Neurocomputing 175 (2016) 348–354

Contents lists available at ScienceDirect

Neurocomputing

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

Simultaneous approximation by spherical neural networks ☆

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

[a] College of Mathematics and Information Science, Wenzhou University, Wenzhou, 325035, PR China
[b] Institute of Metrology and Computational Science, China Jiliang University, Hangzhou 310018, PR China

ARTICLE INFO

Article history:
Received 22 May 2015
Received in revised form
4 August 2015
Accepted 19 October 2015
Communicated by M. Bianchini
Available online 30 October 2015

Keywords:
Spherical neural networks
Simultaneous approximation
Laplace–Beltrami operator
Spherical polynomials

ABSTRACT

Approximation capabilities of the spherical neural networks (SNNs) are considered in this paper. Based on a known Taylor formula, we prove that, for non-polynomial target function, rates of simultaneously approximating the function itself and its (Laplace–Beltrami) derivatives by SNNs is not slower than those by the spherical polynomials (SPs). Then, the simultaneous approximation rates of SPs automatically derive the rates of SNNs.

1. Introduction

Spherical data abound in geodesy, meteorology, astrophysics, geophysics, and other areas [8,9]. For example, the mathematical models of some satellite missions such as the GOCE (Gravity Field and Steady-State Ocean Circulation Explorer) and the CHAMP (Challenging Mini-Satellite Payload for Geophysical Research and Application) which study the gravity potential of the earth are spherical Fredholm integral equations of the first kind [11]. Hence, finding a tool which can efficiently tackle spherical data becomes more and more important.

A basic and classical tool for fitting scattered data on the sphere is the spherical polynomial (SP). Up till now, the approximation capability of SPs has been widely studied [36]. For example, Ditzian [7] deduced a Jackson-type error estimate and its Stechkin-type inverse for SPs; Sloan [34] constructed a hyperinterpolation operator, which is an SP, and deduced the approximation error bound; Dai [5] provided a weighted Jackson inequality for SPs. The main algorithm to implement the SP approximation is the singular value decomposition approach [19]: Expand a function with respect to the orthonormal basis and estimate the corresponding Fourier coefficients. But it is well-known that the spherical harmonic basis is badly localized and incapable of representing local features of the target function, which is important in geophysics applications [10]. Therefore, one turns to find a tool which possesses nice localization performance. Consequently, spherical basis function (SBF) and spherical neural networks (SNNs) come into our sights.

SBF refers to a positive definite function on the $(d+1)$-dimensional unit sphere $\mathbf{S}^d$. Here a positive definite function on $\mathbf{S}^d$ means the matrix $A_\phi:=(\phi(\langle x_i,x_j\rangle))_{i,j=1}^M$ is positive definite [17, Def. 2.7]. The SBF method focuses on using the linear combination of shifts of SBF on the spherical data. Mathematically, the approximant is formed as

$$x\mapsto\sum_{i=1}^{M}c_i\phi(\langle x_i,x\rangle),\quad x\in\mathbf{S}^d,\quad c_i\in\mathbf{R},\tag{1.1}$$

where ϕ is an SBF, x_i is the spherical data and $\langle x,y\rangle$ denotes the inner product in $\mathbf{R}^{d+1}$. There are two topics on the SBF approximation. The first one is the density problem which concerns whether the approximant (1.1) can approximate arbitrary continuous function within arbitrary accuracy, provided the number of spherical data is sufficiently large and the coefficients $\{c_i\}_{i=1}^M$ are appropriately tuned. In the seminal paper [35], Sun and Cheney derived the sufficient and necessary conditions for the density of SBF approximation. The other one called the complexity problem is to determine how many samples are necessary to yield a prescribed degree of approximation by using the SBF approximant (1.1). For this problem, Mhaskar et al. [27] gave an upper bound of approximation by using the positive cubature and Marcinkiewicz–Zygmund inequality. They utilized the summation of the best approximation error of SPs and a residual depending on the smoothness of the SBF to bound the approximation error of the SBF approximant (1.1). Some studies for the SBF approximation can also be found in [4,14,16,21,30,32,33] and references therein.

☆The research was supported by the National Natural Science Foundation of China (Grant no. 61502342, 11401462, 91330118, 61272023).
* Corresponding author.
E-mail address: sblin1983@gmail.com (S. Lin).

http://dx.doi.org/10.1016/j.neucom.2015.10.067
0925-2312/

A typical way to derive the SBF approximant is to use the least squares, which has already been proved to posses perfect approximation capability [12].

Besides the SBF method, [23] also proposed the SNN which formed as

$$\sum_{i=1}^{M} a_i\sigma(\langle w_i, x\rangle + b_i), \tag{1.2}$$

where, $w_i \in \mathbf{R}^{d+1}$, $a_i \in \mathbf{R}$, $b_i \in \mathbf{R}$ are the inner weight, outer weight, threshold of the SNN, respectively, and σ is named as the activation function in the terminology of neural networks [13]. It can be easily found that if σ is positive definite, w_i is the data point and b_i=0, then the SNN defined above coincides with the SBF approximant (1.1). In our previous paper [23], we theoretically proved that there exists an SNN which possesses essentially better approximation capability than the SP and SBF methods in the sense that SNN can deduce a similar approximation error by using much smaller M. However, the results in [23] do not present any hints in selecting the activation function and the related parameters. The aim of the present paper is to pursue the approximation capability of a large range of SNNs and provide a guidance about how to select the activation function and the corresponding parameters.

On the other hand, simultaneous approximation of a function and its derivatives are required in many science and engineering applications. There have been many studies on the problem of simultaneous approximation by neural networks on $\mathbf{R}^{d+1}$[1,22,37,38]. We also pursue the simultaneous approximation capability of SNN in this paper. By using a representation theorem and a Taylor formula, we firstly construct an SNN and use it and its (Laplace–Beltrami) derivatives to approximate SPs and their (Laplace–Beltrami) derivatives. After introducing the best approximation operator on the sphere and using the commutativity of the best approximation operator and the Laplace–Beltrami operator, we then derive the upper bound of simultaneous approximation error of SPs. Under this circumstance, a quantitative upper bound estimate on simultaneous approximation by SNNs can be derived. The obtained results reveal that the simultaneous approximation rate of the constructed SNN depends not only on the number of hidden units used, but also on the smoothness of functions to be approximated. Furthermore, it can be deduced that, for non-polynomial target functions the rate of simultaneous approximation by SNNs is not slower than that by SPs.

The paper is organized as follows. In the next section, some preliminaries together with a representation theorem of SPs will be given. In Section 3, the upper bound of simultaneous approximation by SNNs will be established. In the last sections, we will give the proofs of the main results.

2. Spherical harmonics

Denote by $L^p(\mathbf{S}^d)(1 \le p \le \infty)$ the space of p-th 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

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

We denote by $d\omega$ the surface area element on $\mathbf{S}^d$. The volume of $\mathbf{S}^d$ is denoted by Ω_d, and it is easy to deduce that

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

For integer $k \ge 0$, the restriction to $\mathbf{S}^d$ of a homogeneous harmonic polynomial of degree k on the unit sphere is called a spherical harmonic of degree k. The span of all spherical harmonics of degree k is denoted by $\mathbf{H}_k^d$, and the class of all SPs of degree $k \le n$ is denoted by $\Pi_n{}^d$. It is obvious that $\Pi_n^d = \oplus_{k=0}^n \mathbf{H}_k^d$. The dimension of $\mathbf{H}_k^d$ is given by

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

and that of $\Pi_n{}^d$ is $\sum_{k=0}^n D_k^d = D_n^{d+1} \sim n^d$, where $A \sim B$ denotes that there exist absolute constants C_1 and C_2 such that $C_1 A \le B \le C_2 A$.

Spherical harmonics have an intrinsic characterization. To describe this, we first introduce the Laplace–Beltrami operator Δ, which is defined by [31]:

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

It is well-known that the operator Δ is an elliptic, (unbounded) selfadjoint operator on $L^2(\mathbf{S}^d)$, is invariant under arbitrary coordinate changes, and its spectrum comprises distinct eigenvalues $-\lambda_k := -k(k+d-1)$, $k=0,1,\ldots,$ each having finite multiplicity. The space $\mathbf{H}_k^d$ can be characterized intrinsically as the eigenspace corresponding to λ_k, i.e.

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

Since λ_ks are distinct, and the operator is selfadjoint, the spaces $\mathbf{H}_k^d$ are mutually orthonormal; also, $L^2(\mathbf{S}^d) = \text{closure}\{\oplus_k \mathbf{H}_k^d\}$. 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 well-known addition formula [36] is given by

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

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 relation

$$\int_{-1}^{1} P_k^{d+1}(t)P_j^{d+1}(t)(1-t^2)^{(d-2)/2}\,dt = \frac{\Omega_d}{\Omega_{d-1}D_k^d}\delta_{k,j},$$

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

The following Funk–Hecke formula [36] plays an important role in computing the eigenvalues of the kernel $\phi \in L^1([-1,1])$.

$$\int_{\mathbf{S}^d} \phi(\langle x,y\rangle)H_k(y)\,d\omega(y) = B(\phi,k)H_k(x), \tag{2.3}$$

where

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

In order to reveal the simultaneous approximation capability of SNNs, we need the following representation theorem, which was proven in [23].

Lemma 1. *Let $n \in \mathbf{N}$. Then for any $P_n \in \Pi_n^d$, there exists a set of points $\{a_k\}_{k=1}^{D_n^d} \subset \mathbf{S}^d$ and a set of univariate polynomials $\{g_k\}_{k=1}^{D_n^d}$ defined on $[-1,1]$ with degrees not larger than n such that*

$$P_n(x) = \sum_{k=1}^{D_n^d} g_k(\langle a_k, x\rangle), \quad x \in \mathbf{S}^d. \tag{2.4}$$

From the classical representation theorem (Theorem 3 of [31]), we know that every SP can be represented by a combination of