Preprints.org

Article

Not peer-reviewed version

Convergence Rate of Regularized Regression Associated with Zonal Translation Networks

Xuexue Ran , Baohuai Sheng * , Shuhua Wang

Posted Date: 5 August 2024

doi: 10.20944/preprints202408.0275.v1

Keywords: regularized regression learning; convolution translation network; reproducing property; Marcinkiewicz-Zygmund inequality; quadrature rule; convergence rate

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Convergence Rate of Regularized Regression Associated with Zonal Translation Networks

Xuexue Ran [1], Baohuai Sheng [2,*] and Shuhua Wang [3]

[1] School of Mathematical Physics and Information, Shaoxing University, Shaoxing 312000, China; xuexue6892@163.com
[2] Department of Economic Statistics, School of International Business, Zhejiang Yuexiu University, Shaoxing 312000, China; 20221022@zyufl.edu.cn,bhsheng@usx.edu.cn
[3] School of Information Engineering, Jingdezhen Ceramic University, Jingdezhen 333403, China; w614sh@126.com
* Correspondence: 20221022@zyufl.edu.cn, bhsheng@usx.edu.cn

Abstract: Neural network regularized learning has garnered significant attention in recent years. We give a systematic investigation on the performance of regularized regression associated zonal translation networks. We propose the concept of Marcinkiewicz-Zygmund inequality Setting (MZIS) for the scattered nodes collected from the unit sphere. We show that, under the MZIS, the corresponding convolutional zonal translation network has reproducing property. Based on these facts, we propose a kind of kernel regularized regression learning framework and provide upper bound estimate for the learning rate with the kernel approach. We also give proof for the density of the zonal translation network with spherical Fourier analysis. We provide the approximation error with a K-functional.

Keywords: regularized regression learning; convolution translation network; reproducing property; Marcinkiewicz-Zygmund inequality; quadrature rule; convergence rate

MSC: 41A25

---

1. Introduction

It is known that convolutional neural network provides various models and algorithms to process data model in many fields such as computer vision(see e.g. [1]), natural lagrange processing (see e.g. [2]), and sequence analysis in bio-informatics(see e.g. [3]). The regularized neural network learning has thus become an attractive research topic (see e.g.[4–9]). In the present paper, we shall give theory analysis for the convergence rate of regularized regression associated with zonal translation network on the unit sphere.

Let X be a compact subset in the d-dimensional Euclidean space R^d with the usual norm $\|x\| = \sqrt{\sum_{k=1}^{d} x_k^2}$ for $x = (x_1, x_2, \cdots, x_d) \in R^d$ and Y be a nonempty closed subset contained in $[-M, M]$ for a given $M > 0$. The aim of the regression learning problem is to learn the target function which describes the relationship between the input $x \in X$ and the output $y \in Y$ from a hypothesis function space. In most of the cases, the target function is offered as a set of observations $z = \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m \in Z^m$ which has been drawn independently and identically distributed (i.i.d.) according to a probability joint distribution (measure) $\rho(x,y) = \rho_X(x)\ \rho(y|x)$ on $Z = X \times Y$, where $\rho(y|x)(x \in X)$ is the conditional probability of y for a given x and $\rho_X(x)$ is the marginal probability about x, i.e.,for every integrable function $\varphi :\ X \times Y \to R$ there holds

$$\int_{X\times Y} \varphi(x,y)d\rho = \int_X \Big(\int_Y \varphi(x,y)d\rho(y \mid x) \Big) d\rho_X.$$

For a given normed space $(B, \|\cdot\|_B)$ consisting of real functions on X we define the regularized learning framework with B as

$$f_{z,\lambda} := arg \min_{f\in B} \Big(\mathcal{E}_z(f) + \frac{\lambda}{2}\|f\|_B^2 \Big), \tag{1}$$

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

where $\lambda > 0$ is the regularization parameter,$\mathcal{E}_z(f)$ is the empirical mean

$$\mathcal{E}_z(f) = \frac{1}{m}\sum_{i=1}^{m}(y_i - f(x_i))^2.$$

The optimal target function is the regression function

$$f_\rho(x) = \int_Y y\, d\rho(y|x)$$

satisfying

$$f_\rho = \inf_f \mathcal{E}_\rho(f),$$

where $\mathcal{E}_\rho(f) = \int_Z (y - f(x))^2\, d\rho$ and the inf is taken over all the ρ_X-measurable functions f. Moreover, there holds the famous equality (see e.g.[10])

$$\|f - f_\rho\|^2_{L^2(\rho_X)} = \mathcal{E}_\rho(f) - \mathcal{E}_\rho(f_\rho), \qquad f \in L^2(\rho_X). \tag{2}$$

The choices for the hypothesis space B in (1) are riches. For example, C.P.An et al choose the algebraic polynomial class as B (see [11–13],). In [14], C. De Mol et al chose the dictionary as B. Recently some papers chose the Sobolev space as the hypothesis space B (see [15,16]). By kernel method we mean traditionally replacing B with a reproducing kernel Hilbert space (RKHS) $(H_K, \langle\cdot,\cdot\rangle_K)$ which is a Hilbert space consisting of real functions defined on X and there is a Mercer kernel $K_x(y) = K(x, y)$ on $X \times X$ (i.e., $K_x(y)$ is a continuous and symmetric function on $X \times X$ and for any $n \geq 1$ and any $\{x_1, x_2, \cdots, x_n\} \subset X$ the Mercer matrices $(K(x_i, x_j))_{i,j=1,2,\cdots,n}$ are positive semi-definite) such that

$$f(x) = \langle f, K_x\rangle_K, \qquad \forall x \in X, \forall f \in H_K. \tag{3}$$

and there hold the embeded inequality

$$|f(x)| \leq c\|f\|_K, \qquad \forall x \in X, \forall f \in H_K, \tag{4}$$

where c is a constant independent of f and x. There are two results about the optimal solution $f_{z,\lambda}$. The reproducing property (3) yields the representation

$$f_{z,\lambda}(x) = \sum_{k=1}^{m} c_k K_{x_k}(x), \qquad \forall x \in X. \tag{5}$$

The embeded inequality (4) yields the inequality

$$|f_{z,\lambda}(x)| \leq \sqrt{\frac{2M}{\lambda}} \qquad \forall x \in X. \tag{6}$$

Representation (5) is the theory basis for kernel regularized regression (see e.g.[17,18]). Inequality (6) is the key inequality for bounding the learning rate with covering number method (see e.g.[19–21]). For other skills of the kernel method one can cite [10,22–24] et al.

It is particularly important to mention here that translation networks have recently been used for the hypothesis space of regularized learning (see e.g. [25,26]). From the view of approximation theory, a simple single layer translation network with m neurons is a function space produced by translating a given function ϕ which can be written as

$$\Delta^X_{\phi,\overline{X}} = \{\sum_{i=1}^{m} c_i\, T_{x_i}(\phi, \cdot) : \qquad c_i \in R, x_i \in X, i = 1, 2, \cdots, m\},$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

where $\overline{X} = \{x_i\}_{i=1}^m \subset X$ is a given node set, and for a given $x \in X$ $T_x(\phi,\ y)$ is a translation operator corresponding to X. For example, when $X = R^d$ and $X = [-\pi, \pi]^d$, we choose $T_x(\phi,\ y)$ as the usual convolution translation operator $\phi(x-y)$ for the ϕ defined on R^d or an 2π-periodic function ϕ (see [27,28]. When $X = S^{d-1} = \{x \in R^d : \|x\| = 1\}$ is the unit sphere in R^d, one can choose $T_x(\phi,\ y)$ as the zonal translation operator $\phi(xy)$ for a given ϕ defined on the interval $[-1,1]$ (see [29]). In [30], we defined a kind of translation operator $T_x(\phi,\ y)$ for $X = [-1,1]$.

It is easy to see that the approximation ability and the construction of a translation network depend upon the nodes set $\overline{X} = \{x_i\}_{i=1}^m$ (see e.g. [31–33]). On the other hand, according to the view of [34], the quadrature rule and the Marcinkiewicz-Zygmund (M-Z) inequality associated with $\overline{X}$ also influence the construction of the translation network $\Delta_{\phi,\overline{X}}$. For a bounded closed set $\Omega \subset R^d$ with measure $d\omega$ satisfying $\int_\Omega d\omega = V < +\infty$. We denote by $P_n \subset L^2(\Omega)$ the linear space of polynomials on Ω of degree at most n, equipped with the L^2-inner product $\langle v,\ z\rangle = \int_\Omega vzd\omega$. The m-point quadrature rule (QR) is

$$\int_\Omega g\ d\omega \approx \sum_{j=1}^m w_j\ g(x_j), \tag{7}$$

where $\overline{X} = \{x_j\}_{j=1}^m \subset \Omega$ and weights w_j are all positive for $j = 1, 2, \cdots, m$. We say the QR (7) has polynomial exactness n if

$$\int_\Omega g\ d\omega = \sum_{j=1}^m w_j\ g(x_j), \qquad \forall g \in P_n. \tag{8}$$

The Marcinkiewicz-Zygmund (MZ) inequality based on a set $\overline{X} = \{x_j\}_{j=1}^m \subset \Omega$ is

$$\left(\sum_{j=1}^m w_j\ |g(x_j)|^2\right)^{\frac{1}{2}} \sim \left(\int_\Omega |g(\omega)|^2\ d\omega\right)^{\frac{1}{2}}, \qquad \forall g \in P_n, \tag{9}$$

where the weights w_j in (9) may be not the same as the w_j in (7) and (8).

Accord to the view of [34], the quadrature rule (QR) follows automatically from the Marcinkiewicz-Zygmund (MZ) inequality. H.N.Mhaskar et al gave a method of transition from MZ inequality to polynomial exact GR in [35]. So the MZ inequality (9) is an important features for describing the nodes set $\overline{X}$. Since this reason the node set $\overline{X} = \{x_j\}_{j=1}^m \subset \Omega$ which yields an MZ inequality is given a specially terminology called Marcinkiewicz-Zygmynd Family (MZF) (see [34,36–38]). However, from these literatures we know that the MZFs do not totally coincide with the Lagrange interpolation nodes in the case of $d > 1$. The hyperinterpolations are then developed with the help of exact QR (see [39–43]) and are applied to approximation theory and regularized learning (see e.g.[11–13,44]). On the other hand, we find the problem of polynomial exact QR is investigated individually (see e,g, [45,46]). The concept of spherical t-design was first defined in [47] and is given investigations by many papers subsequently, one can see the classical references [48,49]. We say $T_t = \{x_i\}_{i=1}^{|T_t|} \subset S^{d-1}$ is a spherical t-design if

$$\frac{1}{|T_t|}\sum_{i=1}^{|T_t|} \pi(x_i) = \frac{1}{\omega_{d-1}}\int_{S^{d-1}} \pi(x)d\omega(x), \tag{10}$$

where ω_{d-1} is the volume of S^{d-1} and $\pi(x)$ is a spherical polynomial with degree t. Moreover, in many applications the polynomial exact QR and the MZFS has been used as assumptions. For example, C.P.An et al gave approximation order for the hyperinterpolation approximation under the assumptions that (8), (10) and the MZ inequality (9) hold (see [50,51]). Also, in [25], Lin et al gave

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

investigations on regularized regression associated with zonal translation network by assuming the nodes set $\overline{X} = \{x_j\}_{j=1}^m \subset S^{d-1}$ is a type of spherical t-design.

Polynomial exact QR is also a good tool in approximation theory. For example, we had ever used it to bound the norm for some Mercer matrices (see [52–55]). In particular, H.N.Mhaskar et al use polynomial exact QR to construct the first periodic translation operators (see [27]) and the zonal translation network operators (see [29]). Along this line, the translation operators defined on the unit ball, on the Euclidean space R^d and on the interval $[-1,1]$ are constructed (see [28,30,56]).

Above investigations encourage us to use (8), (10) and (9) as hypothetical conditions to describe the approximation ability of zonal translation networks $\Delta_{\phi,\overline{X}}^X$. To ensure the single layer translation network $\Delta_{\phi,\overline{X}}^X$ can approximation the constant function, $\Delta_{\phi,\overline{X}}^X$ is modified as

$$N_{\phi,\overline{X}}^X = \{\sum_{i=1}^m c_i\, T_{x_i}(\phi,\ \cdot) + c_0 : \qquad c_0, c_i \in R, x_i \in X, i = 1, 2, \cdots, m\}. \tag{11}$$

In the case of $T_y(\phi,\ y) = \sigma(w^\top x + b)$ and $w, x \in R^d, b \in R$, R.D.Nowak et al used (11) to design regularized learning frameworks (see [57]). An algorithm is provided by S.B.Lin et al [26] for designing such kind of network and is applied to construct regularized learning algorithms. In [5] $N_{\phi,\overline{X}}^X$ is used to construct deep neural network learning frameworks. The same type of investigations are given in [58], [59] and [60].

In the present paper, we shall design the translation networks $N_{\phi,\overline{X}}^X$ by taking $X = S^{d-1}$, assuming $\overline{X} = \Omega^{(m)} = \{x_j\}_{j=1}^m \subset S^{d-1}$ satisfies equalities (8) and (9), and choosing the zonal translation $T_x(\phi,\ y) = \phi(xy)$ with ϕ being a given integrable function ϕ on $[-1,1]$. Under these assumptions we shall provide a learning framework with $N_{\phi,\Omega^{(m)}}^{S^{d-1}}$ as the hypothesis space and give error analysis.

The contributions of the present paper are two folds. First, after absorbing the ideas of [34,36–38] and the successful experience of [11,25,27,29,50,51,61,62], we propose the concept of Marcinkiewicz-Zygmund Inequality Setting (MZIS) for the scattered nodes on the sphere unit, based on this as an assumption we show the reproducing property for the convolutional zonal translation network associated with the scattered nodes $\Omega^{(m)}$. Second, we give investigation on the kernel regularized neural network learning by combining classical the kernel approach and the convex analysis method, according to this method, the convergence rate given can be dimensional independent and capacity independent. Since the translation networks are produced by the zonal translations of convolutional kernels, we call them the convolutional zonal translation networks.

The paper is organized as follows. In Section 2 we first show the density for the zonal translations class and then show the reproducing property for the translation network $\Delta_{\phi,\Omega^{(m)}}^{S^{d-1}}$. In Section 3, we shall provide some results in the present paper, for example, a new regression learning framework and a learning setting, the error decomposition for the error analysis,and an estimate for the convergence rate. In Section 4, we shall give some lemmas which are used to prove the main results. The proofs for all the theorems and propositions are given in Section 5.

Throughout the paper, we write $A = O(B)$ if there is a positive constant C independent of A and B such that $A \leq CB$.In particular, by $A = O(1)$ we show A is a bounded quantity. We write $A \sim B$ if both $A = O(B)$ and $B = O(A)$.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

2. Some Properties of the Translation Network on the Unit Sphere

Let $\phi \in L^1_{W_\eta} = \{\phi : \|\phi\|_{1,W_\eta} = \int_{-1}^{1} |\phi(x)|W_\eta(x)dx < +\infty\}, W_\eta(x) = (1-x^2)^{\eta-\frac{1}{2}}, \eta > -\frac{1}{2}$. Then H.N.Mhaskar et al constructed in [29] a sequence of approximation operators to show that the zonal translation class

$$\begin{aligned}\Delta_\phi^{S^{d-1}} &= cl\{\phi(xy) : \quad y \in S^{d-1}\} \cup \{1\} \\ &= cl\{\sum_{i=1}^{m} c_i\, \phi(x_i \cdot) + c_0 : \qquad c_0, c_i \in R, x_i \in S^{d-1}, i = 1,2,\cdots,m; m = 1,2,\cdots\}\end{aligned}$$

is density in $L^p(S^{d-1})(1 \le p \le +\infty)$ if $\widehat{a_l^\eta(\phi)} \neq 0$ for all $l = 0,1,2,\cdots$, where

$$\widehat{a_l^\eta(\phi)} = c_\eta \int_{-1}^{1} \phi(x) \frac{C_l^\eta(x)}{C_l^\eta(1)} W_\eta(x)dx, \ \ \eta = \frac{d-2}{2}$$

and $C_n^\eta(x) = p_n^{(\eta-\frac{1}{2},\eta-\frac{1}{2})}(x)$ is the n-th Legendre polynomial satisfies the orthogonal relation

$$c_\eta \int_{-1}^{1} C_n^\eta(x) C_m^\eta(x) W_\eta(x)dx = h_n^\eta \delta_{n,m}.$$

with $c_\eta = \frac{\Gamma(\eta+1)}{\Gamma(\frac{1}{2})\Gamma(\eta+\frac{1}{2})}$, $h_n^\eta = \frac{\eta}{n+\eta} C_n^\eta(1)$ and it is known that (see it from (B.2.1),(B.2.2)and (B.5.1) of [63])$C_n^\eta(x) \le C_n^\eta(1) = n^{2\eta-1}, x \in [-1,1]$. It follows that

$$\phi(t) = \sum_{l=0}^{+\infty} \widehat{a_l^\eta(\phi)}\, \frac{n+\eta}{\eta} C_l^\eta(t) = \sum_{l=0}^{+\infty} \widehat{a_l^\eta(\phi)}\, Z_l^\eta(t), \tag{12}$$

where $Z_l^\eta(t) = \frac{l+\eta}{\eta} C_l^\eta(t), \eta = \frac{d-2}{2}$.

For a given real number $p \ge 1$,we denote by $L^p(\rho_X)$ the class of ρ_X-measurable functions f satisfying $\|f\|_{L^p(\rho_X)} = (\int_X |f(x)|^p d\rho_X) < +\infty$.

Let P_n^d denote the space of all homogeneous polynomials of degree n in d variables. We denote by $L^p(S^{d-1})$ the class of all measurable functions defined on S^{d-1} with the finite norm

$$\|f\|_{p,S^{d-1}} = \begin{cases} \left(\int_{S^{d-1}} |f(x)|^p d\sigma(x)\right)^{\frac{1}{p}}, & 1 \le p < +\infty, \\ \max_{x\in S^{d-1}} |f(x)|, & p = +\infty, \end{cases}$$

and for $p = +\infty$ we assume that $L^{+\infty}(S^{d-1})$ is the space $C(S^{d-1})$ of continuous functions on S^{d-1} with the uniform norm.

For a given integer $n \ge 0$, the restriction to S^{d-1} of a homogeneous harmonic polynomial of degree n is called a spherical harmonics of degree n. If $Y \in P_n^d$, then $Y(x) = \|x\|^n Y(x^{'}), x^{'} = \frac{x}{\|x\|}$, so that Y is determined by its restriction on the unit sphere. Let $H_n(S^{d-1})$ denote the space of spherical harmonics of degree n. Then

$$\dim H_n(S^{d-1}) = \binom{n+d-2}{n} + \binom{n+d-3}{n-1}, \qquad n = 1,2,3,\cdots,$$

Spherical harmonics of different degrees are orthogonal on the unit sphere. For further properties about harmonics one can refer to [64].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

For $n = 0, 1, 2, \cdots$, let $\{Y_l^n : 1 \le l \le \dim H_n(S^{d-1})\}$ be an orthonormal basis of $H_n(S^{d-1})$. Then

$$\frac{1}{\omega_{d-1}} \int_{S^{d-1}} Y_l^n(\xi)\, Y_{l'}^m(\xi)\, d\sigma(\xi) = \delta_{l,l'}\delta_{m,n},$$

where ω_{d-1} denotes the surface area of S^{d-1} and $\omega_{d-1} = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$. Furthermore, by (1.2.8) in [63] we have

$$\sum_{j=1}^{\dim H_n(S^{d-1})} Y_l^n(x)\, Y_l^n(y) = \frac{n+\eta}{\eta}\, C_n^\eta(xy), \qquad x, y \in S^{d-1}, \tag{13}$$

where $C_n^\eta(t)$ is the n-th generalized Legendre polynomial the same as in (12). Combining (12) and (13) we have

$$\phi(xy) = \sum_{l=0}^{+\infty} \widehat{a_l^\eta(\phi)} \frac{l+\eta}{\eta}\, C_l^\eta(x\cdot y) = \sum_{l=0}^{+\infty} \widehat{a_l^\eta(\phi)} \sum_{k=1}^{\dim H_l(S^{d-1})} Y_k^l(x)\, Y_k^l(y),\ x, y \in S^{d-1}. \tag{14}$$

2.1. Density

We first give a general discrimination method for density.

Proposition 2.1. (see Lemma 1 in Chapter 18 of [65]) For a subset V in a normed linear space E, the following two properties are equivalent:

(a) V is fundamental in E (that is, its linear span is dense in E).

(b) $V^\perp = \{0\}$ (that is, 0 is the only element of E^* that annihilates V).

Based on this proposition, we can show the density of $\Delta_\phi^{S^{d-1}}$.

Theorem 2.1. Let $\phi \in L^2_{W_\eta}$ satisfy $\widehat{a_l^\eta(\phi)} > 0$ for all $l = 0, 1, 2, \cdots$. Then $\Delta_\phi^{S^{d-1}}$ is dense in $L^2(S^{d-1})$.

Proof. See the proof in Section 5.

2.2. Reproducing Property

We first restate a proposition.

Proposition 2.2. For any given $n \ge 1$ there exist a finite subset $\Omega^{(n)} \subset S^{d-1}$ and corresponding positive numbers $\{\mu_\omega : \omega \in \Omega^{(n)}\}$ such that

$$\int_{S^{d-1}} f(x)\, d\sigma(x) = \sum_{\omega\in\Omega^{(n)}} \mu_\omega^{(n)} f(\omega), \quad f \in H_{3n}(S^{d-1}), \tag{15}$$

and for some $1 \le p < +\infty$

$$\int_{S^{d-1}} |f(x)|^p d\sigma(x) \sim \sum_{\omega\in\Omega^{(n)}} \mu_\omega^{(n)} |f(\omega)|^p, \quad f \in H_n(S^{d-1}). \tag{16}$$

Moreover, for any $m \ge n$ and $p \ge 1$ there exists a constant $c_{p,d} > 0$ such that

$$\sum_{\omega\in\Omega^{(n)}} \mu_\omega^{(n)} |f(\omega)|^p \le c_{p,d} \left(\frac{m}{n}\right)^{d-1} \int_{S^{d-1}} |f(x)|^p d\sigma(x), \quad f \in H_m(S^{d-1}). \tag{17}$$

Proof. (15)-(16) were proved by H.N.Mhaskar et al in [61] and now have been extended to other domains (see e.g.[66,67]). Inequality (17) is proved by [68].

(15)-(16) show the existence of $\Omega^{(n)} \subset S^{d-1}$ which has the polynomial exact QR (8) and satisfies MZ inequality (9). Inequality (17) often goes with (15) and (16) and are needed in discussing approximation order(see e.g. [62]).

Based above analysis we propose the following definition.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

Definition 2.1(*Marcinkiewicz-Zygmund inequalities setting (MZIS)*). We say a given finite node set $\Omega^{(n)} \subset S^{d-1}$ forms a Marcinkiewicz-Zygmund inequality setting if (15)-(16) and (17) simultaneously hold.

Let $\phi \in L^2_{W_\eta}$. For a given finite set $\Omega^{(n)} \subset S^{d-1}$ satisfy Definition 2.1 and the positive numbers $\{\mu_\omega : \omega \in \Omega^{(n)}\}$ are defined as in Proposition 2.2. We define a zonal translation network by

$$H_{\Omega^{(n)}} := cl\{f(x) = \sum_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k) + c_0 : \ c_k \in R, k = 0, 1, 2, \cdots, |\Omega^{(n)}|\},$$

where $T_x(\phi)(y) = \phi(xy)$. Then it is easy to see that

$$H_{\Omega^{(n)}} = H_\phi^{(n)} \bigoplus R,$$

where for $A, B \subset R$ and $A \bigcap B = \{0\}$ we define $A \bigoplus B = \{a + b : a \in A, b \in B\}$ and

$$H_\phi^{(n)} := cl\{f(x) = \sum_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k) : \ c_k \in R, k = 0, 1, 2, \cdots, |\Omega^{(n)}|\}.$$

For $f(x) = \sum_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k), g(x) = \sum_{x_k \in \Omega^{(n)}} d_k \mu_k^{(n)} T_x(\phi)(x_k)$ we define a bivariate operation as

$$\langle f, g \rangle_\phi = \sum_{x_k \in \Omega^{(n)}} c_k \, d_k \, \mu_k^{(n)}$$

and

$$\|f\|_\phi = \left(\sum_{x_k \in \Omega^{(n)}} \mu_k^{(n)} |c_k|^2 \right)^{\frac{1}{2}}.$$

Since (14), we have by Theorem 4 in Chapter 17 of [65] that the matrix $\left(T_{x_i}(\phi, x_j)\right)_{i,j=1,2,\cdots,|\Omega^{(n)}|}$ is positive definite for a given n. It follows that the vector $c = \{\mu_k^{(n)} \ c_k\}_{x_k \in \Omega^{(n)}}$ is uniqueness. Then $(H_\phi^{(n)}, \| \cdot \|_\phi)$ is a Hilbert space which is isometric isomorphism with $l^2(\Omega^{(n)})$,where

$$l^2(\Omega^{(n)}) = \{c = \{\mu_k^{(n)} \ c_k\}_{x_k \in \Omega^{(n)}} : \quad \|c\|_{l^2(\Omega^{(n)})} = \left(\sum_{x_k \in \Omega^{(n)}} \mu_k^{(n)} |c_k|^2 \right)^{\frac{1}{2}} < +\infty\}.$$

Let $C([-1, 1])$ be the set of all continuous functions defined on $[-1, 1]$ and

$$\|\phi\|_{C([-1,1])} = \sup_{x \in [-1,1]} |\phi(x)|.$$

We then have the following proposition.

Proposition 2.3. Let $\phi \in C[-1, 1]$ satisfy $\widehat{a_n^\eta(\phi)} > 0$ for all $n \geq 1$ and $\Omega^{(n)} \subset S^{d-1}$ be an MZIS. Then $(H_\phi^{(n)}, \| \cdot \|_\phi)$ is a reproducing kernel Hilbert space associated with kernel

$$K_x^*(\phi)(y) = K^*(\phi, x, y) = \sum_{x_k \in \Omega^{(n)}} \mu_k^{(n)} T_{x_k}(\phi, x) T_{x_k}(\phi, y), \tag{18}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

i.e.,

$$f(x) = \left\langle f,\ K_x^*(\phi)(\cdot) \right\rangle_\phi \qquad \forall x \in S^{d-1}, \forall f \in H_\phi^{(n)} \tag{19}$$

and there is a constant $k^* > 0$ such that

$$|f(x)| \le k^* \|f\|_\phi \qquad \forall f \in H_\phi^{(n)}, \forall x \in S^{d-1}. \tag{20}$$

Proof. See the proof in Section 5.

Corollary 2.1. Under the assumption of Proposition 2.3, $H_{\Omega^{(n)}}$ is a reproducing kernel Hilbert space associated with the inner product defined by

$$\langle f, g \rangle_{H_{\Omega^{(n)}}} = \langle f_1,\ g_1 \rangle_\phi + c_0 d_0,$$

where

$$f(x) = f_1(x) + c_0, \qquad g(x) = g_1(x) + d_0,$$
$$f_1(x) = \sum_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k), \qquad g_1(x) = \sum_{x_k \in \Omega^{(n)}} d_k \mu_k^{(n)} T_x(\phi)(x_k)\}.$$

and the corresponding reproducing kernel $K_x(\phi)(y)$ is

$$K_x(\phi)(y) = K_x^*(\phi)(y) + 1, \qquad x, y \in S^{d-1}.$$

Furthermore, there is a constant $\kappa > 0$ such that

$$|f(x)| \le \kappa \|f\|_{H_{\Omega^{(n)}}}, \qquad \forall f \in_{H_{\Omega^{(n)}}}, \forall x \in S^{d-1}. \tag{21}$$

Proof. The results can be obtained from Proposition 2.3, Lemma 4.2 and the fact that the real set R is a reproducing kernel Hilbert space whose reproducing kernel is 1 and the inner product is the usual product for two real numbers.

3. Convergence Rate of the Kernel Regularized Regression

We now make a combination of the zonal translation network with the kernel regularized regression.

3.1. Learning Framework

For a set of observations $z = \{(x_i, y_i)\}_{i=1}^m$ drawn i.i.d. according to a joint distribution $\rho(x, y)$ on $Z = S^{d-1} \times Y, Y = [-M, M], M > 0$ is a given real number, satisfying $\rho(x, y) = \rho(y|x)\rho_{S^{d-1}}(x)$ we define a regularized framework as

$$f_{z,\lambda}^{\Omega^{(m)}} := arg \min_{f \in H_{\Omega^{(m)}}} \left(\mathcal{E}_z(f) + \frac{\lambda}{2} \|f\|_{H_{\Omega^{(m)}}}^2 \right), \tag{22}$$

where $\lambda = \lambda(m)$ is the regularization parameter, and

$$\mathcal{E}_z(f) = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2.$$

The general integral model of (22) is

$$f_{\rho,\lambda}^{\Omega^{(m)}} := arg \min_{f \in H_{\Omega^{(m)}}} \left(\mathcal{E}_\rho(f) + \frac{\lambda}{2} \|f\|_{H_{\Omega^{(m)}}}^2 \right), \tag{23}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

where

$$\mathcal{E}_\rho(f) = \int_Z (y - f(x))^2 d\rho.$$

To show the convergence analysis for (22) we need to bound the error

$$\|f_{z,\lambda}^{\Omega^{(m)}} - f_\rho\|_{L^2(\rho_{S^{d-1}})},$$

which is an approximation problem, whose convergence rate depends upon the approximation ability of $\mathcal{H}_{\Omega^{(m)}}$. An error decomposition will be given in Section 3.2.

3.2. Error Decomposition

By (2) and the definition of $f_{\rho,\lambda}^{\Omega^{(m)}}$ we have the following decompositions

$$\begin{aligned}
& \|f_{z,\lambda}^{\Omega^{(m)}} - f_\rho\|_{L^2(\rho_{S^{d-1}})} \\
\leq\ & \|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\|_{L^2(\rho_{S^{d-1}})} + \|f_{\rho,\lambda}^{\Omega^{(m)}} - f_\rho\|_{L^2(\rho_{S^{d-1}})} \\
=\ & \|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\|_{L^2(\rho_{S^{d-1}})} + \sqrt{\mathcal{E}_\rho(f_{\rho,\lambda}^{\Omega^{(m)}}) - \mathcal{E}_\rho(f_\rho) + \frac{\lambda}{2}\|f_{\rho,\lambda}^{\Omega^{(m)}}\|^2_{H_{\Omega^{(m)}}}} \\
\leq\ & \|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\|_{L^2(\rho_{S^{d-1}})} + D^{\Omega^{(m)}}(f_\rho, \lambda)_{L^2(\rho_{S^{d-1}})},
\end{aligned} \tag{24}$$

where

$$D^{\Omega^{(m)}}(f_\rho, \lambda) = \inf_{g \in H_{\Omega^{(m)}}} \left(\|g - f_\rho\|_{L^2(\rho_{S^{d-1}})} + \sqrt{\frac{\lambda}{2}}\|g\|_{H_{\Omega^{(m)}}} \right).$$

We have used the fact that for $a > 0, b > 0, c > 0$ and $0 < p \leq 1$ there holds

$$(a + b + c)^p \leq a^p + b^p + c^p.$$

The K-functional $D^{\Omega^{(m)}}(f_\rho, \lambda)$ shows the approximation error whose decay have been described in [69]. So the main estimate we need to deal with is the sample error

$$\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\|_{L^2(\rho_{S^{d-1}})}.$$

3.3. The Convergence Rate

To give a capacity independent generalization error for algorithm (22), we need some concepts of convex analysis.

Gâteaux differentiable. Let $(H, \|\cdot\|_H)$ be a Hilbert space, $F(f) : H \to R \cup \{\mp\infty\}$ be a real function. We say F is *Gâteaux* differentiable at $f_0 \in H$ if there is an $\xi \in H$ such that for any $g \in H$ there holds

$$\lim_{t \to 0} \frac{F(f_0 + tg) - F(f_0)}{t} = \langle g, \xi \rangle_H$$

and write $F_G'(f_0) = \xi$ or $\nabla_f F(f) = \xi$. It known that for a differentiable convex function $F(f)$ on H $f_0 = arg \min_{f \in H} F(f)$ if and only if $(\nabla_f F(f))|_{f=f_0} = 0$. (see Proposition 17.4 in [70]).

In learning theory, to form an explicit learning rate, one often assume the K-functional has a power convergence rate, i.e., we assume that there exists a $0 < \beta \leq 1$ such that

$$D^{\Omega^{(m)}}(f_\rho,\ \lambda) = O(\lambda^\beta), \qquad \lambda \to 0^+ (m \to +\infty). \tag{25}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

Since $H_{\Omega^{(m)}}$ is a reproducing kernel Hilbert space, the decay of $D^{\Omega^{(m)}}(f_\rho, \lambda)$ can be discussed with the method of [69]. In particular, it is shown in Theorem 2.3 of [71] that one can choose an MZIS $\Omega^{(m)} \subset S^{d-1}, \phi \in L_\eta^{+\infty}$ and suitable real numbers α, β in such a way that

$$D^{\Omega^{(m)}}\Big(f_\rho, \frac{1}{2^{\alpha\, m}}\Big) = O(\frac{1}{2^{m\beta}})$$

if $\partial_\beta f_\rho \in L^2(S^{d-1})$, where we say $\partial_\beta f_\rho \in L^2(S^{d-1})$ if there is a function $\varphi \in L^2(S^{d-1})$ such that

$$\varphi(x) \sim \sum_{l=0}^{+\infty} l^\beta Y_l(f,x), \quad Y_l(f,x) = \sum_{k=1}^{\dim H_l(S^{d-1})} \hat{f}_{k,l}\; Y_k^l(x), \hat{f}_{k,l} = \int_{S^{d-1}} f(u) Y_k^l(u) d\sigma(u),$$

which shows that (25) is reasonable.

Theorem 3.1. *Let $\phi \in C[-1,1]$ satisfy $\widehat{a_l^\eta(\phi)} > 0$ for all $l \geq 0$ and $\Omega^{(m)} \subset S^{d-1}$ be an MZIS. Then for any $\delta \in (0,1)$, with confidence $1-\delta$, hold*

$$\begin{aligned} &\Big\| f_{z,\lambda}^{\Omega^{(m)}} - f_\rho \Big\|_{L^2(\rho_{S^{d-1}})} \\ \leq\;& 4\kappa \Big(\frac{M}{\lambda\sqrt{m}} + \frac{D^{\Omega^{(m)}}(f_\rho,\lambda)_{L^2(\rho_{S^{d-1}})}}{\lambda^{\frac{3}{2}}\sqrt{m}} \Big) log\frac{4}{\delta} + D^{\Omega^{(m)}}(f_\rho,\lambda)_{L^2(\rho_{S^{d-1}})}. \end{aligned} \tag{26}$$

Proof. See the proof in Section 5.

3.4. Conclusion and Discussion

We propose the concept of MZIS for the scattered node set on the unit sphere, with which show that the related convolutional zonal translation network is a reproducing kernel Hilbert space, and show the convergence rate for the kernel regularized least square regression model. We give further analysis to show the advantages of the present paper.

(1)The zonal translation networks which we have chosen is a finite dimensional reproducing kernel Hilbert space, our discussions belong to the scope of kernel method, which is a combination and application of (neural) translation networks to learning theory.

(2)Compared with the existing convergence rate estimate of neural network learning, our upper estimates are dimensional independent (see Theorem in [25], Theorem 3.1 in [72], Theorem 7 in [73], Theorem 1 in [26]).

(3) It is hopeful that the node set in (15)-(16) and (17) may be replaced by a set of uniform distribution observations or a set of random samples (see [72,73]).

(4)The density derivation for the zonal translation network is qualitative, the density is described with the Fourier orthogonal coefficients. We think that this method can be extended to other domains such as the unit ball, the Euclidean space R^d and $R_+ = [0, +\infty)$ et al.

(5)It is hopeful that with the help of the concept MZIS one may show reproducing property for the deep translation networks and thus give investigations for the performance of deep convolutional translation learning with the kernel method (see e.g.[6,7,59]).

4. Some Lemmas

To prove the main results, we need some lemmas.

Lemma 4.1.*Let $(H, \|\cdot\|)$ be a Hilbert space and ξ be a random variable on (Z,ρ) with values in H and $\{z_i\}_{i=1}^m$ be independent samples drawers of ρ.Assume $\|\xi\|_H \leq \widetilde{M} < +\infty$ almost surely. Then, for any $0 < \delta < 1$, with confidence $1-\delta$,it holds*

$$\Big\| \frac{1}{m} \sum_{i=1}^m \xi(z_i) - E(\xi) \Big\|_H \leq \frac{2\widetilde{M}\log(\frac{2}{\delta})}{\sqrt{m}} \tag{27}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

Proof. Find it from [74].

Lemma 4.2. *Let $(H, \|\cdot\|_H)$ be a Hilbert space over X with respect to kernel K. If $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ are closed subspaces of H such that $E \perp F$ and $E \bigoplus F = H$, then $K = L + M$, where L and M are the reproducing kernels of E and F respectively. Moreover, for $h = e + f, e \in E, f \in F$, we have*

$$\|h\|_H = \left(\|e\|_E^2 + \|f\|_F^2\right)^{\frac{1}{2}}.$$

Proof. See Corollary 1 in Chapter 31 of [65] or the Theorem in Section 6 in part I of [75].

Lemma 4.3. *There hold the following equalities:*

$$\nabla_f \mathcal{E}_z(f)(\cdot) = -\frac{2}{m}\sum_{i=1}^m (y_i - f(x_i))K_{x_i}(\phi)(\cdot), \qquad f \in \mathcal{H}_{\Omega^{(m)}}. \tag{28}$$

and

$$\nabla_f \mathcal{E}_\rho(f)(\cdot) = -2\int_Z (y - f(x))K_x(\phi)(\cdot)\, d\rho, \qquad f \in \mathcal{H}_{\Omega^{(m)}}. \tag{29}$$

Proof of (28). By the equality

$$a^2 - b^2 = 2(a-b)\,b + (a-b)^2, \qquad a \in R,\ b \in R \tag{30}$$

we have

$$\begin{aligned} & \lim_{t\to 0}\frac{\mathcal{E}_z(f+tg) - \mathcal{E}_z(f)}{t} \\ = & \lim_{t\to 0}\frac{\frac{1}{m}\sum_{i=1}^m(-2ty_i - f(x_i))g(x_i) + t^2 g(x_i)^2}{t} \\ = & -\frac{2}{m}\sum_{i=1}^m (y_i - f(x_i))g(x_i) \end{aligned}$$

Since $g(x) = \langle g,\ K_x(\phi;\cdot)\rangle_{\mathcal{H}_{\Omega^{(m)}}}$ and the definition of *Gâteaux* derivative, we have by above equality that

$$\lim_{t\to 0}\frac{\mathcal{E}_z(f+tg) - \mathcal{E}_z(f)}{t} = \left\langle g,\ -\frac{2}{m}\sum_{i=1}^m (y_i - f(x_i))g(x_i)K_{x_i}(\phi)(\cdot)\right\rangle_{\mathcal{H}_{\Omega^{(m)}}}.$$

We then have (28). By the same way we can have (29).

Lemma 4.4. Let $(H, \langle\cdot\rangle_H)$ be a Hilbert space consisting of real functions on X. Then

$$\|f\|_H^2 - \|g\|_H^2 = 2\langle f-g, g\rangle_H + \|f-g\|_H^2, \qquad \forall f, g \in H \tag{31}$$

and

$$\nabla_f(\|f\|_H^2)(\cdot) = 2f(\cdot), \qquad \forall f \in H. \tag{32}$$

Proof. (31) is the deformation of the parallelogram formula. (32) can be shown with (31).

Lemma 4.5. (22) has a unique solution $f_{z,\lambda}^{\Omega^{(m)}}$ and (23) has a unique solution $f_{\rho,\lambda}^{\Omega^{(m)}}$. Moreover,
(i) There hold the bounds

$$\|f_{\rho,\lambda}^{\Omega^{(m)}}\|_{C(S^{d-1})} \le \frac{2\kappa\, D^{\Omega^{(m)}}(f_\rho)}{\sqrt{\lambda}}, \tag{33}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

where κ is defined as in (21).

(ii) There holds the equality

$$\lambda f_{z,\lambda}^{\Omega^{(m)}}(\cdot) = \frac{2}{m}\sum_{i=1}^{m} (y_i - f_{z,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot) \tag{34}$$

and the equality

$$\lambda f_{\rho,\lambda}^{\Omega^{(m)}}(\cdot) = 2\int_Z (y - f_{\rho,\lambda}^{\Omega^{(m)}}(x))K_x(\phi)(\cdot)\, d\rho \tag{35}$$

Proof.Proof of (i). Since $\mathcal{E}_\rho(f_{\rho,\lambda}^{\Omega^{(m)}}) \geq \mathcal{E}_\rho(f_\rho)$, we have by (2) that

$$\begin{aligned}
\frac{\lambda}{2}\,\|f_{\rho,\lambda}^{\Omega^{(m)}}\|^2_{\mathcal{H}_{\Omega^{(m)}}} &\leq \mathcal{E}_\rho(f_{\rho,\lambda}^{\Omega^{(m)}}) - \mathcal{E}_\rho(f_\rho) + \frac{\lambda}{2}\,\|f_{\rho,\lambda}^{\Omega^{(m)}}\|^2_{H_{\Omega^{(m)}}} \\
&= \inf_{g\in H_{\Omega^{(m)}}} \left(\mathcal{E}_\rho(g) - \mathcal{E}_\rho(f_\rho) + \frac{\lambda}{2}\,\|g\|^2_{H_{\Omega^{(m)}}}\right) \\
&= \inf_{g\in H_{\Omega^{(m)}}} \left(\|g - f_\rho\|^2_{L^2(\rho_{S^{d-1}})} + \frac{\lambda}{2}\,\|g\|^2_{H_{\Omega^{(m)}}}\right).
\end{aligned} \tag{36}$$

By (36) we have (33).

Proof of (ii). By the definition of $f_{z,\lambda}^{\Omega^{(m)}}$ and (32) we have

$$0 = \nabla_f\left(\mathcal{E}_z(f) + \frac{\lambda}{2}\,\|f\|^2_{\mathcal{H}_{\Omega^{(m)}}}\right)\Big|_{f=f_{z,\lambda}^{\Omega^{(m)}}},$$

i.e.,

$$\begin{aligned}
0 &= \nabla_f\mathcal{E}_z(f)\big|_{f=f_{z,\lambda}^{\Omega^{(m)}}} + \lambda\nabla_f(\frac{1}{2}\,\|f\|^2_{\mathcal{H}_{\Omega^{(m)}}})\big|_{f=f_{z,\lambda}^{\Omega^{(m)}}} \\
&= -\frac{2}{m}\sum_{i=1}^{m}(y_i - f_{z,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot) + \lambda f_{z,\lambda}^{\Omega^{(m)}}(\cdot).
\end{aligned}$$

Hence, (34) holds. We can prove (35) in the same way.

Lemma 4.6. *The solutions $f_{z,\lambda}^{\Omega^{(m)}}$ and $f_{\rho,\lambda}^{\Omega^{(m)}}$ satisfy the inequality*

$$\left\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\right\|_{\mathcal{H}_{\Omega^{(m)}}} \leq \frac{2\,A(z)}{\lambda}, \tag{37}$$

where

$$A(z) = \left\|\int_Z (y - f_{\rho,\lambda}^{\Omega^{(m)}}(x))K_x(\phi)(\cdot)\, d\rho - \frac{1}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot)\right\|_{\mathcal{H}_{\Omega^{(m)}}}.$$

Proof. By (30) we have

$$a^2 - b^2 \geq 2(a-b)\,b, \qquad a\in R,\ b\in R$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

It follows that

$$\begin{aligned} & \mathcal{E}_z(f_{z,\lambda}^{\Omega^{(m)}}) - \mathcal{E}_z(f_{\rho,\lambda}^{\Omega^{(m)}}) \\ = & \frac{1}{m}\sum_{i=1}^{m}\left[(y_i - f_{z,\lambda}^{\Omega^{(m)}}(x_i))^2 - (y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))^2\right] \\ \geq & -\frac{2}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i)) \times \left(f_{z,\lambda}^{\Omega^{(m)}}(x_i) - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i)\right) \\ = & \left\langle f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}},\ -\frac{2}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot)\right\rangle_{\mathcal{H}_{\Omega^{(m)}}}, \end{aligned} \tag{38}$$

where we have used the reproducing property

$$f_{z,\lambda}^{\Omega^{(m)}}(x_i) - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i) = \left\langle f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}},\ K_{x_i}(\phi)(\cdot)\right\rangle_{\mathcal{H}_{\Omega^{(m)}}}.$$

By the definition of $f_{z,\lambda}^{\Omega^{(m)}}$ we have

$$0 \geq \left(\mathcal{E}_z(f_{z,\lambda}^{\Omega^{(m)}}) + \frac{\lambda}{2}\|f_{z,\lambda}^{\Omega^{(m)}}\|^2_{\mathcal{H}_{\Omega^{(m)}}}\right) - \left(\mathcal{E}_z(f_{\rho,\lambda}^{\Omega^{(m)}}) + \frac{\lambda}{2}\|f_{\rho,\lambda}^{\Omega^{(m)}}\|^2_{\mathcal{H}_{\Omega^{(m)}}}\right).$$

On the other hand, by above inequality, (38) and (31) we have

$$\begin{aligned} 0 \geq & \left(\mathcal{E}_z(f_{z,\lambda}^{\Omega^{(m)}}) - \mathcal{E}_z(f_{\rho,\lambda}^{\Omega^{(m)}})\right) + \frac{\lambda}{2}\left(\|f_{z,\lambda}^{\Omega^{(m)}}\|^2_{\mathcal{H}_{\Omega^{(m)}}} - \|f_{\rho,\lambda}^{\Omega^{(m)}}\|^2_{\mathcal{H}_{\Omega^{(m)}}}\right) \\ \geq & \left\langle f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}},\ -\frac{2}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot)\right\rangle_{\mathcal{H}_{\Omega^{(m)}}} \\ + & \left\langle f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}},\ \lambda j_q(f_{\rho,\lambda}^{\Omega^{(m)}})\right\rangle_{\mathcal{H}_{\Omega^{(m)}}} + \lambda\left\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\right\|^2_{\mathcal{H}_{\Omega^{(m)}}}. \end{aligned}$$

Since (35), we have

$$\begin{aligned} 0 \geq\ & 2\left\langle f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}},\ \int_Z (y - f_{\rho,\lambda}^{\Omega^{(m)}}(x))K_x(\phi)(\cdot)\, d\rho \right. \\ & \left. -\frac{1}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot)\right\rangle_{\mathcal{H}_{\Omega^{(m)}}} + \lambda\left\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\right\|^2_{\mathcal{H}_{\Omega^{(m)}}}. \end{aligned}$$

By the Cauchy inequality we have

$$\begin{aligned} & \lambda\left\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\right\|^2_{\mathcal{H}_{\Omega^{(m)}}} \\ \leq\ & 2\left\langle f_{\rho,\lambda}^{\Omega^{(m)}} - f_{z,\lambda}^{\Omega^{(m)}},\ \int_Z (y - f_{\rho,\lambda}^{\Omega^{(m)}}(x))K_x(\phi)(\cdot)\, d\rho \right. \\ & \left. -\frac{1}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot)\right\rangle_{\mathcal{H}_{\Omega^{(m)}}} \\ \leq\ & \left\| \int_Z (y - f_{\rho,\lambda}^{\Omega^{(m)}}(x))K_x(\phi)(\cdot)\, d\rho \right. \\ & \left. -\frac{1}{m}\sum_{i=1}^{m}(y_i - f_{\rho,\lambda}^{\Omega^{(m)}}(x_i))K_{x_i}(\phi)(\cdot)\right\|_{\mathcal{H}_{\Omega^{(m)}}} \times 2\left\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\right\|_{\mathcal{H}_{\Omega^{(m)}}}. \end{aligned}$$

We then have (37).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

Lemma 4.7. *For the solutions $f_{z,\lambda}^{\Omega^{(m)}}$ and $f_{\rho,\lambda}^{\Omega^{(m)}}$ we have*

$$\left\| f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}} \right\|_{\mathcal{H}_{\Omega^{(m)}}} \leq 4\kappa \Big(\frac{M}{\lambda\sqrt{m}} + \frac{D^{\Omega^{(m)}}(f_\rho,\lambda)_{L^2(\rho_{S^{d-1}})}}{\lambda^{\frac{3}{2}}\sqrt{m}} \Big) log \frac{4}{\delta}. \tag{39}$$

Proof. It is easy to see that

$$\begin{aligned} A(z) &\leq \Big\| \int_Z y\, K_x(\phi)(\cdot)\, d\rho - \frac{1}{m}\sum_{i=1}^m y_i\, K_{x_i}(\phi)(\cdot) \Big\|_{\mathcal{H}_{\Omega^{(m)}}} \\ &+ \Big\| \int_Z f_{\rho,\lambda}^{\Omega^{(m)}}(x) K_x(\phi)(\cdot)\, d\rho_{S^{d-1}} - \frac{1}{m}\sum_{i=1}^m f_{\rho,\lambda}^{\Omega^{(m)}}(x_i) K_{x_i}(\phi)(\cdot) \Big\|_{\mathcal{H}_{\Omega^{(m)}}} \\ &= B(z) + C(z). \end{aligned} \tag{40}$$

Take $\xi(x,y)(\cdot) = y\, K_x(\phi)(\cdot)$. Then,

$$\|\xi(x,y)(\cdot)\|_{\mathcal{H}_{\Omega^{(m)}}} = |y||K_x(\phi)(x)| \leq M\kappa. \tag{41}$$

Combining (41) with (27) we have, with confidence $1-\delta$, that

$$B(z) \leq \frac{2M\kappa}{\sqrt{m}}\, log \frac{2}{\delta}. \tag{42}$$

Also, take $\eta(x)(\cdot) = f_{\rho,\lambda}^{\Omega^{(m)}}(x) K_x(\phi)(\cdot)$. Then

$$\begin{aligned} \|\eta(x)\|_{\mathcal{H}_{\Omega^{(m)}}} &= |f_{\rho,\lambda}^{\Omega^{(m)}}(x)||K_x(\phi)(x)| \\ &\leq \frac{4\kappa\, D^{\Omega^{(m)}}(f_\rho,\lambda)_{L^2(\rho_{S^{d-1}})}}{\sqrt{\lambda}} \end{aligned} \tag{43}$$

Combing (43) with (27) we have, with confidence $1-\delta$, that

$$C(z) \leq \frac{4\kappa\, D^{\Omega^{(m)}}(f_\rho,\lambda)_{L^2(\rho_{S^{d-1}})}}{\sqrt{\lambda}\, m} log \frac{2}{\delta}. \tag{44}$$

Collecting (44), (42),(40) and (37), we arrive (39).

5. Proof for Theorems and Propositions

Proof of Theorem 2.1. If $\Delta_\phi^{S^{d-1}}$ is not dense in $L^2(S^{d-1})$, i.e.,

$$clspan(\Delta_\phi^{S^{d-1}}) \neq L^2(S^{d-1}).$$

Then by Proposition 2.1 we know $(clspan(\Delta_\phi^{S^{d-1}}))^\perp \neq \{0\}$ and there is a nonzero functional $F \in L^2(S^{d-1})$ such that

$$F(f) = 0, \qquad f \in clspan(\Delta_\phi^{S^{d-1}}).$$

It follows that $F(\phi(\cdot y)) = 0$ for all $y \in S^{d-1}$. By the Riesz representation Theorem, F corresponds to a nonzero $h \in L^2(S^{d-1})$ in such a way that

$$F(f) = \int_{S^{d-1}} f(x)h(x)d\sigma(x), \qquad \forall f \in L^2(S^{d-1}).$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

Consequently, $\int_{S^{d-1}} \phi(xy)h(y)d\sigma(y) = 0, \quad \forall x \in S^{d-1}$, which gives

$$\int_{S^{d-1}} \Big(\int_{S^{d-1}} \phi(xy)h(y)d\sigma(y) \Big) h(x)d\sigma(x) = 0. \tag{45}$$

Combining (45) with (14) we have

$$\sum_{l=0}^{+\infty} \widehat{a_l^{\eta}(\phi)} \sum_{k=1}^{\dim H_l(S^{d-1})} \left(\int_{S^{d-1}} h(y)Y_k^l(y)d\sigma(y) \right)^2 = 0.$$

It follows that $\int_{S^{d-1}} h(y)Y_k^l(y)d\sigma(y) = 0$ for all $l \geq 0$. Therefore, $h = 0$. We have obtained a contradiction.

Proof of Proposition 2.3. By the the definition of $\langle \cdot, \cdot \rangle_\phi$ and the definition kernel $K_x^*(\phi, y)$ in (18) we have for any $f(x) = \sum\limits_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k)$ that

$$\langle f, K_x^*(\phi, \cdot) \rangle_\phi = \sum_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k) = f(x).$$

(19) then holds. We now show (20). In fact, by the Cauchy's inequality we have

$$\begin{aligned} |f(x)| &= \Big| \sum_{x_k \in \Omega^{(n)}} c_k \mu_k^{(n)} T_x(\phi)(x_k) \Big| \\ &\leq \|c\|_{L^2(\Omega^{(n)})} \left(\sum_{x_k \in \Omega^{(n)}} \mu_k^{(n)} |T_x(\phi)(x_k)|^2 \right)^{\frac{1}{2}} \\ &= \|f\|_\phi \left(\sum_{x_k \in \Omega^{(n)}} \mu_k^{(n)} |\phi(x_k \cdot x)|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

On the other hand, by the localized kernel theory about the Jacobi polynomials in [76] (see Lemma 2.5 in [76]), for a given n, we have an operator $V_n : L^1_{W_\eta} \to \Pi_n[-1,1]$ (where $\Pi_n[-1,1]$ is the set of all the algebraic polynomials of order $\leq n$ defined on $[-1,1]$ and $\eta = \frac{d-2}{2}$) such that $V_n(\phi) \in \Pi_{2n}[-1,1]$, $V_n(g) = g$ for any $g \in \Pi_n[-1,1]$ and there is constant $c > 0$ such that

$$\|V_n(\phi)\|_{C([-1,1])} \leq c\|\phi\|_{C([-1,1])}, \qquad \|V_n(\phi) - \phi\|_{C([-1,1])} \leq cE_n(\phi)_{C([-1,1])},$$

where

$$E_n(\phi)_{C([-1,1])} = \inf_{p \in \Pi_n[-1,1]} \|\phi - p\|_{C([-1,1])}.$$

It follows by the Minkowski inequality and inequality (17) that

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

$$
\begin{aligned}
&\Big(\sum_{x_k\in\Omega^{(n)}} \mu_k^{(n)}|\phi(x_k\cdot x)|^2\Big)^{\frac{1}{2}}\\
\leq\ &\Big(\sum_{x_k\in\Omega^{(n)}} \mu_k^{(n)}|(\phi - V_n(\phi))(x_k\cdot x)|^2\Big)^{\frac{1}{2}} + \Big(\sum_{x_k\in\Omega^{(n)}} \mu_k^{(n)}|V_n(\phi)(x_k\cdot x)|^2\Big)^{\frac{1}{2}}\\
\leq\ & c\,E_n(\phi)_{C([-1,1])}\big(\sum_{x_k\in\Omega^{(n)}} \mu_k^{(n)}\big)^{\frac{1}{2}} + O(1)\Big(\int_{S^{d-1}}|V_n(\phi)(x\cdot y)|^2\,d\sigma(x)\Big)^{\frac{1}{2}}\\
=\ & O(1)E_n(\phi)_{C([-1,1])} + O(1)\Big(\int_{-1}^{1}|V_n(\phi)(u)|^2\,W_\eta(u)du\Big)^{\frac{1}{2}}\\
=\ & O(1).
\end{aligned}
$$

(20) thus holds, where we have used the Funk-Hecke formula (see (1.2.11) in [63])

$$
\int_{S^{d-1}} f(x\cdot y)Y_n(y)d\sigma(y) = \lambda_n(f)\,Y_n(x), \qquad Y_n\in H_n(S^{d-1})
$$

and

$$
\lambda_n(f) = \omega_{d-1}\int_{-1}^{1} f(t)\frac{C^\eta(t)}{C^\eta(1)}dt, \quad \eta = \frac{d-1}{2}.
$$

Proof of Theorem 3.2. Since (39) we have by (21) that

$$
\begin{aligned}
&\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\|_{L^2(\rho_{S^{d-1}})}\\
\leq\ &\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\|_{C(S^{S^{d-1}})}\\
\leq\ &\kappa\Big\|f_{z,\lambda}^{\Omega^{(m)}} - f_{\rho,\lambda}^{\Omega^{(m)}}\Big\|_{\mathcal{H}_{\Omega^{(m)}}}\\
\leq\ &4\kappa\Big(\frac{M}{\lambda\sqrt{m}} + \frac{D^{\Omega^{(m)}}(f_\rho,\lambda)_{L^2(\rho_{S^{d-1}})}}{\lambda^{\frac{3}{2}}\sqrt{m}}\Big)log\frac{4}{\delta}.
\end{aligned}
\tag{46}
$$

Taking (46) into (24) we have (26).

Author Contributions: Conceptualization, Ran. X.X.; methodology, Sheng B.H.; validation, Ran. X.X.; formal analysis, Ran. X.X.; resources, Sheng B.H. and Wang S.H.; writing—original draft preparation, Ran. X.X. and Sheng B.H.; writing—review and editing, Sheng B.H. and Wang S.H.; supervision, Sheng B.H.. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China under Grants No.61877039, the NSFC/RGC Joint Research Scheme (Project No. 12061160462 and $N_$CityU102/20) of China and Natural Science Foundation of Jiangxi Province of China(20232BAB201021).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks.*Communications of the ACM.*, **2017**, *60(6)*, 84-90.
2. Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey,W.; Cao, Y.; Gao, Q. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv* **2016**, arXiv:1609.08144v2 [cs.CL].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

3. Alipanahi, B.; Delong, A.; Weirauch, M.T.; Frey, B.J. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning.*Nature Biotechnology.* **2015**,*33(8)*, 831-838.
4. Chui, C.K. ; Lin, S.B.; Zhou, D. X. Construction of neural networks for realization of localized deep learning. *arXiv***2018** , arXiv:1803.03503 [cs.LG]. https://doi.org/10.48550/arXiv.1803.03503.
5. Chui, C. K. ; Lin, S.B.; Zhou, D.X. Deep neural networks for rotation-invariance approximation and learning. *Anal. and Appl.* **2019**, *17(5)*, 737-772.
6. Fang,Z.Y.; Feng, H.; Huang, S.; Zhou, D.X. Theory of deep convolutional neural networks II: spherical analysis. *Neural Networks.* **2020**, *131*, 154-162.
7. Feng, H.; Huang, S.; Zhou, D.X. Generalization analysis of CNNs for classification on spheres.*IEEE Transactions on Neural Networks and Learning Systems.* Doi: 10.1109/TNNLS.2021.3134675.
8. Zhou, D. X. Deep distributed convolutional neural networks: universality. *Anal. Appl.* **2018**, *(16)*, 895-919.
9. Zhou, D. X. Universality of deep convolutional neural networks. *Appl. Comput. Harmon. Anal.* **2020**, *48(2)*, 787-794.
10. Cucker, F.; Smale, S. On the mathematical foundations of learning. *Bull.Amer.Math.Soc.***2001**, *39*, 1-49.
11. An, C.P.; Chen, X.J. ; Sloan, I.H.; Womersley, R.S. Regulatized least squares approximations on the sphere using spherical designs. *SIAM J.Numer.Anal.* **2012**, *50(3)*, 1513-1534.
12. An, C.P.; Wu, H.N. Lasso hyperinterpolation over general regions.*SIAM J. Sci. Comput.* **2021**, *43(601)*, A3967-A3991.
13. An, C.P.; Ran, J.S. Hard thresholding hyperinterpolation over general regions.*arXiv* **2023**, arXiv:2209.14634v2 [math.NA].
14. De Mol, C.; De Vito, E. ; Rosasco, L. Elastic-net regularization in learning theory. *J. Complexity*,**2009**, *25*, 201-230.
15. Fischer, S.; Steinwart, I. Sobolev norm learning rates for regularized least-squares algorithms. *J.Mach.Learn.Res.***2020**, *21(1)*, 8464-8501.
16. Lai, J.F.; Li, Z.F. ; Huang, D.G. ; Lin, Q. The optimality of kernel classifiers in Sobolev space.*arXiv* **2024**, arXiv:2402.01148v1 [math.ST].
17. Sun, H.W.; Wu, Q. Least square regression with indefinite kernels and coefficient regularization. *Appl. Comput. Harmon. Anal.* **2011**,*30*, 96-109.
18. Wu, Q. ; Zhou, D.X. Learning with sample dependent hypothesis spaces. *Comput. Math. Appl.* **2008**, *56* , 2896-2907.
19. Chen, H.; Wu, J.T.; Chen, D.R. Semi-supervised learning for regression based on the diffusion matrix (in Chinese).*Sci Sin Math.* **2014**, *44*, 399-408.
20. Sun, X.J.; Sheng, B.H. The learning rate of kernel regularized regression associated with a correntropy-induced loss.*Adv. in Math.(Beijing)* **2024**, *53(3)*, 633-652.
21. Wu, Q.; Zhou, D.X. Analysis of support vector machine classification. *J. Comput.Anal. Appl.* **2006**, *8(2)*, 99-119.
22. Cucker, F. ; Zhou, D. X. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press, New York, 2007.
23. Sheng, B. H. Reproducing property of bounded linear operators and kernel regularized least square regressions. *Int. J. Wavelets Multiresolution Inf. Process.* **2024** *2450013 (42 pages)*, Doi: 10.1142/S0219691324500139.
24. Steinwart I.; and Christmann, A. *Support Vector Machines*,Springer-Verlag, New York, 2008.
25. Lin, S.B.; Wang, D.; Zhou, D.X. Sketching with spherical designs for noisy data fitting on spheres. *SIAM J.Sci.Comput.* **2024**, *46(1)*, A313-A337.
26. Lin, S.B.; Zeng, J.S.; Zhang, X.Q. Constructive neural network learning. *IEEE Trans. on Cybernetics*. **2019**, *49(1)*, 221-232.
27. Mhaskar, H.N.; Micchelli, C.A. Degree of approximation by neural and translation networks with single hidden layer. *Adv. Appl. Math.* **1995**, *16(1)*, 151-183.
28. Sheng, B.H.; Zhou,S.P.; Li, H.T. On approximation by tramslation networks in $L^p(R^k)$ spaces.*Adv. in Math.(Beijing)* **2007**, *36(1)*, 29-38.
29. Mhaskar, H. N.; Narcowich, F. J.; Ward, J. D.Approximation properties of zonal function networks using scattered data on the sphere. *Adv. Comput. Math.* **1999**,*11(2-3)*, 121-137.
30. Sheng, B. H. On approximation by reproducing kernel spaces in weighted L^p-spaces. *J. Syst. Sci. Complex.* **2007**, *20 (4)*, 623-638.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

31. Narcowich, F.J.; Ward, J.D.; Wendland, H. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constr.Approx.* **2006**, *24*, 175-186.
32. Narcowich, F.J.; Ward, J.D. Scattered data interpolation on spheres: error estimates and locally supported basis functions. *SIAM J.Math.Anal.* **2002**, *33(6)*, 1393-1410.
33. Narcowich,F.J.; Sun, X.P.; Ward, J.D.; Wendland, H. Direct and inverse Sobolev error estimates for scattered data interpolation via spherical basis functions.*Found.Comput.Math.***2007**, *7(3)*,369-390.
34. Gröchenig, K. Sampling, Marcinkiewicz-Zygmund inequalities, approximation and quadrature rules. *J.Approx.Theory* **2020**, *257* , 105455.
35. Mhaskar, H.N.; Narcowich, F.J.; Sivakumar, N.; Ward, J.D. Approximation with interpolatory constraints. *Proc.Amer.Math. Soc.* **2001**,*130(5)*, 1355-1364.
36. Marzo, J. Marcinkiewicz-Zygmund inequalities and interpolation by spherical harmonics. *J. Funct. Anal.* **2007**, *250*, 559-587.
37. Marzo, J.; Pridhnani, B. Sufficiant conditions for sampling and interpolation on the sphere. *Constr. Approx.* **2014**, *40*, 241-257.
38. Wang, H.P. Marcinkiewicz-Zygmund inequalities and interpolation by spherical polynomials with respect to doubling weights. *J.Math.Anal.Appl.* **2015**, *423*,1630-1649.
39. Gia, T.L.; Slon, I.H. The nuiform norm of hyperinterpolation on the unit sphere in an arbitrary number of dimensions em Constr. Approx. **2001**, *17*, 249-265.
40. Sloan, I.H. Polynomial interpolation and hyperinterpolation over general regions. *J.Approx.Theory* **1995** , *83*, 238-254.
41. Sloan, I.H.; Womersley, R.S. Constructive polynomial approximation on the sphere. *J.Approx. Theory* **2000**, *103*, 91-118.
42. Wang, H.P. Optimal lower estimates for the worst case cubature error and the approximation by hyperinterpolation operators in the Sobolev space sertting on the sphere. *Int. J. Wavelets Multiresolution Inf. Process.***2009**, *7(6)*, 813-823.
43. Wang, H.P.; Wang, K.; Wang, X.L. On the norm of the hyperinterpolation operator on the d-dimensional cube. *Comput. Appl.* **2014**, *68*, 632-638.
44. Sloan, I.H.; Womersley, R.S. Filtered hyperinterpolation: a constructive polynomial approximation on the sphere. *Int.J.Geomath.* **2012**, *3*, 95-117.
45. Bondarenko, A.; Radchenko, D.; Viazovska, M. Well-seperated spherical designs. *Constr. Approx.***2015**, *41*, 93-112.
46. Hesse, K.; Womersley, R.S. Numerical integration with polynomial exactness over a spherical cap. *Adv.Math.Math.* **2012**, *36(3)*, 451-483.
47. Delsarte, P.; Goethals, J.M.; Seidel, J.J. Spherical codes and designs. *Geom.Dedicata*. **1977**, *6*, 363-388.
48. An, C.P.; Chen, X.J.; Sloan, I.H.; Womersley, R.S. Well conditioned spherical designs for integration and interpolation on the two-sphere. *SIAM J. Numer.Anal.* **2010**, *48(6)*, 2135-2157.
49. Chen, X.; Frommer, A.; Lang, B. Computational existence proof for spherical t-designs. *Numer. Math.* **2010**, *117(2)*,289-305.
50. An, C.P.; Wu, H.N. Bypassing the quadrature exactness assumption of hyperinterpolation on the sphere. *J.Complexity*. **2024**, *80*, 101789.
51. An, C.P.; Wu, H.N. On the quadrature exactness in hyperinterpolation. *BIT Numer. Math.* **2022**, *62*, 1899-1919.
52. Sheng, B.H. The covering numbers for some periodic reproducing kernel spaces *Acta Math. Scientia* **2009**, *29A(6)*, 1590-1600.
53. Sheng, B.H. Estimate the norm of the Mercer kernel matrices with discrete orthogonal transforms. *Acta Math. Hungar*. **2009**, *122(4)*, 339-355.
54. Sheng, B.H.; Wang, J.L.; Li, P. On the covering number of some Mercer kernel Hilbert spaces. *J. Complexity* **2008** , *24(2)*, 241-258.
55. Zhang, C.P.; Sheng, B.H.; Chen, Z.X. Applications of Bernstein -Durrmeyer operators in estimating the norm of Mercer kernel matrices *Anal. Theory and Appl.***2008**, *24(1)*,74-86.
56. Sun, X. J. ; Sheng, B. H.; Liu, L.; Pan, X. L. On the density of translation networks defined on the unit ball. *Math. Found. Comput.*, **2023**, doi:10.3934/mfc.2023017.
57. Parhi, R.; Nowak, R.D. Banach space representer theorems for neural networks and ridge splines. *J. Mach. Learn Res.* **2021**, *22(43)*, 1-40.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 August 2024 doi:10.20944/preprints202408.0275.v1

58. Oono, K.; Suzuki, Y.J. Approximation and non-parameteric estimate of ResNet-type convolutional neural networks.*arXiv* **2023**, arXiv:1903.10047v4 [stat.ML].
59. Shen, G.H.; Jiao, Y.L.; Lin, Y.Y.; Huang, J. Non-asymptotic excess risk bounds for classification with deep convolutional neural networks. *arXiv* **2021**, arXiv:2105.00292 [cs.LG].
60. Mallat, S. Understanding deep convolutional networks. *Phil. Trans. R. Soc.* **2016**, *374A*, 20150203.
61. Mhaskar, H.N.; Narcowich, F.J.; Ward, J.D. Spherical Marcinkiewicz-Zygmund inequalities and positive quadratue.*Math.Comput.***2001**, *70*,1113-1130. (Corrigendum:*Math.Comp.* **2001**,*(71)*, 453-454.)
62. Wang, H.P.; Wang, K. Optimal recovery of Besov classes of generalized smoothness and Sobolev class on the sphere. *J. Complexity* **2016**, *32(1)* , 40-52.
63. Dai, F.; Xu, Y. *Approximation Theory and Harmonic Analysis on Spheres and Balls.* Springer, New York, 2013.
64. Müller, C. *Spherical Harmonic*, Springer-Verlag, Berlin, 1966.
65. Cheney, W.; Light, W.*A Course in Approximation Theory*, China Machine Press, Beijing, 2004.
66. Brown, G.; Dai, F. Approximation of smooth functions on compact two-point homogeneous spaces. *J. Funct. Anal.* **2005**, *220*, 401-422.
67. Dai, F.; Wang, H.P. Positive cubature formulas and Marcinkiewicz-Zygmund inequalities on spherical caps. *Constr. Approx.* **2010** , *31(1)*, 1-36.
68. Dai, F. On generalized hyperinterpolation on the sphere. *Proc. Amer. Math. Soc.* **2006**, *134(10)*, 2931-2941.
69. Sheng, B. H.; Wang, J. L. Moduli of smoothness, K-functionals and Jackson-type inequalities associated with kernel function approximation in learning theory. *Anal. Appl.* **2024**, https: doi.org/10.1142/S021953052450009X.
70. Bauschke, H. H.; Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2010.
71. Sheng, B.H.; Xiang, D.H. The convergence rate for a K-functional in learning theory. *J. Inequal. Appl.* **2010**, *Artical ID 249507*, *18 pages*.
72. Lin, S.B.; Wang, Y.G.; Zhou, D.X. Distributed filtered hyperinterpolation for noisy data on the sphere.*SIAM J. Numer. Anal.* **2021**, *59(2)*, 634-659.
73. Montúfar, G.; Wang, Y.G. Distributed learning via filtered hyperinterpolation on manifolds.*Found.Comput.Math.* **2022** , *22*, 1219-1271.
74. Smale, S.; Zhou, D.X. Learning theory estimates via integral operators and their applications. *Constr. Approx.* **2007**, *26*, 153-172.
75. Aronszajn, N. Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **1950**, *68*, 337-404.
76. Kyriazis, G.; Petrushev, P.; Xu, Y. Jacobi decomposition of weighted Triebel-Lizorkin and Besov spaces. *arXiv* **2006**, arXiv:math/0610624 [math.CA].

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.