}^d$ such that

$$\int_{\mathbf{S}^d} \sum_{k=1}^{d_s^d} g_k(\langle a_k, x\rangle)Y_{j,i}(x)\mathrm{d}\omega(x) = Q_{j,i},$$

$j = 0, \ldots, s, i = 1, \ldots, d_j^d$.

It follows from the Funk–Hecke formula (2.5) that

$$\int_{\mathbf{S}^d} \sum_{k=1}^{d_s^d} g_k(a_k \cdot x)Y_{j,i}(x)\mathrm{d}\omega(x) = \sum_{k=1}^{d_s^d} B(g_k, j)Y_{j,i}(a_k).$$

If we set

$$z_{ji}^{km} := \begin{cases} Y_{j,i}(a_k), & m = j, \\ 0, & m \neq j, 0 \leq m \leq n, \end{cases}$$

then it is sufficient to prove that there exist a set of points $\mathcal{A} := \{a_1, a_2, \ldots, a_N\} \subset \mathbf{S}^d$ and a set of univariate polynomials $\{g_k\}_{k=1}^{d_s^d}$ defined on $[-1, 1]$ with degrees not larger than $s$ such that

$$\sum_{m=0}^{s}\sum_{k=1}^{d_s^d} z_{ji}^{km}B(g_k, m) = Q_{j,i}, \quad j = 0, \ldots, n,\ i = 1, \ldots, d_j^d.$$

Construct the matrix

$$Z(\mathcal{A}) := \left(z_{ji}^{km}\right)_{j=0,\ldots,s,\ i=1,\ldots,d_j^d}^{k=1,\ldots,d_s^d,\ m=0,\ldots,s},$$

where $(k, m)$ is the indicator number of the column of the matrix $Z(\mathcal{A})$, and $(j, i)$ is the indicator number of the row of the matrix $Z(\mathcal{A})$. If we construct the vectors

$$\mathcal{B} := (B(g_k, m))_{k=1,\ldots,d_s^d,\ m=0,\ldots,s},$$

$$\mathcal{Q} := (Q_{j,i})_{j=0,\ldots,n,\ i=1,\ldots,d_j^d},$$

then it is sufficient to prove that there exists $\mathcal{A}$ such that the system of equations with variables $B(g_k, m)$

$$Z(\mathcal{A})\mathcal{B} = \mathcal{Q}$$

is solvable. Then by using the method of Maiorov (2003), we can deduce (2.6) easily. □

## 3. Upper bound of approximation

Let $C(\mathbf{S}^d)$ be the set of continuous functions on $\mathbf{S}^d$. In this section, we prove that for any $f \in C(\mathbf{S}^d)$ and any $\varepsilon > 0$, there exists an SNN, $N_{\phi,n}$, with analytic, strictly increasing and sigmoidal activation function and $n \sim s^{d-1}$ neurons such that

$$\|f - N_{\phi,n}\| \leq C\mathrm{dist}(f, \Pi_s^d, C(\mathbf{S}^d)) + \varepsilon, \tag{3.1}$$

where $\|\cdot\|$ denotes the uniform norm on $\mathbf{S}^d$.

The following Lemma 3.1 proved by Maiorov and Pinkus (1999) will play a crucial role in our proof.

**Lemma 3.1.** *There exists a function $\phi$ which is real analytic, strictly increasing, and sigmoidal satisfying the following: for every $g \in C([-1, 1])$, there exists a sequence of natural number $\{n_k\}_{k=1}^\infty$ and three sequences of real number $\{a_{n_k}\}_{k=1}^\infty$, $\{b_{n_k}\}_{k=1}^\infty$ and $\{c_{n_k}\}_{k=1}^\infty$ such that*

$$\lim_{k\to\infty} \|g - a_{n_k}\phi(\cdot - 8n_k + 1) - b_{n_k}\phi(\cdot - 8n_k + 5) \\ - c_{n_k}\phi(\cdot + 4n_k + 1)\|_{C[-1,1]} = 0. \tag{3.2}$$

Now we give the main result of this section.

**Theorem 3.1.** *Let $s \in \mathbf{N}, n \sim s^{d-1}$. Then for any $f \in C(\mathbf{S}^d)$ and arbitrary $\varepsilon > 0$, there exists an SNN,*

$$N_{\phi,n}(x) = \sum_{k=0}^{n} c_k\phi(\langle w_k, x\rangle + b_k), \quad w_k \in \mathbf{R}^{d+1}, b_k, c_k \in \mathbf{R},$$

*where $\phi$ is analytic, strictly increasing, and sigmoidal such that*

$$\|f - N_{\phi,n}\| \leq C\mathrm{dist}(f, \Pi_s^d, C(\mathbf{S}^d)) + \varepsilon. \tag{3.3}$$

**Proof.** It follows from the representation formula (2.6) that for arbitrary $P_s \in \Pi_s^d$, there holds

$$P_s(x) = \sum_{i=1}^{d_s^d} g_i(\langle \xi_i, x\rangle), \tag{3.4}$$

where $g_j(j = 1, \ldots, d_s^d)$ are univariate polynomials defined on $[-1, 1]$ with degrees not larger than $s$ and $\{\xi_i\}_{i=1}^{d_s^d} \subset \mathbf{S}^d$. Then we use Lemma 3.1 to obtain (3.3). For arbitrary $\varepsilon > 0$, by (3.2), there exists an analytic, strictly increasing, and sigmoidal function $\phi$ and $k_0$ such that for any $k \geq k_0$, there holds

$$|g_j(t) - a^i_{n_k}\phi(t - 8n_k + 1) - b^i_{n_k}\phi(t - 8n_k + 5) - c^i_{n_k}\phi(t + 4n_k + 1)| \leq \frac{\varepsilon}{2d^d_s}. \quad (3.5)$$

We construct an SNNs, $N_{\phi,n}$, with $n = 3d^d_s \sim s^{d-1}$ neurons as

$$N_{\phi,n}(x) := \sum_{i=1}^{d^d_s} a^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle - 8n_{k_0} + 1) + b^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle - 8n_{k_0} + 5) + c^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle + 4n_{k_0} + 1). \quad (3.6)$$

Then by (3.4)–(3.6), we get

$$|P_s(x) - N_{\phi,n}(x)| = \left|\sum_{i=1}^{d^d_s} g_i(\langle\xi_i, x\rangle) - \sum_{i=1}^{d^d_s} a^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle - 8n_{k_0} + 1) - b^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle - 8n_{k_0} + 5) - c^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle + 4n_{k_0} + 1)\right|$$
$$\leq \sum_{i=1}^{d^d_s}\left|g_i(\langle\xi_i, x\rangle) - a^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle - 8n_{k_0} + 1) - b^i_{n_k}\phi(\langle\xi_i, x\rangle - 8n_{k_0} + 5) - c^i_{n_{k_0}}\phi(\langle\xi_i, x\rangle + 4n_{k_0} + 1)\right|$$
$$\leq \sum_{i=1}^{d^d_s}\frac{\varepsilon}{2d^d_s} = \frac{\varepsilon}{2}.$$

Thus, we obtain that for arbitrary $P_s \in \Pi^d_s$, there exists an SNN, $N_{\phi,n}$, with analytic, strictly increasing, sigmoidal activation function and $n \sim s^{d-1}$ neurons such that for arbitrary $\varepsilon > 0$, there holds

$$|P_s(x) - N_{\phi,n}(x)| \leq \frac{\varepsilon}{2}. \quad (3.7)$$

If we choose $P_s(x)$ satisfying

$$|f(x) - P_s(x)| \leq \mathrm{dist}(f, \Pi^d_s, C(\mathbf{S}^d)) + \frac{\varepsilon}{2},$$

then we have

$$|f(x) - N_{\phi,n}(x)| \leq |f(x) - P_s(x)| + |P_s(x) - N_{\phi,n}(x)| \leq \mathrm{dist}(f, \Pi^d_s, C(\mathbf{S}^d)) + \varepsilon.$$

This finishes the proof of Theorem 3.1. □

The above Theorem 3.1 uses the best approximation rate of SPs to describe the approximation rate of SNNs. It can be found that in order to obtain a prescribed approximation accuracy, the number of parameters needed in SNNs is much less than that of SPs. (Indeed the number of parameters needed in SNNs is asymptotically as $s^{d-1}$ and that of SPs is asymptotically as $s^d$.) Thus it can be seen that as far as the approximation capacity is concerned, SNNs are somewhat superior to SPs. The following three corollaries can be easily deduced from Theorem 3.1 and its proof, for the sake of brevity, we omit the details.

**Corollary 3.1.** *If the assumptions of Theorem 3.1 are fulfilled, then for any $f \in C(\mathbf{S}^d)$, there holds*

$$\mathrm{dist}(f, \Phi_{\phi,n}, C(\mathbf{S}^d)) \leq \mathrm{dist}(f, \Pi^d_s, C(\mathbf{S}^d)). \quad (3.8)$$

**Corollary 3.2.** *Let $s \in \mathbf{N}$, $n \sim s^{d-1}$. Then for any $f \in L^2(\mathbf{S}^d)$, there exists an analytic, strictly increasing, and sigmoidal function $\phi$ such that*

$$\mathrm{dist}(f, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) \leq \mathrm{dist}(f, \Pi^d_s, L^2(\mathbf{S}^d)).$$

The above two corollaries gave a comparison between approximation by SNNs and SPs under different metrics. In order to study the approximation rate of SNNs, we also need to introduce the approximation rate of SPs. The upper bound of approximation functions belonging to $W^2_{2r}$ by SPs has been established by Pawelke (1972), i.e.

$$\mathrm{dist}(f, \Pi^d_s, L^2(\mathbf{S}^d)) \leq Cs^{-2r}, \quad r \in \mathbf{N}. \quad (3.9)$$

This together with Corollary 3.2 yields the following Corollary 3.3.

**Corollary 3.3.** *Let $s \in \mathbf{N}$, $n \sim s^{d-1}$. Then for any $f \in W^2_{2r}$, there exists an analytic, strictly increasing, and sigmoidal function $\phi$ such that*

$$\mathrm{dist}(f, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) \leq \mathrm{dist}(f, \Pi^d_s, L^2(\mathbf{S}^d)) \leq Cn^{-\frac{2r}{d-1}}. \quad (3.10)$$

## 4. Lower bound of approximation

In this section, motivated by Maiorov (1999), we prove that the upper and lower rates of approximation by SNNs are identical, which behave asymptotically as $n^{-\frac{2r}{d-1}}$.

Let the vector set $E^m$ consisting of all vectors $\varepsilon := (\varepsilon_1, \ldots, \varepsilon_m)$, $m \in \mathbf{N}$ with coordinates $\varepsilon_1, \ldots, \varepsilon_m = \pm 1$, i.e.,

$$E^m := \{\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m) : \varepsilon_i = \pm 1, i = 1, 2, \ldots, m\}.$$

Let $m$, $s$, $p$ and $q$ be natural numbers. Let $\pi_{ij}(\sigma)$, $i = 1, \ldots, m; j = 1, \ldots, q$ be any algebraic polynomials with real coefficients in the variables $\sigma = (\sigma_1, \ldots, \sigma_p) \in \mathbf{R}^p$, each of degree $s$. Construct the polynomials in the $p + q$ variables $b = (b_1, \ldots, b_q) \in \mathbf{R}^q$ and $\sigma = (\sigma_1, \ldots, \sigma_p) \in \mathbf{R}^p$

$$\pi_i(b, \sigma) = \sum_{j=1}^{q} b_j\pi_{ij}(\sigma), \quad i = 1, \ldots, m.$$

Construct in $\mathbf{R}^m$ a polynomial manifold

$$\mathcal{P}_{m,s,p,q} := \{\pi(b, \sigma) = (\pi_1(b, \sigma), \ldots, \pi_m(b, \sigma)) : (b, \sigma) \in \mathbf{R}^q \times \mathbf{R}^p\}.$$

The following Lemma 4.1 was given in Maiorov (1999).

**Lemma 4.1.** *Let $m, p, q, s$ be integers such that $p + q \leq m/2$*

$$p\log_2(4s) + (p + 2)\log_2(p + q + 1) + (p + q)\log_2\left(\frac{2em}{p + q}\right) \leq \frac{m}{4}. \quad (4.1)$$

*Then there exists a vector $\epsilon \in E^m$ and an absolute constant $C > 0$ such that*

$$\mathrm{dist}(\epsilon, \mathcal{P}_{m,s,p,q}, l^2) \geq Cm^{1/2}. \quad (4.2)$$

Now we state our main result in this section.

**Theorem 4.1.** *Let $r \geq 0$, $d \geq 2$, then for any $\phi \in L^2(\mathbf{R})$ there exists a constant $C$ depending only on $d$ such that*

$$\mathrm{dist}(W^2_{2r}, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) \geq Cn^{-\frac{2r}{d-1}}. \quad (4.3)$$

**Proof.** Let $n$ and $s$ be any natural numbers. Set $m = d^{d+1}_s$. Thus there exists a constant $C_1 \leq 2$ such that

$$m = C_1 s^d. \quad (4.4)$$

Consider the set consisting of SPs

$$\mathcal{F}_{s,d} := \left\{ h(x) = \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \epsilon(j,i) Y_{j,i}(x) \right\},$$

where $\{\epsilon(j,i) : j = 0, \ldots, s, i = 1, \ldots, d_s^d\} \subset E^m$. It is obvious that $\mathcal{F}_{s,d}$ is a subset of $\Pi_s^d$, hence it satisfies the well-known Bernstein inequality (see Wang & Li, 2000)

$$\|\Delta^r h(\cdot)\|_2 \le C_2 s^{2r} \|h(\cdot)\|_2. \tag{4.5}$$

On the other hand, by Parseval equality we have

$$\|h(\cdot)\|_2^2 = \left\| \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \epsilon(j,i) Y_{j,i}(\cdot) \right\|_2^2 = \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} |\epsilon(j,i)|^2 = m.$$

The above equality together with (4.5) yields that

$$h^*(x) := \frac{1}{C_2 s^{2r} m^{\frac{1}{2}}} h(x) \in W_{2r}^2. \tag{4.6}$$

At first we estimate the distance from the set $\Phi_{\phi,n}$ to $\mathcal{F}_{s,d}$

$$\mathrm{dist}(\mathcal{F}_{s,d}, \Phi_{\phi,n}, L^2) = \sup_{h \in \mathcal{F}_{s,d}} \inf_{g \in \Phi_{\phi,n}} \|h(\cdot) - g(\cdot)\|_2.$$

Let

$$h(x) = \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \varepsilon(j,i) Y_{j,i}(x)$$

be an arbitrary function from $\mathcal{F}_{s,d}$ and let

$$g(x) = \sum_{k=1}^{n} c_k \phi(\langle w_k, x\rangle + b_k), \quad w_k \in \mathbf{R}^{d+1}, \; b_k, c_k \in \mathbf{R}$$

be an arbitrary function from $\Phi_{\phi,n}$. It is obvious that we can rewrite $g(x)$ as

$$g(x) = \sum_{k=1}^{n} c_k \phi(a_k \langle x_k, x\rangle + b_k), \quad x_k \in \mathbf{S}^d, \; a_k, b_k, c_k \in \mathbf{R}.$$

Since $\phi \in L^2(\mathbf{R})$, we have $g \in L^2(\mathbf{S}^d)$, then it follows from Parseval equality that

$$\|h(\cdot) - g(\cdot)\|_2^2 = \left\| \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \epsilon(j,i) Y_{j,i}(\cdot) - g(\cdot) \right\|_2^2$$

$$= \left\| \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \epsilon(j,i) Y_{j,i}(\cdot) - \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \langle g, Y_{j,i}\rangle Y_{j,i}(\cdot) - \sum_{j=s+1}^{\infty} \sum_{i=1}^{d_j^d} \langle g, Y_{j,i}\rangle Y_{j,i}(\cdot) \right\|_2^2$$

$$= \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} |\epsilon(j,i) - \langle g, Y_{j,i}\rangle_2|^2 + \sum_{j=s+1}^{\infty} \sum_{i=1}^{d_j^d} |\langle g, Y_{j,i}\rangle_2|^2$$

$$\ge \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} |\epsilon(j,i) - \langle g, Y_{j,i}\rangle_2|^2.$$

Fix indices $i, j$, and consider the inner product $\langle g, Y_{j,i}\rangle_2$. By Funk–Hecke formula (2.5), we have

$$\langle g, Y_{j,i}\rangle_2 = \sum_{k=1}^{n} \langle c_k \phi(a_k \langle x_k, \cdot\rangle + b_k), Y_{j,i}(\cdot)\rangle_2$$

$$= \sum_{k=1}^{n} \Omega_{d-1} \int_{-1}^{1} P_j^{d+1}(t) c_k \phi(a_k t + b_k)(1-t^2)^{\frac{d-2}{2}} \mathrm{d}t Y_{j,i}(x_k).$$

If we set $\sigma_k$ be the $(d+1) \times (d+1)$ orthogonal matrix with the corresponding determinant being 1 such that $x_k = \sigma_k \mathbf{e}$, $\mathbf{e} := (0, \ldots, 0, 1)$, and let

$$b_{k,l}(\phi) := \begin{cases} \Omega_{d-1} \int_{-1}^{1} P_j^{d+1}(t) c_k \phi(a_k t + b_k)(1-t^2)^{\frac{d-2}{2}} \mathrm{d}t, & l = j \\ 0, & l \ne j. \end{cases}$$

Then we have

$$\sum_{l=0}^{s} b_{k,l}(\phi) = \Omega_{d-1} \int_{-1}^{1} P_j^{d+1}(t) c_k \phi(a_k t + b_k)(1-t^2)^{\frac{d-2}{2}} \mathrm{d}t,$$

$j = 0, \ldots, s$.

Thus,

$$\langle g, Y_{j,i}\rangle_2 = \sum_{k=1}^{n} \sum_{l=0}^{s} b_{k,l}(\phi) Y_{j,i}(\sigma_k \mathbf{e}). \tag{4.7}$$

From (4.7) it follows that

$$\inf_{g \in \Phi_{\phi,n}} \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \left|\epsilon(j,i) - \langle g, Y_{j,i}\rangle_2\right|^2$$

$$= \inf_{b_{k,l}, \sigma_k} \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} \left| \epsilon(j,i) - \sum_{k=1}^{n} \sum_{l=0}^{s} b_{k,l}(\phi) Y_{j,i}(\sigma_k \mathbf{e}) \right|^2, \tag{4.8}$$

where the infimum is calculated over all collections of matrices $\sigma_1, \ldots, \sigma_n$, and all collections of $b_{1,0}, \ldots, b_{n,s}$. Obviously, the dimension of an $(d+1) \times (d+1)$ orthogonal matrix is $\frac{(d+1)(d+2)}{2}$, which together with the fact that $Y_{j,i}$ is an SP with degree $j$, $0 \le j \le s$ yields that $Y_{j,i}(\sigma_k e)$ is an algebraic polynomial with $\frac{(d+1)(d+2)}{2}$ variables and degree not larger than $s$. For the sake of brevity, we set

$$\pi_{j,i}(b, \sigma) = \sum_{k=1}^{n} \sum_{l=0}^{s} b_{k,l}(\phi) Y_{j,i}(\sigma_k \mathbf{e})$$

then $\pi_{j,i}(b, \sigma)$ is an algebraic with degree not larger than $s$, $b \in \mathbf{R}^{n(s+1)}$, $\sigma \in \mathbf{R}^{\frac{n(d+1)(d+2)}{2}}$. Therefore, we have

$$(\mathrm{dist}(\mathcal{F}_{n,d}, \Phi_{\phi,n}, L^2(\mathbf{S}^d)))^2$$

$$\ge \max_{\epsilon_{j,i} \in E^m} \inf_{b \in \mathbf{R}^{n(s+1)}, \sigma \in \mathbf{R}^{n(d+2)(d+1)/2}} \sum_{j=0}^{s} \sum_{i=1}^{d_j^d} |\epsilon_{j,i} - \pi_{j,i}(b,\sigma)|^2.$$

Since $m = d_s^{d+1}$ we can rearrange the sequence $\{(j,i), j = 0, \ldots, s, i = 1, \ldots, d_j^d\}$ as $\{k\}_{k=1}^m$. Thus

$$(\mathrm{dist}(\mathcal{F}_{n,d}, \Phi_{\phi,n}, L^2(\mathbf{S}^d)))^2$$

$$\ge \max_{\epsilon_k \in E^m} \inf_{b \in \mathbf{R}^{n(s+1)}, \sigma \in \mathbf{R}^{n(d+2)(d+1)/2}} \sum_{k=1}^{m} |\epsilon_k - \pi_k(b,\sigma)|^2. \tag{4.9}$$

Let $p = n(d+2)(d+1)/2$, $q = n(s+1)$ and

$$\mathcal{P}_{m,s,p,q} = \{\pi(b,\sigma) = (\pi_1(b,\sigma), \ldots, \pi_m(b,\sigma)) : (b,\sigma) \in \mathbf{R}^q \times \mathbf{R}^p\}.$$

Note that $m = C_1 s^d$. For $s \geq 4d^2$, we set $n = C_3 s^{d-1}$, where $C_3 := \frac{C_1}{2000ed^3}$. Then

$$
\begin{aligned}
p+q &= \frac{n(d+2)(d+1)}{2} + n(s+1) \\
&\leq 2n(s+1) \leq 2C_3 s^{d-1}(s+1) \leq \frac{C_1}{2} s^d = \frac{m}{2}, \\
p\log_2(4s) &= \frac{n(d+2)(d+1)}{2}\log_2(4s) \leq 2d^2 n \log_2(4s) \\
&\leq 8d^2 C_3 s^d \leq \frac{C_1 s^d}{12} = \frac{m}{12}, \\
(p+2)\log_2(p+q+1) &= \left(\frac{n(d+2)(d+1)}{2} + 2\right) \\
&\quad \times \log_2\left(\frac{n(d+2)(d+1)}{2} + n(s+1) + 1\right) \\
&\leq 5d^2 n \log_2(2d^2 n + 2ns) \leq 5d^2 C_3 s^{d-1}\log_2(4C_3 s^d) \\
&\leq 20 C_3^2 d^3 s^d \\
&\leq \frac{C_1 s^d}{12} = \frac{m}{12},
\end{aligned}
$$

and

$$
\begin{aligned}
(p+q)\log_2\left(\frac{2em}{p+q}\right) &= \left(\frac{n(d+2)(d+1)}{2} + n(s+1)\right)\log_2 \\
&\quad \times \left(\frac{2em}{\frac{n(d+2)(d+1)}{2} + n(s+1)}\right) \\
&\leq (4d^2 n + n(s+1)) 2\log_2\left(\frac{2eC_1 s^d}{ns}\right)^{\frac{1}{2}} \\
&\leq (8C_3 s^d)\frac{2e\sqrt{C_1}}{\sqrt{C_3}} \\
&\leq \frac{C_1 s^d}{12} = \frac{m}{12}.
\end{aligned}
$$

Thus the assumptions of Lemma 4.1 are fulfilled. Then from (4.2) we have

$$\mathrm{dist}(\mathcal{F}_{n,d}, \Phi_{\phi,n}, L^2) \geq Cm^{1/2} \geq Cs^{d/2}. \tag{4.10}$$

From (4.6) and (4.10), it follows that

$$
\begin{aligned}
\mathrm{dist}(W_{2r}^2, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) &\geq \mathrm{dist}(C_2^{-1} s^{-2r} m^{-\frac{1}{2}} \mathcal{F}_{n,d}, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) \\
&\geq Cs^{-2r-d/2}\mathrm{dist}(\mathcal{F}_{n,d}, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) \geq Cs^{-2r-d/2} \times s^{d/2} \\
&= Cs^{-2r} \sim n^{-2r/(d-1)}.
\end{aligned}
$$

This completes the proof of Theorem 4.1. □

## 5. Conclusions and remarks

In Section 3, we got the upper bound of approximation by SNNs with analytic, strictly increasing, sigmoidal activation function. In Section 4 we also deduce the lower bound of approximation by SNNs with square integrable activation function. Combining these we obtain the following Theorem 5.1.

**Theorem 5.1.** *If $n \sim s^{d-1}$, then there exists a $\phi : \mathbf{R} \to \mathbf{R}$ being analytic, strictly increasing, sigmoidal and square Lebesgue integrable such that*

$$\mathrm{dist}(W_{2r}^2, \Phi_{\phi,n}, L^2(\mathbf{S}^d)) \sim \mathrm{dist}(W_{2r}^2, \Pi_s^d, L^2(\mathbf{S}^d)) \sim n^{-\frac{2r}{d-1}}. \tag{5.1}$$

From Theorem 5.1 it follows the following two assertions. On the one hand, if the target functions satisfying some smoothness conditions, then the approximation rate of SNNs is faster than that of SPs. On the other hand, the approximation rate of SNNs we deduced cannot be improved. In other word, this rate is optimal. As shown above, the essential rate of approximation by SNNs and that by SPs are identical in $W_{2r}^2$ if the number of neurons $n$ and the degree of SPs $s$ satisfy $n \sim s^{d-1}$. Hence, as tools of dealing with spherical data, the SNNs constructed in this paper possess the following three advantages. The first one is that SNNs have superior approximation ability to SPs, since only $\mathcal{O}(s^{d-1})$ parameters are used in SNNs to get the approximation rate $\mathcal{O}\left(\frac{1}{s}\right)$ while which of SPs are $\mathcal{O}(s^d)$. The second one is that there exist some fast algorithms such as backprogram (BP) and extreme learning machine (ELM) to compute the parameters of SNNs. The third one is that SNNs are linear combinations of univariate functions, thus they provide a possibility to circumvent the curse of dimensionality.

We are also interested in finding a larger space than $W_{2r}^2$, which also satisfies (5.1). We consider a subset of $L^2(\mathbf{S}^d)$, $\mathcal{B}^{2r}$, which has some properties same as the Besov space. It was stated in Section 2 that $\{Y_{j,i}, i = 1, \ldots, d_j^d\}$ is orthogonal basis of $\mathbf{H}_j^d$. For any natural $N$ we denote the set of multi-indices $\Xi$ as

$$\Xi := \{(j,i) : j = 2^N + 1, \ldots, 2^{N+1}, i = 1, \ldots, d_j^d\}.$$

Introduce the subspace

$$\Psi_N := \mathrm{span}\{Y_{j,i} : (j,i) \in \Xi_N\}.$$

Let

$$G_N^{2r} := \left\{ \sum_{(j,i)\in\Xi_N} C(j,i) Y_{j,i} \in \Psi_N : \left(\sum_{(j,i)\in\Xi} |C(j,i)|^2\right)^{1/2} \leq 2^{-2rN} \right\}.$$

Then we define $\mathcal{B}^{2r}$ as

$$\mathcal{B}^{2r} := \left\{ f : f = \sum_{N=0}^{\infty} f_N, f_N \in G_N^r, N = 0, 1, \ldots \right\}.$$

For every $f \in \mathcal{B}^{2r}$ it is not difficult to prove that the class $\mathcal{B}^{2r}$ is essentially equivalent to the class $\mathcal{F}^r$, consisting of all functions $f$ for which the best approximation by SPs of degree $2^N$ satisfies the inequality

$$\mathrm{dist}(f, \Pi_{2^N}^d, L_2) \leq 2^{-2rN} \quad (N = 0, 1, \ldots).$$

Then by (3.1) we know that $W_{2r}^2 \subset \mathcal{B}^{2r}$, thus we have

$$\mathrm{dist}(W_{2r}^2, \Phi_{\phi,n}, L^2) \leq \mathrm{dist}(\mathcal{B}^{2r}, \Phi_{\phi,n}, L^2). \tag{5.2}$$

Moreover, from the definition of $\mathcal{B}^{2r}$, we obtain that

$$\mathrm{dist}(\mathcal{B}^{2r}, \Pi_s^d, L^2) \leq Cs^{-2r}. \tag{5.3}$$

Thus by (5.1)–(5.3) we obtain

**Theorem 5.2.** *Under the assumptions of Theorem 5.1, we have*

$$\mathrm{dist}(\mathcal{B}^{2r}, \Phi_{\phi,n}, L^2) \sim \mathrm{dist}(\mathcal{B}^{2r}, \Pi_s^d, L^2) \sim n^{-\frac{2r}{d-1}}. \tag{5.4}$$

The above theorems give some theoretical analysis of the SNN method. It has been shown that by using SNN, the approximation rate can be essentially improved. Now we turn to state some potential applications of the SNN method. It can be found in Freeden and Perevrzev (2001) and Freeden et al. (1998); Freeden, Glockner, and Thalhammer (1999); Freeden, Michel, and Nutz (2002) etc. that in order to study the gravity potential of the earth by the high–low GPS satellite and satellite tracking approach, it is sufficient to solve a spherical Fredholm integral equation of the first kind, i.e.

$$\int_{\mathbf{S}^d} f(y)k(x,y)\mathrm{d}\omega(y) = g(x), \tag{5.5}$$

where $k(x, y)$ is the radial kernel of the first kind (see Chapter 5 of Freeden et al. (1998) or Freeden et al. (1999)). By using the well-known collocation method with the assumption that the measurements are made in $\{x_j, g(x_j)\}_{j=1}^m$, ones need to solve the following system of equations

$$\int_{\mathbf{S}^d} f(y)k(x_j,y)\mathrm{d}\omega(y) = g(x_j), \quad j = 1, \ldots, m. \tag{5.6}$$

By the proposed SNN method, we can first approximates the target potential function $f(y)$ by SNN, i.e.

$$\sum_{i=1}^n c_i\phi(\langle w_i, y\rangle + \theta_i). \tag{5.7}$$

Thus the purpose of the high–low GPS satellite and satellite tracking mission is to compute the parameter of the SNN (5.7). It follows from (5.6) that the parameters of the SNN satisfy

$$\sum_{i=1}^n c_i \int_{\mathbf{S}^d} \phi(\langle w_i, y\rangle + \theta_i)k(x_j, y) = g(x_j), \quad j = 1, \ldots, m.$$

Then by using the well-known spherical cubature formula (Mhaskar et al., 1999), the parameters of the SNN satisfy

$$\sum_{i=1}^n c_i \sum_{k=1}^N \lambda_k c_i \phi(\langle w_i, z_k\rangle + \theta_i)k(x_j, z_k) \asymp g(x_j),$$
$$j = 1, \ldots, m, \tag{5.8}$$

where $\{z_k\}_{k=1}^N$ is the set of cubature points, $N$ is a natural number, $\{\lambda_k\}_{k=1}^N$ is the set of cubature weights and $a_k \asymp b$ denotes that $\lim_{k\to\infty} a_k = b$. All of cubature points and cubature weights can be deuced by some algorithms (Gia & Mhaskar, 2008). Therefore, the parameters of the SNN can be solved from (5.8), and the well-known ELM algorithm (Zhu, Qin, Suganthan, & Huang, 2005) can do it directly.

## Acknowledgments

The authors wish to thank the referees for their helpful suggestions. The research was supported by the National 973 Project (2007CB311002) and the National Natural Science Foundation of China (Nos. 90818020, 60873206).

## References

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, *39*, 930–945.
Bulsari, A. (1993). Some analytical solutions to the general approximation problem for feedforward neural networks. *Neural Networks*, *6*, 991–996.
Cao, F. L., Lin, S. B., & Xu, Z. B. (2010). The best approximation for polynomials and neural networks on the unit sphere. Manuscript.
Chen, T. P., & Chen, H. (1995). Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical system. *IEEE Transactions on Neural Networks*, *6*, 911–917.
Chui, C. K., & Li, X. (1992). Approximation by ridge functions and neural networks with one hidden layer. *Journal of Approximation Theory*, *70*, 131–141.
Cybenko, G. (1989). Approximation by superpositions of sigmoidal function. *Mathematics of Control, Signals, and Systems*, *2*, 303–314.
Dai, F., & Ditzian, Z. (2008). Jackson inequality for Banach spaces on the sphere. *Acta Mathematica Hungarica*, *118*(1–2), 171–195.
Ferrari, S., & Stengel, R. F. (2005). Smooth function approximation using neural networks. *IEEE Transactions on Neural Networks*, *16*, 24–38.
Freeden, W., Gervens, T., & Schreiner, M. (1998). *Constructive approximation on the sphere*. Oxford: Calderon Press.
Freeden, W., Glockner, O., & Thalhammer, M. (1999). Multiscale gravitational field recovery from GPS satellite-to-satellite tracking. *Studia Geophysica et Geodaetica*, *43*, 229–264.
Freeden, W., Michel, V., & Nutz, H. (2002). Satellite-to-satellite tracking and satellite gravity gradiometry (Advanced techniques for high-resolution geopotential field determination). *Journal of Engineering Mathematics*, *43*, 19–56.
Freeden, W., & Perevrzev, S. (2001). Spherical Tikhonov regularization wavelets in satellite gravity gradiometry with random noise. *Journal of Geometry*, *74*, 730–736.
Funahashi, K. I. (1989). On the approximate realization of continuous mapping by neural networks. *Neural Networks*, *2*, 183–192.
Gia, Q. T. Le, & Mhaskar, H. N. (2008). Localized linear polynomial operators and quadrature formulas on the spheres. *SIAM Journal on Numerical Analysis*, *47*, 440–466.
Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, *3*, 551–560.
Konovalov, V. N., Leviatan, D., & Maiorov, V. E. (2008). Approximation by polynomials and ridge functions of the classes of *s*-monotone radial functions. *Journal of Approximation Theory*, *152*, 20–51.
Konovalov, V. N., Leviatan, D., & Maiorov, V. E. (2009). Approximation of Sobolev classes by polynomials and ridge functions. *Journal of Approximation Theory*, *159*, 97–108.
Korain, P. (1993). On the complexity of approximating mapping using neural networks. *Neural Networks*, *6*, 449–653.
Leshno, M., Lin, V. Y., Pinks, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximat any function. *Neural Networks*, *6*, 861–867.
Lizorkin, I. P., & Nikol'skiĭ, S. M. (1983). A theorem concerning approximation on the sphere. *Analysis Mathematica*, *9*, 207–221.
Maiorov, V. E. (1999). On best approximation by ridge functions. *Journal of Approximation Theory*, *99*, 68–94.
Maiorov, V. E. (2003). On best approximation of classes by radial functions. *Journal of Approximation Theory*, *120*, 36–70.
Maiorov, V. E., & Meir, R. S. (1998). Approximation bounds for smooth function in $C(\mathbf{R}^d)$ by neural and mixture networks. *IEEE Transactions on Neural Networks*, *9*, 969–978.
Maiorov, V. E., & Pinkus, A. (1999). Lower bounds for approximation by MLP neural networks. *Neurocomputing*, *25*, 81–91.
Makovoz, Y. (1998). Uniform approximation by nerual networks. *Journal of Approximation Theory*, *95*, 215–228.
Mhaskar, H. N. (1996). Neural networks for optimal approximation of smooth and analytic functions. *Neural Computation*, *8*, 164–177.
Mhaskar, H. N. (2006). Weighted quadrature formulas and approximation by zonal function networks on the sphere. *Journal of Complexity*, *22*, 348–370.
Mhaskar, H. N., & Micchelli, C. A. (1995). Degree of approximation by neural networks with a single hidden layer. *Advances in Applied Mathematics*, *16*, 151–183.
Mhaskar, H. N., Narcowich, F. J., & Ward, J. D. (1999). Approximation properties of zonal function networks using scattered data on the sphere. *Advances in Computational Mathematics*, *11*, 121–137.
Mhaskar, H. N., Narcowich, F. J., & Ward, J. D. (2003). Zonal function network frames on the sphere. *Neural Networks*, *16*, 183–203.
Müller, C. (1966). *Lecture notes in mathematics*: *Vol. 17*. *Spherical harmonics*. Berlin: Springer.
Narcowich, F. J., Sun, X. P., Ward, J. D., & Wendland, H. (2007). Direct and inverse sobolev error estimates forscattered data interpolation via spherical basis functions. *Foundations of Computational Mathematics*, *7*, 369–370.
Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, *3*, 246–257.
Park, J., & Sandberg, I. W. (1993). Approximation and radial-basis-function networks. *Neural Computation*, *5*, 305–316.
Pawelke, S. (1972). Über die approximationsordnung bei kugelfunktionen und algebraischen polynomen. *The Tôhoku Mathematical Journal*, *24*, 473–486.
Petrushev, P. P. (1999). Approximation by ridge functions and neural networks. *SIAM Journal on Mathematical Analysis*, *30*, 155–189.
Sun, X. P., & Cheney, E. W. (1997). Fundamental sets of continuous functions on spheres. *Constructive Approximation*, *13*, 245–250.
Suzuki, S. (1998). Constructive function approximation by three-layer artificial neural networks. *Neural Networks*, *11*, 1049–1058.
Wang, K. Y., & Li, L. Q. (2000). *Harmonic analysis and approximation on the unit sphere*. Beijing: Science Press.
Xie, T. F., & Cao, F. L. (2010). The errors in simultaneous approximation by feed-forward neural networks. *Neurocomputing*, *73*, 903–907.
Xu, Z. B., & Cao, F. L. (2004). The essential order of approximation for neural networks. *Science in China. Series F*, *47*, 97–112.
Zhu, Y. Q., Qin, A. K., Suganthan, P. N., & Huang, G. B. (2005). Evolutionary extreme learning machine. *Pattern Recognition*, *38*, 1759–1763.