Neural Networks 63 (2015) 57–65

Contents lists available at ScienceDirect

Neural Networks

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

Jackson-type inequalities for spherical neural networks with doubling weights[☆]

Shaobo Lin[a], Jinshan Zeng[b,*], Lin Xu[b], Zongben Xu[b]

[a] Institute of Intelligent System and Decision, Wenzhou University, Wenzhou 325035, China

[b] Institute for Information and System Sciences, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China

ARTICLE INFO

Article history:
Received 16 April 2014
Received in revised form 5 August 2014
Accepted 10 November 2014
Available online 20 November 2014

Keywords:
Spherical neural networks
Doubling weight
Minimal Riesz energy points
Jackson-type inequality

ABSTRACT

Recently, the spherical data processing has emerged in many applications and attracted a lot of attention. Among all the methods for dealing with the spherical data, the spherical neural networks (SNNs) method has been recognized as a very efficient tool due to SNNs possess both good approximation capability and spacial localization property. For better localized approximant, weighted approximation should be considered since different areas of the sphere may play different roles in the approximation process. In this paper, using the minimal Riesz energy points and the spherical cap average operator, we first construct a class of well-localized SNNs with a bounded sigmoidal activation function, and then study their approximation capabilities. More specifically, we establish a Jackson-type error estimate for the weighted SNNs approximation in the metric of L^p space for the well developed doubling weights.

1. Introduction

Spherical data abound in geophysics, astrophysics, computer graphics and other areas. These data are used to synthesize a function with a certain purpose (regression, classification, clustering, etc.). Generally, to implement a synthesizing process, a set of parameterized functions called the approximation model should be specified in advance. Such a model is commonly selected to encode the prior knowledge of the data, and has significant influence on the performance of the synthesizing process. To measure the performance of the selected model, the approximation capability is one of the most important factors.

The spherical approximation model can be generally represented as a span of some basis functions,

$$\mathcal{H}_n := \left\{ \sum_{i=1}^{n} a_i g_i(x) : a_i \in \mathbf{R} \right\},$$

where $g_i (i = 1, \ldots, n)$ are functions defined on the unit sphere $\mathbf{S}^d \subset \mathbf{R}^{d+1}$. The approximation capability and other attributions of $\mathcal{H}_n$ are obviously determined by the properties and structures of g_i. For example, if $g_i(x) = \phi(x_i \cdot x)$, where ϕ is a positive definite radial function (Wendland, 2005) and $\{x_i\}_{i=1}^n$ is a set of spherical data, then $\mathcal{H}_n$ is the spherical basis function model (Narcowich & Ward, 2002). Up till now, we have witnessed enormous emergence of spherical approximation models such as spherical polynomials (SPs) (Brown & Dai, 2005; Dai, 2006a, 2006b; Ditzian, 2004; Wang & Li, 2000; Xu, 2005), spherical basis functions (SBFs) (Freeden, Gervens, & Schreiner, 1998; Jetter, Stökler, & Ward, 1999; Mhaskar, 2006; Mhaskar, Narcowich, & Ward, 1999; Narcowich, Sun, Ward, & Wendland, 2007; Narcowich & Ward, 2002) and spherical neural networks (SNNs) (Lin, Cao, Chang, & Xu, 2012; Lin, Cao, & Xu, 2011; Lin, Zeng, & Xu, 2014; Mhaskar, Narcowich, & Ward, 2000, 2003). All of them have been proved to possess universal approximation property (Lin et al., 2011; Sun & Cheney, 1997; Wang & Li, 2000). However, on one hand, the localization theory (Freeden & Michel, 1999) shows that SPs are rarely spacial localized and therefore incapable of representing local features of some phenomena, which are particularly important in the studies of geodesy and geophysics (Freeden et al., 1998). On the other hand, the data-dependent property of SBFs leads to poor approximation capability provided the spherical data are "badly" located (Levesley & Sun, 2005).

☆ The research was supported by the National 973 Programming (2013CB329404), the Key Program of National Natural Science Foundation of China (Grant No. 11131006) and the National Natural Science Foundation of China (Grant No. 11401462).

* Corresponding author.
E-mail address: jsh.zeng@gmail.com (J. Zeng).

http://dx.doi.org/10.1016/j.neunet.2014.11.002
0893-6080/

Under this circumstance, SNNs come into our sights. SNNs can be mathematically expressed as

$$N_n(x) := \sum_{i=1}^{n} c_i\sigma(h_i(x)), \quad x \in \mathbf{S}^d, \tag{1.1}$$

where $h_i : \mathbf{S}^d \to \mathbf{R}$ is a hidden processing function, $c_i \in \mathbf{R}$ is an output weight, and $\sigma : \mathbf{R} \to \mathbf{R}$ is an activation function. We denote by $\Phi_{\sigma,n}$ the family of SNNs formed as (1.1), i.e.,

$$\Phi_{\sigma,n} := \left\{ N_n(x) = \sum_{i=1}^{n} c_i\sigma(h_i(x)) : c_i \in \mathbf{R} \right\}.$$

It is easy to see that SBFs are special types of SNNs in which $h_i(x) = x_i \cdot x$ and σ is (conditional) positive definite. Another widely used SNNs are to take $h_i(x) = w_i \cdot x + b_i$, where $w_i \in \mathbf{R}^{d+1}$ and $b_i \in \mathbf{R}$. For such SNNs, Lin et al. (2011) found that there exists an activation function such that the approximation capability of SNNs can be essentially better than that of SBFs. More recently, Lin et al. (2012) constructed a well-localized SNN via introducing a hidden processing function. Such a hidden processing function is generated according to a general distance. Furthermore, they provided a pointwise Jackson-type error estimate for the constructed SNNs.

Although, the SNNs constructed in Lin et al. (2012) (see also Lin et al., 2014) have been proved to possess both good approximation capability and localized property, there are still certain flaws that should be remedied. At first, the SNNs proposed in Lin et al. (2012) were constructed based on the spherical data, which makes the approximation capabilities of the corresponding SNNs depend heavily on the geometric distribution of the data. Secondly, the approximation result was only available to continuous functions. At last, as a well localized approximant, weighted approximation capability should be analyzed since different areas of the sphere may play different roles in the approximation process.

The main purpose of this paper is to construct a similar SNNs approximation model to overcome the above flaws, and then develop the corresponding theoretical results. The main novelty can be concluded as the following three aspects. Firstly, the SNNs constructed in this paper are based on the minimal s-Riesz energy points (Kuijlaars & Saff, 1998) with $s \geq d-1$. Thus, the approximation capability of the SNNs depends on the number of neurons rather than the geometric distribution of the spherical data. Secondly, using the well developed spherical cap average operator (Ditzian & Runovskii, 2000), we prove that the constructed SNNs can approximate arbitrary pth ($1 \leq p < \infty$) Lebesgue integrable functions defined on the sphere. Thirdly, we study the weighted approximation capability of the well spacial localized SNNs. Specifically, we establish several Jackson-type inequalities for the constructed SNNs with respect to the so-called doubling weight (Dai, 2006a).

The rest of this paper is organized as follows. In the next section, we provide some preliminaries including doubling weight, minimal s-Riesz energy points, and spherical cap average operator. In Section 3, we construct the SNNs and deduce weighted Jackson-type inequalities. In the last section, we present the proofs.

2. Preliminaries

2.1. Doubling weight

A weight function W on $\mathbf{S}^d$ is called a doubling weight (Dai, 2006a, 2006b) if there exists a constant $L > 0$ (called the doubling constant) such that, for any $x \in \mathbf{S}^d$ and $t > 0$,

$$W(D(x,2t)) := \frac{1}{\Omega_d}\int_{D(x,2t)} W(y)d\omega(y) \leq L\frac{1}{\Omega_d}\int_{D(x,t)} W(y)d\omega(y) = LW(D(x,t)), \tag{2.1}$$

where $D(z,\theta)$ denotes the spherical cap with center $z \in \mathbf{S}^d$ and radius $\theta > 0$:

$$D(z,\theta) := \{y \in \mathbf{S}^d : \arccos y \cdot z \leq \theta\},$$

and $d\omega$ denotes the surface area element on $\mathbf{S}^d$. Furthermore, we denote by Ω_d and $D(\theta)$ the volumes of $\mathbf{S}^d$ and $D(z,\theta)$, respectively. A simple computation yields that

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

and

$$D(\theta) := \Omega_{d-1}\int_0^{\theta} \sin^{d-1} t dt \sim \theta^d.$$

Many of the weights such as the generalized Jacobi weights satisfy the doubling condition (2.1). We refer the readers to Erdélyi (1999, 2003) and Mastroianni and Totik (1999, 2000, 2001) for more information on the doubling weights.

To measure the approximation capability, we need introduce the definition of weighted moduli of smoothness. Denote by $L^p(\mathbf{S}^d)$ ($1 \leq p \leq \infty$) the space of pth Lebesgue integrable functions on $\mathbf{S}^d$ endowed with the norms

$$\|f\| := \|f(\cdot)\|_{L^\infty(\mathbf{S}^d)} := \operatorname*{ess\,sup}_{x\in\mathbf{S}^d} |f(x)|, \quad p = \infty,$$

and

$$\|f\|_p := \|f(\cdot)\|_{L^p(\mathbf{S}^d)} := \left\{\int_{\mathbf{S}^d} |f(x)|^p d\omega(x)\right\}^{1/p} < \infty, \quad 1 \leq p < \infty.$$

Let $SO(d+1)$ be the group of rotations on $\mathbf{R}^{d+1}$. For $\rho \in SO(d+1)$ and integer $r > 0$, we define

$$T_\rho f(x) = f(\rho x),\ \Delta_\rho^r = (I - T_\rho)^r f,$$

where I denotes the identity operator. For an integer $r > 0$ and a weight function w, the weighted rth-order moduli of smoothness on $\mathbf{S}^d$ (Dai, 2006b) is defined by

$$\omega^r(f,t)_{p,w} := \sup_{\rho\in O_t} \|w(\Delta_\rho^r f)\|_p, \quad t > 0, 1 \leq p \leq \infty,$$

where $O_t := \{\rho \in SO(d+1) : \max_{x\in\mathbf{S}^d} \arccos(x)\cdot\rho x \leq t\}$. If $r = 1$, then we write $\omega(f,t)_{p,w}$ instead of $\omega^1(f,t)_{p,w}$. Furthermore, if $w = 1$, then we use $\omega(f,t)_p$ for the sake of brevity.

If we set $W_0 = W_1(x)$ and

$$W_m(x) = m^d\int_{D(x,1/m)} W(y)d\omega(y), \quad m = 1, 2, \ldots, x \in \mathbb{S}^d, \tag{2.2}$$

then the following two lemmas can be found in Dai (2006b, Lemma 2.1) and Dai (2006b, Lemma 2.2), respectively.

Lemma 2.1. *There exists a number $l > 0$ depending only on the doubling constant of W such that, for arbitrary $x, y \in \mathbf{S}^d$,*

$$W_m(x) \leq C(1 + m\arccos(x\cdot y))^l W_m(y), \quad m = 0, 1, \ldots,$$

where $C > 0$ depends only on the doubling weight constant of W.

Lemma 2.2. *For $1 \leq p \leq \infty$, $a > 1$ and $t > 0$, we have*

$$\omega(f,at)_{p,W_m} \leq C(p,W)(1 + mat)^l a\omega(f,t)_{p,W_m}, \quad m = 0, 1, 2, \ldots,$$

with $l > 0$ the same as that in Lemma 2.1.

2.2. Minimal Riesz energy points on the sphere

The Riesz s-energy ($s \geq 0$) associated with Ξ_n, $E_s(\Xi_n)$, is defined by (see Hardin & Saff, 2004 for example)

$$E_s(\Xi_n) := \begin{cases} \sum_{i\neq j} |x_i - x_j|^{-s}, & \text{if } s > 0 \\ \sum_{i\neq j} -\log |x_i - x_j|, & \text{if } s = 0. \end{cases}$$

Here $|\cdot|$ denotes the Euclidean norm in $\mathbf{R}^{d+1}$. We use $\mathcal{E}_s(\mathbf{S}^d, n)$ to denote the n-point minimal s-energy over $\mathbf{S}^d$, which is defined as

$$\mathcal{E}_s(\mathbf{S}^d, n) := \inf_{\Xi_n \in \mathbf{S}^d} E_s(\Xi_n), \tag{2.3}$$

where the infimum is taken over all n-points subsets of $\mathbf{S}^d$. If $\Xi_n \subset \mathbf{S}^d$ satisfies

$$E_s(\Xi_n) = \mathcal{E}_s(\mathbf{S}^d, n),$$

then Ξ_n is called a minimal s-energy configuration, and the points in Ξ_n are called minimal s-energy points. The determination of a minimal s-energy configuration and its corresponding minimal s-energy on $\mathbf{S}^d$ is very important since this problem will emerge in many fields of scientific researches such as physics, chemistry and computer science. For further background, we refer the readers to Dahlberg (1978), Hardin and Saff (2004, 2005), Kuijlaars and Saff (1998), Kuijlaars, Saff, and Sun (2007) and references therein. According to a series of work in Dahlberg (1978), Kuijlaars and Saff (1998), Kuijlaars et al. (2007) and Lin et al. (2012), it demonstrates that all minimal s-energy configurations with $s \geq d-1$ satisfy the following two lemmas.

Lemma 2.3. *Let $\Xi_n := \{\xi_i\}_{i=1}^n$ be the minimal s-energy configuration with $s \geq d-1$, then there exists a constant c_d depending only on d such that*

$$\mathbf{S}^d \subset \bigcup_{i=1}^n D\left(\xi_i, c_d n^{-1/d}\right).$$

Lemma 2.4. *Let $\Xi_n := \{\xi_i\}_{i=1}^n$ be the minimal s-energy configuration with $s \geq d-1$, then we have*

$$\left|\left\{\xi \in \Xi_n : \xi \in D(\xi_i, c_d n^{-1/d})\right\}\right| \leq C_d, \tag{2.4}$$

where $|A|$ denotes the cardinal norm of the set A, and C_d is a constant depending only on d.

2.3. Spherical cap average operator

For a function $f \in L^1(\mathbf{S}^d)$ and $h \in \mathbf{R}$, the spherical cap average operator is defined by

$$B_h(f, x) := f_h(x) := \frac{1}{D(h)} \int_{D(x,h)} f(y) d\omega(y). \tag{2.5}$$

The properties of $B_h(f)$ were developed well in Ditzian and Runovskii (2000). Particularly, a strong inverse inequality of $B_h(f)$ was established in Ditzian and Runovskii (2000). In this subsection, we deduce some properties of $B_h(f)$ concerning doubling weight.

Lemma 2.5. *Let $1 \leq p < \infty$ and f_h be defined as in (2.5). Then, for arbitrary $f \in L^p(\mathbf{S}^d)$, it holds*

$$\|f - f_h\|_{p,W_m} \leq \omega(f, h)_{p,W_m}.$$

Lemma 2.6. *Let $1 \leq p < \infty$ and f_h be defined as in (2.5). Suppose further $\Xi_n := \{x_1, \ldots, x_n\}$ is the minimal s-energy configuration with $s \geq d-1$. Then, for arbitrary $f \in L^p(\mathbf{S}^d)$, it holds*

$$\sum_{i=1}^n \int_{D(x_i,h)} |f_h(x_i) - f_h(x)|^p W_m(x) d\omega(x) \leq C(1+2mh)^l \left(\omega(f, h)_{p,W_m}\right)^p,$$

where l is defined in Lemma 2.1 and C is a constant depending only on p and d.

Lemma 2.7. *Let $1 \leq p < \infty$ and f_h be defined as in ((2.5)). Suppose further $\Xi_n := \{x_1, \ldots, x_n\}$ is the minimal s-energy configuration with $s \geq d-1$. Then, for arbitrary $f \in L^p(\mathbf{S}^d)$, it holds*

$$D(h) \sum_{i=2}^n |f_h(x_i) - f_h(x_{i-1})|^p W_m(x_i) \leq C(1+mh)^l \left(\omega(f, h)_{p,W_m}\right)^p,$$

where l is defined in Lemma 2.1 and C is a constant depending only on p and d.

Lemma 2.8. *Let $1 \leq p < \infty$ and f_h be defined as in ((2.5)). Suppose further $\Xi_n := \{x_1, \ldots, x_n\}$ is the minimal s-energy configuration with $s \geq d-1$. Then, for arbitrary $f \in L^p(\mathbf{S}^d)$, it holds*

$$D(h) \sum_{j=3}^n \left(\sum_{i=2}^{j-1} |f_h(x_i) - f_h(x_{i-1})|\right)^p W_m(x_j) \leq Cn^{l+p} \left(\omega(f, h)_{p,W_m}\right)^p,$$

and

$$D(h) \sum_{j=2}^{n-1} \left(\sum_{i=j+1}^{N} |f_h(x_i) - f_h(x_{i-1})|\right)^p W_m(x_j) \leq Cn^{l+p} \left(\omega(f, h)_{p,W_m}\right)^p,$$

where l is defined in Lemma 2.1 and C is a constant depending only on p and d.

3. Weighted Jackson-type inequality for SNNs

In this section, we aim to construct a type of well-localized SNNs and study their approximation capabilities.

3.1. Construction of the SNNs

The construction of the SNNs in this paper is based on the minimal Riesz energy points, spherical cap average operator and the general distance proposed in Lin et al. (2012). Let $\Xi_n := \{x_i\}_{i=1}^n$ be the n-point minimal s-energy configuration. The mesh norm (Mhaskar et al., 1999) of Ξ_n is defined by

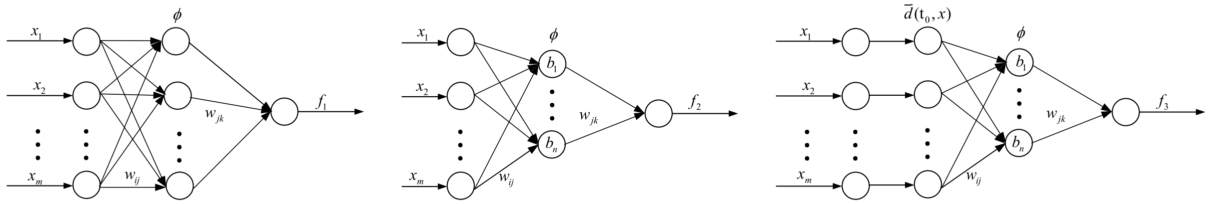
$$h_\Xi := \max_{x\in\mathbf{S}^d} \min_j d(x, x_j),$$

where $d(x, y)$ is the geodesic (great circle) distance between the points x and y on $\mathbf{S}^d$. We rearrange the minimal Riesz energy points such that $d(x_k, x_{k+1}) \leq h_\Xi$, $k = 1, 2, \ldots, n-1$. Since $\mathbf{S}^d \subset \bigcup_{i=1}^n D(x_i, h_\Xi)$, for arbitrary $x \in \mathbf{S}^d$, there exists at least a point $y \in \Xi_n$ such that $x \in D(y, h_\Xi)$. If we set

$$k := \min\{j : x \in D(x_j, h_\Xi)\}, \tag{3.1}$$

then for arbitrary $x \in \mathbf{S}^d$, there exists a unique k satisfying (3.1) such that $x \in D(x_k, h_\Xi)$. For $x \in D(x_{k_0}, h_\Xi)$ and $y \in D(x_{j_0}, h_\Xi)$, we define the general distance by

$$\bar{d}(x, y) := \begin{cases} d(x, y), & k_0 = j_0, \\ \sum_{i=j_0}^{k_0} d(x_i, x_{i+1}) + d(x_{k_0}, x) + d(x_{j_0}, y), & j_0 < k_0, \\ \sum_{i=k_0}^{j_0} d(x_i, x_{i+1}) + d(x_{k_0}, x) + d(x_{j_0}, y), & k_0 < j_0. \end{cases} \tag{3.2}$$



(a) Structure of SBFs. (b) Structure of classical SNNs. (c) Structure of classical SNNs.

Fig. 1. Comparisons with topological structures of SBFs, classical SNNs and SNNs (3.3).

Setting $h = h_{\Xi}$, the SNNs are constructed as

$$N_{n,\sigma}^{A}(x) := f_h(x_1) + \sum_{i=1}^{n-1} \left(f_h(x_{i+1}) - f_h(x_i)\right) \sigma^* \times \left(\overline{d}(x_1, x) - \overline{d}(x_1, x_i)\right), \tag{3.3}$$

where $\sigma^*(t) := \sigma(At)$ and A is a constant depending only on σ and n.

The proposed network $N_{n,\sigma}^{A}$ can be interpreted as a model of feed-forward neural networks with four layers:

- The first one is the input layer with finitely many inputs.
- The second one is the pre-handling layer which transforms an input x into the general distance between x and x_1.
- The third one is the handling layer with n neurons.
- The last one is the output layer.

The topological structures of SBFs, the SNNs proposed in Lin et al. (2011) and the SNNs (3.3) are illustrated in Fig. 1.

It can be seen from Fig. 1(a) that SBFs are special SNNs without thresholds whose neurons are set to be the spherical data. From Fig. 1(b), we can find that the classical SNNs (Lin et al., 2011) are three-layer neural networks whose inner weights and thresholds need adjusting based on the data. Differently, it follows from Fig. 1(c) that the new SNNs (3.3) possess a pre-handling layer. It should be noted that with such an easy-tackled pre-handling process (Lin et al., 2012), we can endow a certain spacial localization property to the SNNs (see (3.6)).

In order to guarantee both good approximation capability and spacial localization property of the SNNs, the activation functions utilized in this paper are assumed to be bounded, sigmoidal and local pth integrable:

(a1) Bounded: $\sup_{t\in\mathbf{R}} |\sigma(t)| \le \|\sigma\| < \infty$.

(a2) Sigmoidal: $\lim_{t\to\infty} \sigma(t) = 1, \quad \lim_{t\to-\infty} \sigma(t) = 0$.

(a3) Local pth integrable: σ is pth integrable on every compact subset of $\mathbf{R}$.

The above three assumptions on the activation function are very mild in the realm of neural networks. Actually, the widely used Heaviside function ϕ and logistic function ψ satisfy our assumptions (a1)–(a3), where

$$\phi(t) := \begin{cases} 1, & t \ge 0, \\ 0, & t < 0, \end{cases}$$

and

$$\psi(t) = \frac{1}{1+e^{-t}}.$$

The definition of the sigmoidal function σ implies that there exists an $A \in \mathbf{Z}_+$ such that

$$|\sigma(t)| \le n^{-(1+l/p)} \quad \text{if } t \le -Ah \tag{3.4}$$

and

$$|\sigma(t) - 1| \le n^{-(1+l/p)} \quad \text{if } t > Ah, \tag{3.5}$$

where l is a constant defined as in Lemma 2.1.

Now, we are in a position to illustrate the spacial localization property of the SNNs (3.3). We can rewrite $N_{n,\sigma}^{A}$ as follows. Define

$$c_1(x) := 1 - \sigma^*(\overline{d}(x_1, x)),$$
$$c_i(x) := \sigma^*\left(\overline{d}(x_1, x) - \overline{d}(x_1, x_{i-1})\right) - \sigma^*\left(\overline{d}(x_1, x) - \overline{d}(x_1, x_i)\right), \quad 2 \le i \le n-1,$$
$$c_n(x) := \sigma^*\left(\overline{d}(x_1, x) - \overline{d}(x_1, x_{n-1})\right).$$

Then

$$N_{n,\sigma}^{A} := \sum_{i=1}^{n} f_h(x_i) c_i(x). \tag{3.6}$$

If σ is the Heaviside function, then for arbitrary $A > 0$, there is only one $c_i(x) \ne 0$. This means that $N_{n,\sigma}^{A}(x) = f_h(x_i)$ for $x \in D(x_i, h)$, and 0 otherwise, which further shows that $N_{n,\sigma}$ is the piece-wise constant function with respect to a certain spherical partition. If σ is taken as the other sigmoidal function (say, $\sigma^*(t) = \frac{1}{1+e^{-At}}$), then we can set A to be sufficiently large such that $c_i(x)$ degenerates fast when $x \notin D(x_i, h)$. As an extreme case, we can set $A = \infty$, and in this case, σ^* is actually the Heaviside function. In summary, the well spacial localization of $N_{n,\sigma}^{A}$ can be realized by selecting either a suitable A or an appropriate activation function σ.

3.2. *Jackson-type error estimate of SNNs approximation*

In this part, we give Theorem 3.1, which describes the approximation capability of the SNNs, $N_{n,\sigma}^{A}$, constructed in Section 3.1.

Theorem 3.1. *Let $1 \le p < \infty$, $m, n \in \mathbf{N}$ satisfying $n \sim m^d$, W_m and $N_{n,\sigma}^{A}$ be defined as in (2.2) and (3.3), and $\Xi_n = \{x_i\}_{i=1}^{n}$ be the minimal s-energy points with $s \ge d-1$. If σ satisfies (a1)–(a3) and A satisfies (3.4) and (3.5), then for arbitrary $f \in L^p(\mathbf{S}^d)$, it holds*

$$\|f - N_{n,\sigma}^{A}\|_{p,W_m} \le C\omega\left(f, n^{-1/d}\right)_{p,W_m},$$

where C is a constant depending only on s, l, d and p.

Regarding the SPs approximation, Dai (2006b) also established the Jackson-type inequalities with doubling weights. Furthermore, it can be found in Dai (2006b, Theorem 1.1) that their inequalities hold for arbitrary rth-order moduli of smoothness. At the first glance, it seems negative since Theorem 3.1 holds only for $r = 1$. However, it should be emphasized that the Jackson-inequality in Theorem 3.1 cannot be improved to $r > 1$, even in the unweighted case. Indeed, using the same method as that in Chen (1993, Theorem 3), we can deduce that

$$C_1\omega\left(f, n^{-1}\right)_p \le \|f - N_{n,\sigma}^{A}\|_{p,1} \le C_2\omega\left(f, n^{-1}\right)_p$$

for $d = 1$. In short, since Theorem 3.1 holds for SNNs with arbitrary activation function satisfying assumptions (a1)–(a3), the error estimate in Theorem 3.1 cannot be essentially improved. It means that more restrictions to σ should be required to deduce the Jackson-type inequality with rth-order moduli of smoothness with $r > 1$. Moreover, it can be noted that the condition $n \sim m^d$

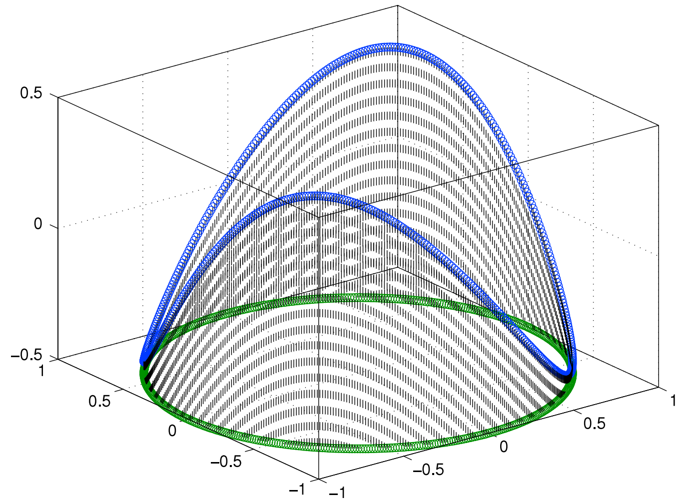


Fig. 2. The underlying function $f(x^1, x^2) := x^1 \sin(x^2)$.

is not stringent since the dimension of SPs of degrees at most m is $\mathcal{O}(m^d)$.

At the end of the section, we give a simple simulation to illustrate the feasibility of the proposed SNNs (3.3). For the underlying function $f(x^1, x^2) := x^1 \sin(x^2)$ (see Fig. 2) defined on $\mathbf{S}^1$, we sampled $n = 200$ equally spaced points which is an extreme case of the minimal Riesz s energy points with $s \to \infty$ (Saff & Kuijlaars, 1997). Furthermore, to compute the integral $f_h(x_i)$, we sampled another 100 points in $D(x_i, h)$ and used $\frac{1}{100}\sum_{i=1}^m f(x_i)$ to approximate $f_h(x_i)$. Under this circumstance, we can construct the SNNs with Heaviside activation function according to (3.3) to approximate the underlying function (see Fig. 3). As shown in Figs. 2 and 3, it can be observed that the constructed SNNs approximate the underlying function very well. Moreover, Fig. 4 shows the error between the underlying function and the constructed SNNs.

4. Proofs

4.1. Proof of Lemma 2.5

From the definition of f_h, it follows that

$$\|f_h - f\|_{p,W_m}^p = \int_{\mathbb{S}^d} |f_h(x) - f(x)|^p W_m(x)d\omega(x)$$
$$= \int_{\mathbb{S}^d} \left|\frac{1}{D(h)}\int_{D(x,h)} f(y)d\omega(y) - f(x)\right|^p W_m(x)d\omega(x)$$
$$= \int_{\mathbb{S}^d} \left|\frac{1}{D(h)}\int_{D(x,h)} (f(y) - f(x))\, d\omega(y)\right|^p W_m(x)d\omega(x).$$

Let $\rho_{x,y} \in SO(d+1)$ such that $y = \rho_{x,y}x$. Let $e_0 = (0, \ldots, 0, 1)$ be the north pole of $\mathbf{S}^d$. By the Hölder inequality, for $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\left|\frac{1}{D(h)}\int_{D(x,h)} (f(y) - f(x))\, d\omega(y)\right|^p$$
$$= \left|\frac{1}{D(h)}\int_{D(x,h)} f(\rho_{e_0,x}y)d\omega(\rho_{e_0,x}y) - f(x)\right|^p$$
$$= \left|\frac{1}{D(h)}\int_{D(e_0,h)} \left(f(\rho_{e_0,x}y) - f(x)\right) d\omega(y)\right|^p$$
$$\leq \left(\frac{1}{D(h)}\right)^p \int_{D(e_0,h)} \left|\left(f(\rho_{e_0,x}y) - f(x)\right)\right|^p d\omega(y) D(h)^{\frac{p}{q}}$$
$$\leq \frac{1}{D(h)}\int_{D(e_0,h)} \left|\left(f(\rho_{e_0,x}y) - f(x)\right)\right|^p d\omega(y).$$

Therefore,

$$\|f_h - f\|_{p,W_m}^p \leq \frac{1}{D(h)}\int_{D(e_0,h)}\int_{\mathbb{S}^d} \left|\left(f(\rho_{e_0,x}y) - f(x)\right)\right|^p \times W_m(x)d\omega(x)d\omega(y).$$

Because $y \in D(e_0, h)$ and $\rho_{x,e_0}x = e_0$, we have $\rho_{e_0,x}y \in D(x, h)$ and further

$$\|f_h - f\|_{p,W_m}^p \leq \left(\omega(f, h)_{p,W_m}\right)^p,$$

which completes the proof of Lemma 2.5. ■

4.2. Proof of Lemma 2.6

By the definition of f_h, we obtain

$$\sum_{i=1}^n \int_{D(x_i,h)} |f_h(x_i) - f_h(x)|^p W_m(x)d\omega(x)$$
$$= \sum_{i=1}^n \int_{D(x_i,h)} \left|\frac{1}{D(h)}\left(\int_{D(x_i,h)} f(y)d\omega(y) - \int_{D(x,h)} f(z)d\omega(z)\right)\right|^p \times W_m(x)d\omega(x).$$

Then,

$$\int_{D(x,h)} f(z)d\omega(z) = \int_{D(x,h)} f(\rho_{x_i,x}y)d\omega(\rho_{x_i,x}y) = \int_{D(x_i,h)} f\left(\rho_{x_i,x}y\right) d\omega(y).$$

Hence, by the Hölder inequality, we have

$$\sum_{i=1}^n \int_{D(x_i,h)} \left|\frac{1}{D(h)}\left(\int_{D(x_i,h)} f(y)d\omega(y) - \int_{D(x,h)} f(z)d\omega(z)\right)\right|^p \times W_m(x)d\omega(x)$$
$$= \sum_{i=1}^n \int_{D(x_i,h)} \left|\frac{1}{D(h)}\left(\int_{D(x_i,h)} \left(f(y) - f\left(\rho_{x_i,x}y\right)\right) d\omega(y)\right)\right|^p \times W_m(x)d\omega(x)$$
$$\leq \frac{1}{D(h)}\sum_{i=1}^n \int_{D(x_i,h)}\int_{D(x_i,h)} \left|f(y) - f\left(\rho_{x_i,x}y\right)\right|^p \times W_m(x)d\omega(y)d\omega(x).$$

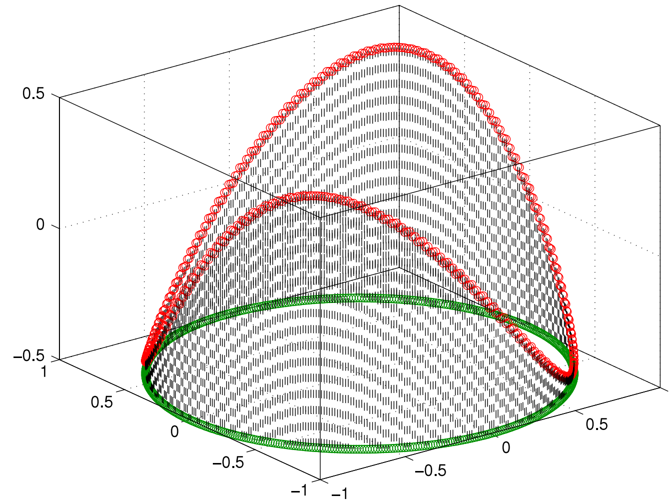


Fig. 3. The constructed SNNs based on 200 neurons.

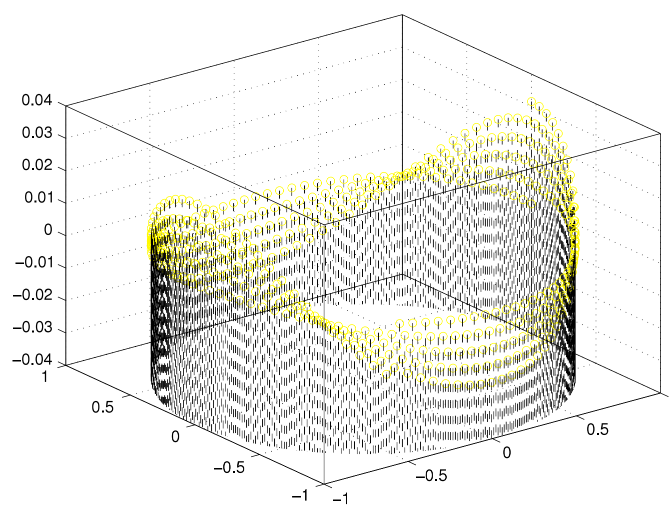


Fig. 4. Error function.

Then, it follows from Lemma 2.1 that the above quantity can be bounded by

$$C\frac{1}{D(h)}\sum_{i=1}^{n}\int_{D(x_i,h)}\int_{D(x_i,h)}\left|f(y)-f\left(\rho_{x,x_i}y\right)\right|^p \times (1+m\arccos(x\cdot y))^l W_m(y)d\omega(y)d\omega(x).$$

Furthermore, since $x, y \in D(x_i, h)$, we have

$$(1+m\arccos(x\cdot y))^l \leq (1+2mh)^l.$$

Thus,

$$\begin{aligned}&\sum_{i=1}^{n}\int_{D(x_i,h)}\left|\frac{1}{D(h)}\left(\int_{D(x_i,h)}f(y)d\omega(y)-\int_{D(x,h)}f(z)d\omega(z)\right)\right|^p \times W_m(x)d\omega(x)\\ &\leq C(1+2mh)^l\sum_{i=1}^{n}\int_{D(x_i,h)}\left|f(y)-f\left(\rho_{x_i,x}y\right)\right|^p W_m(y)d\omega(y)\\ &\leq C(1+2mh)^l\int_{\mathbf{S}^d}\left|f(y)-f\left(\rho_{x_i,x}y\right)\right|^p W_m(y)d\omega(y)\\ &\leq C(1+2mh)^l\left(\omega(f,h)_{p,W_m}\right)^p,\end{aligned}$$

where the last two inequalities can be easily deduced by Lemma 2.4, i.e.,

$$\int_{\mathbf{S}^{d-1}}|g(x)|d\omega(x)\leq\sum_{i=1}^{n}\int_{D(x_i,h)}|g(x)|d\omega(x)\leq 3\int_{\mathbf{S}^{d-1}}|g(x)|d\omega(x)$$

and $\rho_{x,x_i} \in O_h$. This finishes the proof of Lemma 2.6. ∎

4.3. Proof of Lemma 2.7

From the definition of f_h, it follows that

$$\begin{aligned}&D(h)\sum_{i=2}^{n}|f_h(x_i)-f_h(x_{i-1})|^p\\ &=D(h)\sum_{i=2}^{n}\left|\frac{1}{D(h)}\left(\int_{D(x_i,h)}f(y)d\omega(y)-\int_{D(x_{i-1},h)}f(z)d\omega(z)\right)\right|^p.\end{aligned}$$

Since $x_i, x_{i-1} \in D(x_i, 2h)$, we have $\rho_{x_{i-1},x_i} \in O_{2h}$. Then

$$\int_{D(x_{i-1},h)}f(z)d\omega(z)=\int_{D(x_i,h)}f\left(\rho_{x_i,x_{i-1}}y\right)d\omega(y).$$

Therefore, by the Hölder inequality, Lemmas 2.1 and 2.2, we have

$$D(h)\sum_{i=2}^{n}\left|\frac{1}{D(h)}\left(\int_{D(x_i,h)}f(y)d\omega(y)-\int_{D(x_{i-1},h)}f(z)d\omega(z)\right)\right|^p W_m(x_i)$$

$$=D(h)\sum_{i=2}^{n}\left|\frac{1}{D(h)}\int_{D(x_i,h)}\left(f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right)d\omega(y)\right|^p W_m(x_i)$$

$$\leq C(1+mh)^l D(h)\frac{1}{D(h)}\sum_{i=2}^{n}\int_{D(x_i,h)}\left|\left(f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right)\right|^p \times W_m(y)d\omega(y)$$

$$\leq C(1+mh)^l\sum_{i=2}^{n}\int_{D(x_i,h)}\left|\left(f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right)\right|^p W_m(y)d\omega(y)$$

$$\leq C(1+mh)^l\int_{\mathbf{S}^d}\left|\left(f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right)\right|^p W_m(y)d\omega(y)$$

$$\leq C(1+mh)^l\left(\omega(f,2h)_{p,W_m}\right)^p\leq C(1+mh)^l\left(\omega(f,h)_{p,W_m}\right)^p.$$

The proof of Lemma 2.7 is completed. ■

4.4. Proof of Lemma 2.8

We only prove the first inequality. The proof of the second one is similar, we omit the details. By the definition of $f_h(x)$, we can easily calculate that

$$\sum_{i=2}^{j-1}|f_h(x_i)-f_h(x_{i-1})|W_m^{\frac{1}{p}}(x_j)$$

$$=\sum_{i=2}^{j-1}\left|\frac{1}{D(h)}\left(\int_{D(x_i,h)}f(y)d\omega(y)-\int_{D(x_{i-1},h)}f(z)d\omega(z)\right)\right| \times W_m^{\frac{1}{p}}(x_j)$$

$$\leq C(1+jmh)^{\frac{l}{p}}\sum_{i=2}^{j-1}\frac{1}{D(h)}\int_{D(x_i,h)}\left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right| \times W_m^{\frac{1}{p}}(y)d\omega(y)$$

$$\leq C(1+jmh)^{\frac{l}{p}}\frac{1}{D(h)}\int_{D(x_2,h)\cup\ldots\cup D(x_{j-1},h)}\left|f(y)-f\left(\rho_{x_{i-1},x_i}y\right)\right| \times W_m^{\frac{1}{p}}(y)d\omega(y).$$

Thus, by the Hölder inequality, Lemmas 2.2 and 2.4, we have

$$D(h)\sum_{j=3}^{n}\left(\sum_{i=2}^{j-1}|f_h(x_i)-f_h(x_{i-1})|\right)^p W_m(x_j)$$

$$\leq D(h)\sum_{j=3}^{n}C(1+jmh)^l\left(\frac{1}{D(h)}\int_{D(x_2,h)\cup\ldots\cup D(x_{j-1},h)} \times \left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right|W_m^{\frac{1}{p}}(y)d\omega(y)\right)^p$$

$$\leq C(1+nmh)^l\sum_{j=3}^{n}\frac{1}{D(h)^{p-1}}\int_{D(x_2,h)\cup\ldots\cup D(x_{j-1},h)} \times \left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right|^p W_m(y)d\omega(y)\left((j-2)D(h)\right)^{\frac{p}{q}}$$

$$\leq C(1+nmh)^l\sum_{j=3}^{n}(j-2)^{p-1}\int_{D(x_2,h)\cup\ldots\cup D(x_{j-1},h)} \times \left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right|^p W_m(y)d\omega(y)$$

$$\leq C(1+nmh)^l n^{p-1}\sum_{j=3}^{n}\int_{D(x_2,h)\cup\ldots\cup D(x_{j-1},h)} \times \left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right|^p W_m(y)d\omega(y)$$

$$\leq C(1+nmh)^l n^{p-1}\sum_{j=1}^{n}\int_{\mathbf{S}^d}\left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right|^p W_m(y)d\omega(y)$$

$$\leq C(1+nmh)^l n^p\int_{\mathbf{S}^d}\left|f(y)-f\left(\rho_{x_i,x_{i-1}}y\right)\right|^p \times W_m(y)d\omega(y)\leq C(1+nmh)^l n^p\left(\omega(f,2h)_{p,W_m}\right)^p$$

$$\leq C(1+nmh)^l n^p\left(\omega(f,h)_{p,W_m}\right)^p.$$

This completes the proof of Lemma 2.8. ■

4.5. Proof of Theorem 3.1

From the definition of σ^*, we obtain

$$|\sigma^*(t)-1|\leq n^{-(1+l/p)}\quad \text{if } t\geq h\quad\text{and}$$
$$|\sigma^*(t)|\leq n^{-(1+l/p)}\quad \text{if } t\leq -h. \tag{4.1}$$

For any $1\leq j\leq N$, and any $x\in D(x_j,h)$, it follows from Lemma 2.4 that there exists a constant $u\leq C_d$ such that

$$x_{j-u},x_{j-u+1},\ldots,x_j,\ldots,x_{j+u}\in D(x,h).$$

Then it holds

$$\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\geq h,$$
$$\text{for } k=1,2,\ldots,j-u-1,\ j=2,3,\ldots,N$$

and also

$$\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\leq -h,$$
$$\text{for } k=j+u+1,j+u+2,\ldots,N,\ j=1,2,\ldots,N-1.$$

Furthermore, from (4.1), it implies

$$\left|\sum_{k=1}^{j-u-1}\left(f_h(x_{k+1})-f_h(x_k)\right)\left(\sigma^*\left(\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\right)-1\right)\right|$$

$$\leq\frac{1}{n^{1+l/p}}\sum_{k=1}^{j-u-1}|f_h(x_{k+1})-f_h(x_k)|,$$

and

$$\left|\sum_{k=j+u+1}^{n-1}\left(f_h(x_{k+1})-f_h(x_k)\right)\sigma^*\left(\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\right)\right|$$

$$\leq\frac{1}{n^{1+l/p}}\sum_{k=j+u+1}^{n-1}|f_h(x_{k+1})-f_h(x_k)|.$$

Therefore, we have

$$N_{n,\sigma}^A(x)-f_h(x)$$

$$=f_h(x_1)-f_h(x)+\sum_{k=1}^{n-1}\left(f_h(x_{k+1})-f_h(x_k)\right)\sigma^* \times\left(\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\right)$$

$$=f_h(x_1)-f_h(x)+\sum_{k=1}^{j-u-1}\left(f_h(x_{k+1})-f_h(x_k)\right) \times\left(\sigma^*\left(\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\right)-1\right)$$

$$+f_h(x_{j-u})-f_h(x_1)+\sum_{k=j-u}^{j+u}\left(f_h(x_{k+1})-f_h(x_k)\right)\sigma^* \times\left(\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\right)$$

$$+\sum_{k=j+u+1}^{n-1}\left(f_h(x_{k+1})-f_h(x_k)\right)\sigma^*\left(\bar{d}(x_1,x)-\bar{d}(x_1,x_k)\right).$$

Thus, it follows from Lemma 2.3 that

$$\|N_{n,\sigma} - f_h\|_{p,W_m}^p = \int_{\mathbb{S}^d} |N_{n,\sigma}(x) - f_h(x)|^p W_m(x) d\omega(x) \le \sum_{j=1}^n \int_{D(x_j,h)} |N_{n,\sigma}(x) - f_h(x)|^p W_m(x) d\omega(x).$$

If we set $\sum_{k=i_0}^{i_1} |a_k| = 0$ for $i_0 < i_1$, then we have

$$\|N_{n,\sigma} - f_h\|_{p,W_m}^p \le \sum_{j=1}^n \int_{D(x_j,h)} \Bigg| f_h(x_{j-u}) - f_h(x) + \sum_{k=1}^{j-u-1} (f_h(x_{k+1}) - f_h(x_k)) \times \left(\sigma^*\left(\bar{d}(x_1,x) - \bar{d}(x_1,x_k)\right) - 1\right) + \sum_{k=j-u}^{j+u} (f_h(x_{k+1}) - f_h(x_k))\, \sigma^*\left(\bar{d}(x_1,x) - \bar{d}(x_1,x_k)\right) + \sum_{k=j+u+1}^{n-1} (f_h(x_{k+1}) - f_h(y_k))\, \sigma^*\left(\bar{d}(x_1,x) - \bar{d}(x_1,x_k)\right) \Bigg|^p \times W_m(x) d\omega(x).$$

Furthermore, since

$$|a+b+c+d|^p \le 4^p(|a|^p + |b|^p + |c|^p + |d|^p), \quad a,b,c,d \in \mathbb{R},\ 1 \le p < \infty,$$

it holds

$$\begin{aligned}\|N_{n,\sigma} - f_h\|_{p,W_m}^p &\le 4^p \sum_{j=1}^n \int_{D(x_j,h)} \Bigg(\left|f_h(x_{j-u}) - f_h(x)\right|^p + \left| \frac{1}{n} \sum_{k=1}^{j-u-1} |f_h(x_{k+1}) - f_h(x_k)| \right|^p + \|\sigma\|^p \left| \sum_{k=j-u}^{j+u} (f_h(x_{k+1}) - f_h(x_k)) \right|^p + \left| \frac{1}{n} \sum_{k=j+u+1}^{n-1} |f_h(x_{k+1}) - f_h(x_k)| \right|^p \Bigg) W_m(x) d\omega(x) \\ &\le 4^p \sum_{j=1}^n \int_{D(x_j,h)} \left|f_h(x_{j-u}) - f_h(x)\right|^p W_m(x) dx + (4\|\sigma\|)^p \sum_{j=1}^n \int_{D(x_j,h)} \sum_{k=j-u}^{j+u} |f_h(x_{k+1}) - f_h(x_k)|^p W_m(x) d\omega(x) \\ &\quad + (4/n^{1+l/p})^p \sum_{j=1}^n \int_{D(x_j,h)} \left| \sum_{k=1}^{j-u} |f_h(x_{k+1}) - f_h(x_k)| \right|^p W_m(x) d\omega(x) + (4/n^{1+l/p})^p \sum_{j=1}^n \int_{D(x_j,h)} \left| \sum_{k=j+u+1}^{n-1} |f_h(x_{k+1}) - f_h(x_k)| \right|^p \times W_m(x) d\omega(x) \\ &= I_1 + I_2 + I_3 + I_4.\end{aligned}$$

Now, we use Lemmas 2.6–2.8 to bound I_1, I_2, I_3 and I_4, respectively. Since $x \in D(x_j, h)$, it follows from Lemmas 2.2 and 2.6 that

$$I_1 \le C\omega(f,h)_{p,W_m}^p \le C\omega\left(f, n^{-1/d}\right)_{p,W_m}^p. \tag{4.2}$$

To bound I_2, we use Lemma 2.1 to obtain

$$\begin{aligned} I_2 &\le \frac{1}{D(h)} (4\|\sigma\|)^p \sum_{j=1}^n \sum_{k=j-u}^{j+u} \int_{D(x_j,h)} |f_h(x_{k+1}) - f_h(x_k)|^p \times D(h) W_m(x) d\omega(x) \\ &\le C(1+mh)^l D(h) \sum_{i=2}^n |f_h(x_i) - f_h(x_{i-1})|^p W_m(x_i). \end{aligned}$$

Thus, it follows from Lemma 2.7 and $h \sim 1/m \sim n^{1/d}$ that

$$I_2 \le C(1+mh)^{2l} \omega(f,h)_{p,W_m}^p \le C\omega\left(f, n^{-1/d}\right)_{p,W_m}^p. \tag{4.3}$$

To bound I_3, note that

$$\begin{aligned} I_3 &\le Cn^{-(p+l)} \sum_{j=1}^n \int_{D(x_j,h)} \left| \sum_{i=1}^{j-u} |f_h(x_{k+1}) - f_h(x_k)| \right|^p W_m(x) d\omega(x) \\ &\le Cn^{-(p+l)} D(h) \sum_{j=1}^n \left| \sum_{i=1}^{j-u} |f_h(x_{k+1}) - f_h(x_k)| \right|^p W_m(x_j) d\omega(x). \end{aligned}$$

Then it follows from Lemma 2.8 that

$$I_3 \le C\omega(f,h)_{p,W_m}^p C\omega\left(f, n^{-1/d}\right)_{p,W_m}^p. \tag{4.4}$$

Using a similar method as above, we can obtain

$$I_4 \le C(1+2mh)^l \omega(f,h)_{p,W_m}^p C\omega\left(f, n^{-1/d}\right)_{p,W_m}^p. \tag{4.5}$$

Thus, the triangle inequality

$$\|f - N_{n,\sigma}^A\|_{p,W_m} \le \|f - f_h\|_{p,W_m} + \|f_h - N_{n,\sigma}^A\|_{p,W_m}$$

together with Lemma 2.5, (4.2)–(4.5) yield Theorem 3.1 directly. ■

References

Brown, G., & Dai, F. (2005). Approximation of smooth functions on compact two-point homogeneous spaces. *Journal of Functional Analysis*, *220*, 401–423.

Chen, D. B. (1993). Degree of approximation by superpsitions of a sigmoidal function. *Approximation Theory and its Applications*, *9*, 17–28.

Dahlberg, B. (1978). On the distribution of fekete points. *Duke Mathematical Journal*, *45*, 537–542.

Dai, F. (2006a). Multivariate polynomial inequalities with respect to doubling weights and A_∞ weights. *Journal of Functional Analysis*, *235*, 137–170.

Dai, F. (2006b). Jackson-type inequality for doubling weights on the sphere. *Constructive Approximation*, *24*, 91–112.

Ditzian, Z. (2004). Jackson-type inequality on the sphere. *Acta Mathematica Hungarica*, *102*(1–2), 1–35.

Ditzian, Z., & Runovskii, K. (2000). Averages on caps of $\mathbf{S}^{d-1}$. *Journal of Mathematical Analysis and Applications*, *248*, 260–274.

Erdélyi, T. (1999). Notes on inequalities with doubling weights. *Journal of Approximation Theory*, *100*, 60–72.

Erdélyi, T. (2003). Markov-Bernstein-Type inequality for trigonometric polynomials with respect to doublingweights on $[-\omega, \omega]$. *Constructive Approximation*, *19*, 329–338.

Freeden, W., Gervens, T., & Schreiner, M. (1998). *Constructive approximation on the sphere*. Oxford: Calderon Press.

Freeden, W., & Michel, V. (1999). Constructive approximation and numerical methods in geodetic research today — an attempt at a categorization based on an uncertainty principle. *Journal of Geodesy*, *73*, 452–465.

Hardin, D., & Saff, E. (2004). Discretizing manifolds via minimum energy points. *Notices of the American Mathematical Society*, *51*, 1186–1194.

Hardin, D., & Saff, E. (2005). Minimal Riesz energy point configurations for rectifiable *d*-dimensional manifolds. *Advances in Mathematics*, *193*, 174–204.

Jetter, K., Stökler, J., & Ward, J. (1999). Error estimates for scattered data interpolation on spheres. *Mathematics of Computation*, *68*, 733–749.

Kuijlaars, A., & Saff, E. (1998). Asymptotics for minimal discrete energy on the sphere. *Transactions of the American Mathematical Society*, *350*, 523–538.

Kuijlaars, A., Saff, E., & Sun, X. (2007). On separation of minimal Riesz energy points on spheres in Euclidean spaces. *Journal of Computational and Applied Mathematics*, *199*, 172–180.

Levesley, J., & Sun, X. (2005). Approximation in rough native spaces by shifts of smooth kernels on spheres. *J. Approx. Theory*, *133*, 269–283.

Lin, S. B., Cao, F. L., Chang, X. Y., & Xu, Z. B. (2012). A general radial quasi-interpolation operator on the sphere. *Journal of Approximation Theory*, *164*, 1402–1414.

Lin, S. B., Cao, F. L., & Xu, Z. B. (2011). Essential rate for approximation by spherical neural networks. *Neural Networks*, *24*, 752–758.
Lin, S. B., Zeng, J. S., & Xu, Z. B. (2014). Error estimate for spherical neural networks interpolation. *Neural Processing Letters*, http://dx.doi.org/10.1007/s11063-014-9361-x.
Mastroianni, G., & Totik, V. (1999). Jackson type inequalities for doubling weights II. *East Journal on Approximations*, *5*, 101–166.
Mastroianni, G., & Totik, V. (2000). Weighted polynomial iequalities with doubling and A_∞ weights. *Constructive Approximation*, *16*, 37–71.
Mastroianni, G., & Totik, V. (2001). Best approximation and moduli of smoothness for doubling weights. *Journal of Approximation Theory*, *110*, 180–199.
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., 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. (2000). In B. Widrow, L. Guan, K. Paliwa, T. Adall, J. Larsen, E. Wilson, & S. Douglas (Eds.), *Neural network frames on the sphere, in Neural networks for signal processing, X* (pp. 175–184). New York: IEEE.
Mhaskar, H. N., Narcowich, F. J., & Ward, J. D. (2003). Zonal function network frames on the sphere. *Neural Networks*, *16*, 183–203.
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.
Narcowich, F. J., & Ward, J. D. (2002). Scattered data interpolation on spheres: Error estimates and locally supported basis functions. *SIAM Journal on Mathematical Analysis*, *33*, 1393–1410.
Saff, E., & Kuijlaars, A. (1997). Distributing many points on a sphere. *The Mathematical Intelligencer*, *19*, 5–11.
Sun, X. P., & Cheney, E. W. (1997). Fundamental sets of continuous functions on spheres. *Constructive Approximation*, *13*, 245–250.
Wang, K. Y., & Li, L. Q. (2000). *Harmonic analysis and approximation on the unit sphere*. Beijing: Science Press.
Wendland, H. (2005). *Scattered data approximation*. Cambridge: Cambridge University Press.
Xu, Y. (2005). Weighted approximation of functions on the unit sphere. *Constructive Approximation*, *21*, 1–28.