ANZIAM J. 58(2017), 238–246
doi:10.1017/S1446181117000104

APPROXIMATION BY SPHERICAL NEURAL NETWORKS WITH ZONAL FUNCTIONS

ZHIXIANG CHEN[1] and FEILONG CAO[✉2]

(Received 14 July, 2016; accepted 18 November, 2016; first published online 26 April 2017)

Abstract

We address the construction and approximation for feed-forward neural networks (FNNs) with zonal functions on the unit sphere. The filtered de la Vallée-Poussin operator and the spherical quadrature formula are used to construct the spherical FNNs. In particular, the upper and lower bounds of approximation errors by the FNNs are estimated, where the best polynomial approximation of a spherical function is used as a measure of approximation error.

2010 *Mathematics subject classification*: 41A25.

Keywords and phrases: approximation, neural networks, zonal function, error.

1. Introduction

Feed-forward neural networks (FNNs) with single hidden layer are a class of basic neural networks, which can be described mathematically as

$$\mathcal{N}(x) = \sum_{i=1}^{N} c_i \phi(\omega_i \cdot x + \theta_i), \tag{1.1}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is the activation function, $x = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$ is the input, $c_i \in \mathbb{R}$ $(i = 1, \dots, N)$ are the output weights connecting the N nodes, $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in}) \in \mathbb{R}^n$ are the input weights connecting the ith hidden node and the input and $\theta_i \in \mathbb{R}$ $(i = 1, \dots, N)$ are the biases of hidden nodes.

It is well known that FNNs are universal approximators. In many applications in geophysics and metrology, the data are collected over the surface of the Earth by satellite or ground stations. Naturally, the FNNs as in (1.1), defined on the sphere, should be a powerful tool for solving this problem [2, 6, 7, 9]. There are usually

[1]Department of Mathematics, Shaoxing University, Shaoxing 312000, Zhejiang Province, PR China; e-mail: czx@usx.edu.cn.
[2]Department of Applied Mathematics, China Jiliang University, Hangzhou 310018, Zhejiang Province, PR China; e-mail: feilongcao@gmail.com.

two kinds of activation functions used to construct FNNs for the approximation of spherical functions. One is the sigmoidal function [1, 7] and the other is the spherical radial basis function, also called the zonal function, which can be derived from a real function ϕ defined on the interval $[-1, 1]$. Recently, there have been a lot of results about approximation by spherical positive-definite radial basis functions, because this kind of network has the property of interpolation to the target function and has been proved to be very applicable in data fitting, especially in the case of scattered data (see [7, 9] and the references therein). Mhaskar et al. [9] considered another kind of radial basis function, for which none of the Fourier–Legendre coefficients of ϕ is zero, and they discussed the approximation of networks derived from ϕ for functions defined on the unit sphere. In this paper, we further study the construction and approximation of this kind of spherical neural network.

This paper first uses the filtered de la Vallée-Poussin operator and a spherical quadrature formula to construct the explicit FNNs on the unit sphere. Then an upper bound of error is obtained; especially, a lower bound of approximation error for the networks is estimated by using a constructive method.

2. Preliminaries

Let $\mathbb{S}^2 = \{x \in \mathbb{R}^3 \mid |x|_2 = 1\}$ be the two-dimensional unit sphere embedded in $\mathbb{R}^3$, where $|\cdot|_2$ denotes the Euclidean norm. The surface measure on $\mathbb{S}^2$ is denoted by μ and we assume that it is normalized such that $\int_{\mathbb{S}^2} d\mu = 4\pi$. The space $L^2 = L^2(\mathbb{S}^2)$ is the usual Hilbert space of square-integral functions on $\mathbb{S}^2$ with the inner product $(f, g) = \int_{\mathbb{S}^2} f(x)g(x)\, d\mu(x)$ and the norm $\|f\|_2 = \sqrt{(f, f)}$. The space of continuous functions on $\mathbb{S}^2$ is denoted by $C(\mathbb{S}^2)$, which is a Banach space with respect to the supremum norm $\|f\|_\infty = \sup_{x\in\mathbb{S}^2} |f(x)|$.

For an integer $l \geq 0$, the restriction to $\mathbb{S}^2$ of a homogeneous harmonic polynomial of degree l is called a *spherical harmonic* of degree l. The class of all spherical harmonics of degree l is denoted by H_l and the class of all spherical harmonics of degree $l \leq n$ is denoted by Π_n. The spaces H_l are mutually orthogonal and, obviously, $\Pi_n = \bigoplus_{l=0}^{n} H_l$. The dimension of H_n is $2n + 1$ and that of Π_n is $\sum_{l=0}^{n}(2l + 1) = (n + 1)^2$. It is well known that $L^2(\mathbb{S}^2) = \text{closure} \bigoplus_{l=0}^{\infty} H_l$. For any $l \in \mathbb{N}$, the set $\{Y_{l,k} \mid k = 1, 2, \ldots, 2l + 1\}$ denotes a real L^2-orthonormal basis of H_l. And, $\{Y_{l,k} \mid l = 0, 1, \ldots, n, \quad k = 1, 2, \ldots, 2l + 1\}$ forms an L^2-orthonormal basis of Π_n. We have the well-known addition formula

$$\sum_{k=1}^{2l+1} Y_{l,k}(x)Y_{l,k}(y) = \frac{2l + 1}{4\pi} P_l(x \cdot y),$$

where $x \cdot y$ denotes the usual inner product on $\mathbb{R}^3$, and P_l is the Legendre polynomial with the degree l with $P_l(1) = 1$. If $\phi \in L^1[-1, 1]$ and $Y_l \in H_l$, then we have the following Funk–Hecke formula [5]:

$$\int_{\mathbb{S}^2} \phi(x \cdot y)Y_l(y)\, d\mu(y) = 2\pi Y_l(x) \int_{-1}^{1} \phi(t)P_l(t)\, dt = 2\pi\hat{\phi}(l)Y_l(x). \tag{2.1}$$

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

Since the set $\{Y_{l,k}\, l = 0, 1, \ldots ; k = 1, 2, \ldots , 2l + 1\}$ is an orthonormal basis for $L^2(\mathbb{S}^2)$, any function $f \in L^2$ can be expanded into a Fourier–Laplace series with respect to this orthonormal system in the L^2 sense, that is, $f(x) = \sum_{l=0}^{\infty} \sum_{k=1}^{2l+1} \hat{f}_{l,k} Y_{l,k}(x)$, where the Fourier coefficients are given by

$$\hat{f}_{l,k} = (f, Y_{l,k}) = \int_{\mathbb{S}^2} f(x)Y_{l,k}(x)\, d\mu(x).$$

Set $K_l(x, y) = K_l(x \cdot y) = \sum_{k=1}^{2l+1} Y_{l,k}(x)Y_{l,k}(y) = \{(2l + 1)/4\pi\}P_l(x \cdot y)$,

$$h(t) = \begin{cases} 1, & x \in [0, 1), \\ 1 - 2(x - 1)^2, & x \in [1, 3/2), \\ 2(2 - x)^2, & x \in [3/2, 2), \\ 0, & x \in [2, \infty) \end{cases}$$

and introduce the operator

$$V_L f(x) = \int_{\mathbb{S}^2} f(y) \sum_{l=0}^{2L} h\Big(\frac{l}{L}\Big) K_l(x \cdot y)\, d\mu(y) = \sum_{l=0}^{2L} h\Big(\frac{l}{L}\Big) \sum_{k=1}^{2l+1} \hat{f}_{l,k} Y_{l,k}(x). \tag{2.2}$$

Then V_L is a de la Vallée-Poussin operator [10]. As Sloan [10] has pointed out, note that the sequence of operators $V_1, V_2, \ldots , V_L, \ldots$ has the properties: (A) V_L reproduces polynomials with degree up to L, that is, $V_L p = p$ for all $p \in \Pi_L$; (B) the linear operator sequence $V_1, V_2, \ldots$ is bounded uniformly.

If $\phi(t) \in L^1[-1, 1]$, then ϕ has the Fourier–Legendre polynomial expansions

$$\sum_{k=0}^{\infty} \hat{\phi}(k)P_k(t), \tag{2.3}$$

where $\hat{\phi}(k) = \int_{-1}^{1} \phi(t)P_k(t)\, dt$. Let $S_n(\phi) = S_n(\phi, t)$ denote the nth partial sum of (2.3); then the de la Vallée-Poussin operator of ϕ is defined as $h_N(\phi, t) = \{1/(N + 1)\} \sum_{v=N}^{2N} S_v(\phi, t)$, which has the properties [3]: (C) $\widehat{h_N}(\phi)(l) = \hat{\phi}(l), l = 0, 1, 2, \ldots , N$; (D) if $\phi(t) \in C[-1, 1]$, then there exists an absolute positive constant C such that $|h_N(\phi, t) - \phi(t)| \le CE_N(\phi)$, where $E_N(\phi)$ is the Nth best approximation of ϕ in $C[-1, 1]$.

3. The estimates of approximation errors

From the Funk–Hecke formula (2.1) and the property (C) of operator $h_N(\phi)$,

$$Y_{l,k}(x) = \frac{1}{2\pi\widehat{h_N}(\phi)(l)} \int_{\mathbb{S}^2} h_N(\phi, x \cdot y)Y_{l,k}(y)\, d\mu(y) = \frac{1}{2\pi\hat{\phi}(l)} \int_{\mathbb{S}^2} h_N(\phi, x \cdot y)Y_{l,k}(y)\, d\mu(y),$$

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

where $l = 0, 1, \ldots, N$ and $\hat{\phi}(l) \neq 0$. So, from (2.2), it follows that

$$\begin{aligned}
V_{[N/2]}(f, x) &= \sum_{l=0}^{2[N/2]} \sum_{k=1}^{2l+1} \widehat{V_{[N/2]}}(f)(l, k) Y_{l,k}(x) \\
&= \frac{1}{2\pi} \sum_{l=0}^{2[N/2]} \frac{1}{\hat{\phi}(l)} \sum_{k=1}^{2l+1} \int_{\mathbb{S}^2} V_{[N/2]}(f, u) Y_{l,k}(u)\, d\mu(u) \int_{\mathbb{S}^2} h_N(\phi, x \cdot y) Y_{l,k}(y)\, d\mu(y) \\
&= \frac{1}{2\pi} \sum_{l=0}^{2[N/2]} \frac{1}{\hat{\phi}(l)} \int_{\mathbb{S}^2} h_N(\phi, x \cdot y) \Big(\int_{\mathbb{S}^2} V_{[N/2]}(f, u) \sum_{k=1}^{2l+1} Y_{l,k}(u) Y_{l,k}(y)\, d\mu(u) \Big) d\mu(y) \\
&= \frac{1}{8\pi^2} \sum_{l=0}^{2[N/2]} \frac{2l+1}{\hat{\phi}(l)} \int_{\mathbb{S}^2} h_N(\phi, x \cdot y) Y_l(V_{[N/2]}(f), y)\, d\mu(y), \qquad (3.1)
\end{aligned}$$

where $Y_l(V_{[N/2]}(f), y)$ denotes the projections of $V_{[N/2]}(f)$ onto H_l and $[a]$ denotes the largest integer not greater than a.

Applying the quadrature formula [4, Theorem 5.1] to equation (3.1),

$$V_{[N/2]}(f, x) = \frac{1}{8\pi^2} \sum_{l=0}^{2[N/2]} \frac{2l+1}{\hat{\phi}(l)} \sum_{\xi(l)} a_{\xi(l)} h_N(\phi, x \cdot \xi(l)) Y_l(V_{[N/2]}(f), \xi(l)),$$

where $\xi(l)$ denotes some point on the unit sphere, which changes with l, and $a_{\xi(l)}$ are nonnegative numbers. Now we can construct spherical FNNs as

$$T_N(f, x) = \frac{1}{8\pi^2} \sum_{l=0}^{2[N/2]} \frac{2l+1}{\hat{\phi}(l)} \sum_{\xi(l)} a_{\xi(l)} \phi(x \cdot \xi(l)) Y_l(V_{[N/2]}(f), \xi(l)).$$

Then the following theorem, which is related with the upper bound estimate for the constructed zonal function networks, can be established.

THEOREM 3.1. *Let $\phi \in C[-1, 1]$ and the Fourier–Legendre coefficients $\hat{\phi}(l) \neq 0$, $l = 0, 1, \ldots$. If $f \in C(\mathbb{S}^2)$, then*

$$|T_N(f, x) - f(x)| \leq C\Big(E_{[N/2]}(f) + E_N(\phi)(N+1) \max_{0 \leq l \leq N} \frac{1}{|\hat{\phi}(l)|} \|f\|^2\Big).$$

PROOF. Note that

$$|T_N(f, x) - f(x)| \leq |T_N(f, x) - V_{[N/2]}(f, x)| + |V_{[N/2]}(f, x) - f(x)| =: I_1 + I_2.$$

On one hand, from [10], it can be obtained that $I_2 \leq C E_{[N/2]}(f)$. Here and in the following, C denotes an absolute positive constant, and its value may be different at different occurrences, even within the same formula.

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

On the other hand,

$$\begin{aligned} I_1 &\le \frac{1}{8\pi^2} \sum_{l=0}^{2[N/2]} \frac{2l+1}{|\hat{\phi}(l)|} \sum_{\xi(l)} |a_{\xi(l)}||h_N(\phi, x\cdot\xi(l)) - \phi(x\cdot\xi(l))||Y_l(V_{[N/2]}(f), \xi(l))| \\ &\le CE_N(\phi) \sum_{l=0}^{2[N/2]} \frac{2l+1}{|\hat{\phi}(l)|} \sum_{\xi(l)} |a_{\xi(l)}||Y_l(V_{[N/2]}(f), \xi(l))| \\ &\le CE_N(\phi)(N+1) \max_{0\le l\le N} \frac{1}{|\hat{\phi}(l)|} \sum_{l=0}^{2[N/2]} \sum_{\xi(l)} |a_{\xi(l)}||Y_l(V_{[N/2]}(f), \xi(l))|. \end{aligned}$$

Using a result of Filbir and Themistoclakis [4, (5.4) in Theorem 5.1] ($p = \infty$) and the L^2 Marcinkiewicz–Zygmund inequality [4, Theorem 4.2],

$$\begin{aligned} \sum_{\xi(l)} |a_{\xi(l)}||Y_l(V_{[N/2]}(f), \xi(l))| &\le \sum_{\xi(l)} |a_{\xi(l)}| \sum_{\xi(l)} |a_{\xi(l)}||Y_l(V_{[N/2]}(f), \xi(l))|^2 \\ &= 4\pi \sum_{\xi(l)} |a_{\xi(l)}||Y_l(V_{[N/2]}(f), \xi(l))|^2 \\ &\le C \int_{\mathbb{S}^2} \bigg| \sum_{k=1}^{2l+1} \hat{f}_{l,k} Y_{l,k}(x) \bigg|^2 d\mu(x) \\ &= C \sum_{k=1}^{2l+1} |\hat{f}_{l,k}|^2. \end{aligned}$$

Therefore, $I_1 \le CE_N(\phi)(N+1)\max_{0\le l\le N}(1/|\hat{\phi}(l)|)\|f\|^2$. Hence,

$$|T_N(f,x) - f(x)| \le C\Big(E_{[N/2]}(f) + E_N(\phi)(N+1) \max_{0\le l\le N} \frac{1}{|\hat{\phi}(l)|}\|f\|^2\Big).$$

This completes the proof of Theorem 3.1. □

Let $\mathbb{E}^m$ denote a vector set, which consists of all vectors $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)$ ($m \in \mathbb{N}$) with coordinates $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m = \pm 1$, that is, $\mathbb{E}^m = \{\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m) \mid \varepsilon_i = \pm 1, i = 1, 2, \dots, m\}$. Let m, s, p, q be natural numbers and π_{ij} ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, q$) be any algebraic polynomials with real coefficients in the variables $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_p) \in \mathbb{R}^p$ and each of degree s. Now we construct polynomials with variables $\mathbf{b} = (b_1, b_2, \dots, b_q) \in \mathbb{R}^q$ and $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_p) \in \mathbb{R}^p$ such that $\pi_i(\mathbf{b}, \sigma) = \sum_{j=1}^q b_j \pi_{ij}(\sigma)$, $i = 1, 2, \dots, m$. Also, we construct a polynomial manifold in $\mathbb{R}^m$:

$$\mathbb{P}_{m,s,p,q} = \{\pi(\mathbf{b}, \sigma) = (\pi_1(\mathbf{b}, \sigma), \pi_2(\mathbf{b}, \sigma), \dots, \pi_m(\mathbf{b}, \sigma)) \mid (\mathbf{b}, \sigma) \in \mathbb{R}^q \times \mathbb{R}^p\}.$$

To prove our main result in this section, we need the following lemma [8, Theorem 4].

Lemma 3.2. *Let m, p, q, s be integers such that*

$$p + q \le \frac{m}{2} \tag{3.2}$$

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

and

$$p\log_2 4s + (p+2)\log_2(p+q+1) + (p+q)\log_2\left(\frac{2em}{p+q}\right) \le \frac{m}{4}. \tag{3.3}$$

Then there exist a vector $\varepsilon = (\bar{\varepsilon}_1, \bar{\varepsilon}_2, \dots, \bar{\varepsilon}_m) \in \mathbb{E}^m$ *and a constant* $C > 0$ *such that* $\mathrm{dist}(\varepsilon, \mathbb{P}_{m,s,p,q}, l^2) \ge Cm^{1/2}$, *where*

$$\mathrm{dist}(\varepsilon, \mathbb{P}_{m,s,p,q}, l^2) = \inf_{\pi(\mathbf{b},\sigma)\in\mathbb{P}_{m,s,p,q}} \left\{\sum_{j=1}^{m}(\bar{\varepsilon}_j - \pi_j(\mathbf{b},\sigma))^2\right\}^{1/2}.$$

Now we also construct two sets of functions:

$$\mathbb{M} = \{f \in C(\mathbb{S}^2), \|f\|_\infty \le 1\} \tag{3.4}$$

and

$$\mathbb{F}_s = \left\{h \mid h(x) = \sum_{i=0}^{s}\sum_{j=1}^{2i+1}\epsilon(i,j)Y_{i,j}(x)\right\},$$

where $\{\epsilon(i,j) \mid i = 0,1,\dots,s;\ j = 1,2,\dots,2i+1\} \subset \mathbb{E}^{(s+1)^2}$. Then, for any h, the estimate [5, inequality (3.1.4)], $\sup_{x\in\mathbb{S}^2}|Y_{i,j}(x)| \le \{(2i+1)/4\pi\}^{1/2}$, implies that

$$\|h\|_\infty \le \sum_{i=0}^{s}\sum_{j=1}^{2i+1}\max_{x\in\mathbb{S}^2}|Y_{i,j}(x)| \le (2s+3)^{5/2} \le C_0 s^{5/2},$$

so that $h^*(x) = \{1/C_0 s^{5/2}\}h(x) \in \mathbb{M}$. In addition, we denote the set $\Phi_{\phi,n}$ as

$$\Phi_{\phi,n} = \left\{N_{\phi,n} \mid N_{\phi,n}(x) = \sum_{i=1}^{n} c_i\phi(\omega_i\cdot x), x, \omega_i \in \mathbb{S}^2\right\} \tag{3.5}$$

and define the distance of two function sets W from H (with $W, H \subset C(\mathbb{S}^2)$) by

$$\mathrm{dist}(W, H, C(\mathbb{S}^2)) = \sup_{f\in W}\mathrm{dist}(f, H, C(\mathbb{S}^2)) = \sup_{f\in W}\inf_{h\in H}\|f-h\|_\infty.$$

Note that for $h \in \mathbb{F}_s$ and $g \in C(\mathbb{S}^2)$,

$$\begin{aligned}
\|h-g\|_\infty^2 &\ge \frac{1}{4\pi}\int_{\mathbb{S}^2}|h(x)-g(x)|^2\,d\mu(x)\\
&= \frac{1}{4\pi}\left\|\sum_{i=0}^{s}\sum_{j=1}^{2i+1}\epsilon(i,j)Y_{i,j}(\cdot) - g(\cdot)\right\|_2^2\\
&= \frac{1}{4\pi}\left\|\sum_{i=0}^{s}\sum_{j=1}^{2i+1}\epsilon(i,j)Y_{i,j}(\cdot) - \sum_{i=0}^{\infty}\sum_{j=1}^{2i+1}(g, Y_{i,j})Y_{i,j}(\cdot)\right\|_2^2\\
&\ge \frac{1}{4\pi}\left(\sum_{i=0}^{s}\sum_{j=1}^{2i+1}|\epsilon(i,j)-(g,Y_{i,j})|^2 + \sum_{i=s+1}^{\infty}\sum_{j=1}^{2i+1}|(g,Y_{i,j})|^2\right)\\
&\ge \frac{1}{4\pi}\sum_{i=0}^{s}\sum_{j=1}^{2i+1}|\epsilon(i,j)-(g,Y_{i,j})|^2.
\end{aligned}$$

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

By the Funk–Hecke formula (2.1),

$$(g, Y_{i,j}) = \sum_{k=1}^{n} c_k(\phi((\omega_k, \cdot)), Y_{i,j}(\cdot)) = \sum_{k=1}^{n} c_k \int_{\mathbb{S}^2} \phi(\omega_k \cdot x) Y_{i,j}(x)\, d\mu(x)$$
$$= 2\pi \sum_{k=1}^{n} c_k \int_{-1}^{1} \phi(t) P_i(t)\, dt Y_{i,j}(\omega_k).$$

Let σ_k be orthogonal matrices from the group $SO(3)$ [11] for which $\omega_k = \sigma_k e$ with $e = (1, 0, 0)$ and let

$$b_{k,l}(\phi) = \begin{cases} 2\pi c_k \int_{-1}^{1} \phi(t) P_i(t)\, dt, & l = i, \\ 0, & l \neq i. \end{cases}$$

Then $(g, Y_{i,j}) = \sum_{k=1}^{n} \sum_{l=0}^{s} b_{k,l}(\phi) Y_{i,j}(\sigma_k e)$, from which it follows that

$$\inf_{g \in \Phi_{\phi,n}} \sum_{i=0}^{s} \sum_{j=1}^{2i+1} |\epsilon(i, j) - (g, Y_{i,j})|^2 = \inf_{b_{k,l}, \sigma_k} \sum_{i=0}^{s} \sum_{j=1}^{2i+1} \left| \epsilon(i, j) - \sum_{k=1}^{n} \sum_{l=0}^{s} b_{k,l}(\phi) Y_{i,j}(\sigma_k e) \right|,$$

where the infimum is calculated over all collections of matrices $\sigma_1, \sigma_2, \ldots, \sigma_n \in SO(3)$ and $b_{1,0}, b_{1,1}, \ldots, b_{n,s}$. We set $p = 9n, q = n(s + 1)$ and

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

then $\pi_{i,j}(\mathbf{b}, \sigma)$ is a polynomial in the variables $\mathbf{b} = (b_1, b_2, \ldots, b_q) \in \mathbb{R}^q$ and $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_p) \in \mathbb{R}^p$, each of degree not larger than s. So,

$$\{\mathrm{dist}(\mathbb{F}_s, \Phi_{\phi,n}, C(\mathbb{S}^2))\}^2 \geq \max_{\epsilon(i,j) \in \mathbb{E}^{(s+1)^2}} \inf_{\mathbf{b} \in \mathbb{R}^q, \sigma \in \mathbb{R}^p} \sum_{i=0}^{s} \sum_{j=1}^{2i+1} |\epsilon(i, j) - \pi_{i,j}(\mathbf{b}, \sigma)|^2$$
$$= \max_{\epsilon_k \in \mathbb{E}^m} \inf_{\mathbf{b} \in \mathbb{R}^q, \sigma \in \mathbb{R}^p} \sum_{k=1}^{m} |\epsilon_k - \pi_k(\mathbf{b}, \sigma)|.$$

For given $n \in \mathbb{N}$, $p = 9n$, $q = n(s + 1)$, $m = (s + 1)^2$ and $s = 2^{10} n + 31$, we observe that p, q, m and s satisfy the conditions (3.2) and (3.3) of Lemma 3.2 (see the Appendix for details) and thus we get $\{\mathrm{dist}(\mathbb{F}_s, \Phi_{\phi,n}, C(\mathbb{S}^2))\}^2 \geq C s^2$. Therefore,

$$\{\mathrm{dist}(\mathbb{M}, \Phi_{\phi,n}, C(\mathbb{S}^2))\}^2 \geq \left\{ \mathrm{dist}\left(\frac{1}{C_0 s^{5/2}} \mathbb{F}_s, \Phi_{\phi,n}, C(\mathbb{S}^2) \right) \right\}^2 \geq C s^{-1/2} \geq C n^{-1/2},$$

that is, $\mathrm{dist}(\mathbb{M}, \Phi_{\phi,n}, C(\mathbb{S}^2)) \geq C n^{-1/4}$, which proves the following main result of this section.

THEOREM 3.3. *For a given $n \in \mathbb{N}$, for the sets of functions $\mathbb{M}$ and $\Phi_{\phi,n}$ defined as in (3.4) and (3.5), respectively, we have* $\mathrm{dist}(\mathbb{M}, \Phi_{\phi,n}, C(\mathbb{S}^2)) \geq C n^{-1/4}$.

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

4. Conclusion

Neural networks have become an important method in the research areas of numerical analysis, machine learning, big data analytics and artificial intelligence. In this paper, we studied the construction and approximation of spherical networks with zonal functions, and obtained some theoretical results on the approximate capability of networks, which may provide support for the design of networks in real applications. Since the networks in this paper are determined by the input and output weights besides the activation function, the algorithms for the weights and the optimal choice of activation function are the issues we intend to study in a future work.

Appendix

We recall that $p = 9n, q = n(s+1), m = (s+1)^2$ and n is a given integer. So, we can choose s such that

$$\frac{m}{2^{11}} \le p + q = n(s+10) \le \frac{m}{2^{10}}. \tag{A.1}$$

Now we consider the equation $10n + ns = (s^2 + 2s + 1)/x$ satisfied by the variable s. Clearly, the roots of this equation are

$$s_{1,2} = \frac{nx - 2 \pm \sqrt{(2-nx)^2 - 4(1-10nx)}}{2}.$$

Setting $x = 2c$, $s_{1,2} = nc - 1 \pm \sqrt{n^2c^2 + 18nc}$. Considering $c = 2^9, 2^{10}$, respectively, and recalling the inequalities

$$s \ge 2^9 n - 1 + 2^5 \sqrt{2^8 n^2 + 9n} \tag{A.2}$$

and

$$s \le 2^{10} n - 1 + 2^5 \sqrt{2^{10} n^2 + 18n}, \tag{A.3}$$

we choose $s = 2^{10}n + 31$. Obviously, s satisfies (A.2) and (A.3). Thus, $m = (2^{10}n + 2^5)^2$ and we have validated (3.2) for the above m, s, p, q.

From (A.1),

$$(p+q)\log_2\left(\frac{2em}{p+q}\right) \le \frac{m}{2^{10}} \log_2(2^3 \cdot 2^{11}) \le \frac{m}{2^6}. \tag{A.4}$$

We estimate the term $(p+2)\log_2(p+q+1)$ in (3.3) below. Since

$$(p+2)\log_2(p+q+1) \le 2p\log_2 2(p+q) \le 2p\log_2\left(2 \cdot \frac{m}{2^{10}}\right) = 18n\log_2 \frac{m}{2^9},$$

we prove that $18n\log_2(2^{m/2^9}) \le m/16$, that is, $m/2^9 \le 2^{m/(16\times 18n)}$ or $m \le 2^9 \cdot 2^{m/(16\times 18n)}$. Setting $m = (2^{10}n + 2^5)^2$, the above inequality becomes

$$(2^{10}n + 2^5)^2 \le 2^9 \cdot 2^{(2^{10}n+2^5)^2/(16\times 18n)}. \tag{A.5}$$

https://doi.org/10.1017/S1446181117000104 Published online by Cambridge University Press

To obtain (A.5), we only need to prove that $2^{22}n^2 \leq 2^9 \cdot 2^{(2^{20}n^2)/(16\times 18n)}$, that is, $2^{117}n^{18} \leq 2^{2^{15}n}$. Since $2^{14}n \geq 117$, we have $2^{15}n - 117 \geq 2^{14}n$. So, it follows that the inequality $n^{18} \leq 2^{2^{14}n}$ leads to $2^{117}n^{18} \leq 2^{2^{15}n}$. Therefore,

$$(p+2)\log_2(p+q+1) \leq \frac{m}{16}. \tag{A.6}$$

Finally, we estimate $p\log_2 4s$. We will prove that

$$p\log_2 4s \leq \frac{m}{16}. \tag{A.7}$$

Since $p = 9n, m = (2^{10}n + 2^5)^2$ and $s = 2^{10}n + 31$, inequality (A.7) is equivalent to $4(2^{10}n + 31) \leq 2^{(2^{10}n+2^5)^2/(16\times 9n)}$ and we only need to get the inequality $4 \cdot 2^{11}n \leq 2^{(2^{20}n^2)/(16\times 9n)}$, that is,

$$2^{117}n^9 \leq 2^{2^{16}n}. \tag{A.8}$$

Similarly, inequality (A.8) holds, which yields (A.7). Hence, from (A.1), (A.4), (A.6) and (A.7), we observe that the inequalities (3.2) and (3.3) hold for given n and m, s, p, q in Lemma 3.2.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant nos. 61672477, 11401388 and 61640316.

References

[1] F. Cao and Z. Chen, "Approximation by spherical neural networks with sigmoidal function", *J. Comput. Anal. Appl.* **18** (2015) 390–396; https://scholar.google.com.au/scholar?cluster=16133814104893368681&hl=en&as_sdt=0,5&as_vis=1.

[2] F. Cao and S. Lin, "The capability of approximation for neural networks interpolant on the sphere", *Math. Methods Appl. Sci.* **34** (2011) 469–478; doi:10.1002/mma.1373.

[3] C. F. Dunkl and Y. Xu, "Orthogonal polynomials of several variables", in: *Encyclopedia of Mathematics and its Applications* (Cambridge University Press, Cambridge, 2001).

[4] F. Filbir and W. Themistoclakis, "Polynomial approximation on the sphere using scattered data", *Math. Nachr.* **281** (2008) 650–668; doi:10.1002/mana.200710633.

[5] W. Freeden, T. Gervens and M. Schreiner, *Constructive approximation on the sphere* (Oxford University Press, New York, 1998).

[6] S. Lin and F. Cao, "Simultaneous approximation by spherical neural networks", *Neurocomputing* **175** (2016) 348–354; doi:10.1016/j.neucom.2015.10.067.

[7] S. Lin, J. Zeng and Z. Xu, "Error estimate for spherical neural networks interpolation", *Neural Process. Lett.* **42** (2015) 369–379; doi:10.1007/s11063-014-9361-x.

[8] V. E. Maiorov, "On best approximation by ridge functions", *J. Approx. Theory* **99** (1999) 68–94; doi:10.1006/jath.1998.3304.

[9] H. N. Mhaskar, F. J. Narcowich and J. D. Ward, "Approximation properties of zonal function networks using scattered data on the sphere", *Adv. Comput. Math.* **11** (1999) 121–137; doi:10.1023/A:1018967708053.

[10] I. H. Sloan, "Polynomial approximation on spheres – generalizing de la Vallée-Poussin", *Comput. Methods Appl. Math.* **11** (2011) 540–552; doi:10.2478/cmam-2011-0029.

[11] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces* (Princeton University Press, Princeton, NJ, 1971).