Hindawi Publishing Corporation
Journal of Inequalities and Applications
Volume 2010, Article ID 249507, 18 pages
doi:10.1155/2010/249507

*Research Article*

# The Convergence Rate for a $K$-Functional in Learning Theory

**Bao-Huai Sheng[1] and Dao-Hong Xiang[2]**

[1] *Department of Mathematics, Shaoxing University, Shaoxing, Zhejiang 312000, China*
[2] *Department of Mathematics, Zhejiang Normal University, Jinhua, Zhejiang 321004, China*

Correspondence should be addressed to Bao-Huai Sheng, bhsheng@usx.edu.cn

Received 11 November 2009; Accepted 21 February 2010

Academic Editor: Yuming Xing



It is known that in the field of learning theory based on reproducing kernel Hilbert spaces the upper bounds estimate for a $K$-functional is needed. In the present paper, the upper bounds for the $K$-functional on the unit sphere are estimated with spherical harmonics approximation. The results show that convergence rate of the $K$-functional depends upon the smoothness of both the approximated function and the reproducing kernels.

## 1. Introduction

It is known that the goal of learning theory is to approximate a function (or some function features) from data samples.

Let $X$ be a compact subset of $n$-dimensional Euclidean spaces $\mathfrak{R}^n$, $Y \subset \mathfrak{R}$. Then, learning theory is to find a function $f$ related the input $x \in X$ to the output $y \in Y$ (see [1–3]). The function $f(x)$ is determined by a probability distribution $\rho(x, y) = \rho(y \mid x)\rho_X(x)$ on $Z := X \times Y$, where $\rho_X(x)$ is the marginal distribution on $X$ and $\rho(y \mid x)$ is the condition probability of $y$ for a given $x$.

Generally, the distribution $\rho(x, y)$ is known only through a set of sample $z := \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m$ independently drawn according to $\rho(x, y)$. Given a sample $z$, the regression problem based on Support Vector Machine (SVM) learning is to find a function $f_z : X \to Y$ such that $f_z(x)$ is a good estimate of $y$ when a new input $x$ is provided. The binary classification problem based on SVM learning is to find a function $\psi_z : X \to \{-1, 1\}$ which divides $X$ into two parts. Here $\psi_z$ is often induced by a real-valued function $f_z$ with the form of $\psi_z(x) = \mathrm{sgn}(f_z(x))$ where $\mathrm{sgn}(f_z(x)) = 1$ if $f_z(x) \geq 0$, otherwise, $-1$. The functions $f_z$ are often generated from the following Tikhonov regularization scheme (see, e.g., [4–9]) associated with a reproducing kernel Hilbert space (RKHS) $\mathscr{H}_K$ (defined below) and a sample $z$:

$$f_z := f_{z,\lambda,\mathscr{H}_K} = \arg\min_{f\in\mathscr{H}_K}\left(\frac{1}{m}\sum_{i=1}^{m}V_p(y_if(x_i)) + \lambda\|f\|^2_{\mathscr{H}_K}\right), \tag{1.1}$$

where $\lambda > 0$ is a positive constant called the regularization parameter and $V_p(t) = (1-t)^p_+ = \max\{1-t, 0\}^p$ $(p \geq 1)$ called $p$-norm SVM loss.

In addition, the Tikhonov regularization scheme involving offset $b$ (see, e.g., [4, 10, 11]) can be presented below with a similar way to (1.1)

$$f_z := f_{z,\lambda,\mathscr{H}_K} = \arg\min_{f=f^*+b,\, f^*\in\mathscr{H}_K,\, b\in\mathfrak{R}}\left(\frac{1}{m}\sum_{i=1}^{m}V_p(y_if(x_i)) + \lambda\|f^*\|^2_{\mathscr{H}_K}\right). \tag{1.2}$$

We are in a position to define reproducing kernel Hilbert space. A function $K : X\times X \to \mathfrak{R}$ is called a Mercer kernel if it is continuous, symmetric, and positive semidefinite, that is, for any finite set of distinct points $\{x_1, x_2, \ldots, x_l\} \subset X$, the matrix $(K(x_i, x_j))^l_{i,j=1}$ is positive semidefinite.

The reproducing kernel Hilbert space (RKHS) $\mathscr{H}_K$ (see [12]) associated with the Mercer kernel $K$ is defined to be the closure of the linear span of the set of functions $\{K_x := K(x,\cdot) : x \in X\}$ with the inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}_K} = \langle\cdot,\cdot\rangle_K$ satisfying $\langle K_x, K_y\rangle_K = K(x,y)$ and the reproducing property

$$\langle K_x, g\rangle_K = g(x), \quad \forall x \in X,\ g \in \mathscr{H}_K. \tag{1.3}$$

If $g(x) = \sum_i c_iK(x,x_i)$, then $\|g\|^2_K = \sum_{i,j} c_ic_jK(x_i,x_j)$. Denote $C(X)$ as the space of continuous function on $X$ with the norm $\|\cdot\|_\infty$. Let $\kappa := \sqrt{\|K\|_\infty}$. Then the reproducing property tells that

$$\|g\|_\infty \leq \kappa\|g\|_K, \quad \forall g \in \mathscr{H}_K. \tag{1.4}$$

It is easy to see that $\mathscr{H}_K$ is a subset of $C(X)$. We say that $K(x,y)$ is a universal kernel if for any compact subset $X$ $\mathscr{H}_K$ is dense in $C(X)$ (see [13, Page 2652]).

Let $\wedge \subset X$ be a given discrete set of finite points. Then, we may define an RKHS $(\mathscr{H}^\wedge_K, \|\cdot\|_{\mathscr{H}^\wedge_K})$ by the linear span of the set of functions $\{K_x := K(x,\cdot) : x \in \wedge\}$. Then, it is easy to see that $\mathscr{H}^\wedge_K \subset \mathscr{H}_K$ and for any $f \in \mathscr{H}^\wedge_K$ there holds $\|f\|_{\mathscr{H}^\wedge_K} = \|f\|_K$.

Define $\mathcal{E}(f) := \int_Z V_p(yf(x))d\rho(x,y)$ and $f^{V_p}_\rho := \arg\min\mathcal{E}(f)$, where the minimum is taken over all measurable functions. Then, to estimate the explicit learning rate, one needs to estimate the regularization errors (see, e.g., [4, 7, 9, 14])

$$\mathfrak{D}(\lambda) := \inf_{f\in\mathscr{H}_K}\left\{\mathcal{E}(f) - \mathcal{E}\left(f^{V_p}_\rho\right) + \lambda\|f\|^2_{\mathscr{H}_K}\right\}, \tag{1.5}$$

$$\overline{\mathfrak{D}}(\lambda) := \inf_{f=f^*+b,\, f^*\in\mathscr{H}_K,\, b\in\mathfrak{R}}\left\{\mathcal{E}(f) - \mathcal{E}\left(f^{V_p}_\rho\right) + \lambda\|f^*\|^2_{\mathscr{H}_K}\right\}. \tag{1.6}$$

The convergence rate of (1.5) is controlled by the $K$-functional (see, e.g., [9])

$$K(f,\lambda)_{\mathcal{H}_K} = \inf_{g\in\mathcal{H}_K}\left(\|f-g\|_{p,d\rho_X} + \lambda\|g\|_K^2\right), \quad \lambda > 0,\ f \in L^p(d\rho_X), \tag{1.7}$$

and (1.6) is controlled by another $K$-functional (see, e.g., [4])

$$\overline{K}(f,\lambda)_{\overline{\mathcal{H}_K}} = \inf_{g\in\overline{\mathcal{H}_K},\, g=g^*+b,\, g^*\in\mathcal{H}_K,\, b\in\mathfrak{R}}\left(\|f-g\|_{p,d\rho_X} + \lambda\|g^*\|_K^2\right), \tag{1.8}$$

where $\lambda > 0, f \in L^p(d\rho_X) = \{f(x) : \|f\|_{p,d\rho_X} < +\infty\}$ with

$$\|f\|_{p,d\rho_X} = \begin{cases} \left(\int_X |f(x)|^p d\rho_X(x)\right)^{1/p}, & 1 \le p < +\infty, \\ \operatorname*{ess\,sup}_{x\in X}|f(x)|, & p = +\infty. \end{cases} \tag{1.9}$$

We notice that, on one hand, the $K$-functionals (1.7) and (1.8) are the modifications of the $K$-functional of interpolation theory (see [15]) since the interpolation relation (1.4). On the other hand, they are different from the usual $K$-functionals (see e.g., [16–30]) since the term $\|g\|_K^2$. However, they have some similar point. For example, if $K(x,y)$ is a universal kernel, $\mathcal{H}_K$ is dense in $L^p(d\rho_X)$ (see e.g., [31]). Moreover, some classical function spaces such as the polynomial spaces (see [2, 32]) and even some Sobolev spaces may be regarded as RKHS (see e.g., [33]).

In learning theory we often require $K(f,t)_{\mathcal{H}_K} = O(t^\beta)$ and $\overline{K}(f,t)_{\overline{\mathcal{H}_K}} = O(t^\beta)$ for some $\beta > 0$ (see e.g., [1, 7, 14]). Many results on this topic have been achieved. With the weighted Durrmeyer operators [8, 9] showed the decay by taking $K(x,y)$ to be the algebraic polynomials kernels on $[0,1]\times[0,1]$ or on the simplex in $\mathfrak{R}^2$.

However, in general case, the convergence of $K$-functional (1.8) should also be considered since the offset often has influences on the solution of the learning algorithms (see e.g., [6, 11]). Hence, the purpose of this paper is twofold. One is to provide the convergence rates of (1.7) and (1.8) when $K(x,y)$ is a general Mercer kernel on the unit sphere $S^q$ and $1 \le p \le +\infty$. The other is how to construct functions of the type of

$$f(x) = \beta_0 + g^*(x) = \beta_0 + \sum_{i=1}^m \beta_i K(x, x_i), \quad x \in \mathrm{X}, \tag{1.10}$$

to obtain the convergence rate of (1.8). The translation networks constructed in [34–37] have the form of (1.10) and the zonal networks constructed in [38, 39] have the form of (1.10) with $\beta_0 = 0$. So the methods used by these references may be used here to estimate the convergence rates of (1.7) and (1.8) if one can bound the term $\|g^*\|_K^2$.

In the present paper, we shall give the convergence rate of (1.7) and (1.8) for a general kernel defined on the unit sphere $S^n = \{x \in \mathfrak{R}^{n+1} : \|x\|_{\mathfrak{R}^{n+1}} = 1\}$ and $\rho(x,y) = \rho(y \mid x)\mu_n(x)$ with $\mu_n(x)$ being the usual Lebesgue measure on $S^n$. If there is a distortion between $\rho_{S^n}(x)$

and $\mu_n(x)$, the convergence rate of (1.7)-(1.8) in the general case may be obtained according to the way used by [1, 8].

The rest of this paper is organized as follows. In Section 2, we shall restate some notations on spherical harmonics and present the main results. Some useful lemmas dealing with the approximation order for the de la Vallée means of the spherical harmonics, the Gauss integral formula, the Marcinkiewicz-Zygmund with respect to the scattered data obtained by G. Brown and F. Dai and a result on the zonal networks approximation provided by H. N. Mhaskar will be given in Section 3. A kind of weighted norm estimate for the Mercer kernel matrices on the unit sphere will be given in Lemma 3.8. Our main results are proved in the last section.

Throughout the paper, we shall write $A = O(B)$ if there exists a constant $C > 0$ such that $A \leq CB$. We write $A \sim B$ if $A = O(B)$ and $B = O(A)$.

## 2. Notations and Results

To state the results of this paper, we need some notations and results on spherical harmonics.

### 2.1. Notations

For integers $l \geq 0$, $q \geq 1$, the class of all one variable algebraic polynomials of degree $\leq l$ defined on $[-1, 1]$ is denoted by $P_l$, the class of all spherical harmonics of degree $l$ will be denoted by $H_l^q$, and the class of all spherical harmonics of degree $l \leq n$ will be denoted by $\prod_n^q$. The dimension of $H_l^q$ is given by (see [40, Page 65])

$$d_l^q = \dim H_l^q = \begin{cases} \dfrac{2l+q-1}{l+q-1}\dbinom{l+q-1}{q-1}, & l \geq 1, \\ 1, & l = 0, \end{cases} \tag{2.1}$$

and that of $\Pi_n^q$ is $\sum_{l=0}^{n} d_l^q$. One has the following well-known addition formula (see [41, Page 10, Theorem 2]):

$$\sum_{k=1}^{d_l^q} Y_{l,k}(x)Y_{l,k}(y) = \frac{d_l^q}{\omega_q} p_l^{q+1}(xy), \quad l = 0, 1, \ldots, \; x, y \in S^q, \tag{2.2}$$

where $p_l^{q+1}(x)$ is the degree-$l$ generalized Legendre polynomial. The Legendre polynomials are normalized so that $p_l^{q+1}(1) = 1$ and satisfy the orthogonality relations

$$\int_{-1}^{1} p_l^{q+1}(x) p_k^{q+1}(x) W_q(x)dx = \frac{\omega_q}{\omega_{q-1} d_l^q} \delta_{l,k}, \quad W_q(x) = \left(1 - x^2\right)^{q/2-1}. \tag{2.3}$$

Define $L^p(S^q)$ and $L^p_{W_{\alpha,\beta}}$ by taking $\rho_X$ to be the usual volume element of $S^q$ and the Jacobi weights functions $W_{\alpha,\beta}(x) = (1-x)^\alpha(1+x)^\beta$, $\alpha > -1$, $\beta > -1$, respectively. For any $\phi \in L^1_{W_q}$ we have the following relation (see [42, Page 312]):

$$\int_{S^q} \phi(xy)d\mu_q(x) = \omega_{q-1}\int_{-1}^{1} \phi(x)W_q(x)dx. \tag{2.4}$$

The orthogonal projections $Y_k(f,x)$ of a function $f \in L^1(S^q)$ on $H_k^q$ are defined by (see e.g., [43])

$$Y_k(f,x) = \frac{d_k^q}{\omega_q}\int_{S^q} p_k^{q+1}(xy)f(y)d\mu_q(y), \quad x \in S^q, \tag{2.5}$$

where $xy$ denotes the inner product of $x$ and $y$.

### 2.2. Main Results

Let $\phi \in L^1_{W_q}$ satisfy $a_l(\phi) = \omega_{q-1}\int_{-1}^{1}\phi(x)p_l^{q+1}(x)W_q(x)dx > 0$ and $\sum_{l=0}^{+\infty} a_l(\phi)d_l^q < +\infty$. Define

$$K(\phi,x,y) = K(\phi,xy) = \sum_{l=0}^{+\infty} a_l(\phi)\frac{d_l^q}{\omega_q}p_l^{q+1}(xy), \quad a_l(\phi) > 0,\ x,y \in S^q. \tag{2.6}$$

Then, by [44, Chapter 17] we know that $K(\phi,x,y)$ is positive semidefinite on $S^q$ and the right of (2.6) is convergence absolutely and uniformly since $|p_l^{q+1}(x)| \le 1$. Therefore, $K(\phi,x,y)$ is a Mercer kernel on $S^q$. By [13, Theorem 10] we know that $K(\phi,x,y)$ is a universal kernel on $S^q$. We suppose that there is a constant $C_p > 0$ depending only on $p$ such for any $y \in S^q$

$$\|K(\phi,\cdot,y)\|_{p,S^q} \le C_p\|\phi\|_{p,W_q}, \quad 1 \le p \le +\infty,\ \phi \in L^p_{W_q}. \tag{2.7}$$

Given a finite set $\wedge$, we denote by $\sharp\wedge$ the cardinality of $\wedge$. For $r > 0$ and $a \ge 1$ we say that a finite subset $\wedge \subset S^q$ is an $(r,a)$-covering of $S^q$ if

$$S^q \subset \bigcup_{\omega\in\wedge} B(\omega,r), \qquad \max_{\omega\in\wedge}\sharp(\wedge \cap B(\omega,r)) \le a, \tag{2.8}$$

where $B(\omega,r) = \{x \in S^q : d(\omega,x) \le r\}$ with $d(\omega,x) := \arccos(\omega,x)$ being the geodesic distance between $x$ and $\omega$.

Let $S \geq 1$ be an integer, $a = \{a_n\}_{n=0}^{+\infty}$ a sequence of real numbers. Define forward difference operators by $\Delta a_n = \Delta^1 a_n := a_{n+1} - a_n$, $\Delta^r := \Delta(\Delta^{r-1} a_n)$, $r = 2, 3, \ldots$, $\Delta^0 a_n := a_n$,

$$\begin{aligned} |a|_S &:= \sup_{0 \leq r \leq S,\, v \geq 0} (v+1)^r |\Delta^r a_v|, \\ |a|_S^* &= \sum_{r=1}^{S} \sum_{v=0}^{+\infty} (v+1)^{r-1} |\Delta^r a_v|. \end{aligned} \tag{2.9}$$

We say a finite subset $\wedge \subset S^q$ is a subset of interpolatory type if for any real numbers $\{y_\omega\}_{\omega \in \wedge}$ there is a $p \in \Pi_n^q$ such that $p(\omega) = y_\omega$, $\omega \in \wedge$. This kind of subsets may be found from [45, 46].

Let $\mathcal{B}_S$ be the set of all sequence $a$ for which $|a|_S < +\infty$ and $\mathcal{B}_S^*$ the set of all sequence $a$ for which $|a|_S + |a|_S^* < +\infty$.

Let $\beta > 0$ be a real number, $f \in L^p(S^q)$. Then, we say $\partial_\beta f \in L^p(S^q)$ if there is a function $\varphi \in L^p(S^q)$ such that

$$\varphi(x) \sim \sum_{k=0}^{+\infty} k^\beta Y_k(f, x), \quad x \in S^q. \tag{2.10}$$

We now give the results of this paper.

**Theorem 2.1.** *If there is a constant $\gamma_0 > 0$ depending only on $q$ such that $\wedge$ is a subset of interpolatory type and a $(\delta/6(2^m), a)$-covering of $S^q$ satisfying $0 < \delta < \gamma_0/a$ with $a \geq 1$ and $m$ being a given positive integer. $S > q$ is an integer. $\beta > 0$ is a real number such that there is $\gamma' \geq \gamma + 2q - 1$ and $\beta - \gamma' > q$, $\phi \in L_{W_q}^\infty$ satisfies $\{(l+1)^\beta a_l(\phi)\}_{l=0}^{+\infty} \in \mathcal{B}_S$ and $\{(l+1)^{-\beta} a_l^{-1}(\phi)\}_{l=0}^{+\infty} \in \mathcal{B}_S^*$. $\mathscr{H}_{K(\phi)}^\wedge$ is the reproducing kernel space reproduced by $\wedge$ and the kernel (2.6). $\partial_\gamma f \in L^p(S^q)$. Then there is a constant $C_{p,q} > 0$ depending only on $p$ and $q$ and a function $g_{m,\phi}(x) = C_0 + g_{m,\phi}^*(x)$ with $g_{m,\phi}^*(x) \in \mathscr{H}_{K(\phi)}^\wedge$ and $C_0$ a constant such that*

$$\left\| f - g_{m,\phi} \right\|_{p,S^q} = O\left( \frac{1}{2^{m\gamma}} \right), \tag{2.11}$$

$$\left\| g_{m,\phi}^* \right\|_{\mathscr{H}_{K(\phi)}^\wedge}^2 = O\left( 2^{2m(2q-1)} \right). \tag{2.12}$$

The functions $\phi$ satisfying the conditions of Theorem 2.1 may be found in [39, Page 357].

**Corollary 2.2.** *Under the conditions of Theorem 2.1. If $\partial_\gamma f \in L^p(S^q)$, then*

$$\overline{K}\left( f, \frac{1}{2^{2m\gamma'}} \right)_{\overline{\mathscr{H}_{K(\phi)}^\wedge}} = O\left( \frac{1}{2^{m\gamma}} \right). \tag{2.13}$$

Corollary 2.2 shows that the convergence rate of the $K$-functional (1.8) is controlled by the smoothness of both the reproducing kernels and the approximated function $f$.

**Theorem 2.3.** *If there is a constant $\gamma > 0$ depending only on $q$ such that $\wedge$ is a subset of interpolatory type and a $(\delta/6(2^m), a)$-covering of $S^q$ satisfying $0 < \delta < \gamma/a$ with $a \geq 1$ and $m$ being a given positive integer. $\mathscr{H}^{\wedge}_{K(\phi)}$ is the reproducing kernel space reproducing by $\wedge$ and the kernel* (2.6) *with $\phi \in L^{\infty}_{W_q}$ satisfying $\{(l+1)^{\beta} a_l(\phi)\} \in \mathcal{B}_S$ and $\{(1/(l+1)^{\beta}) a_l^{-1}(\phi)\} \in \mathcal{B}_S^*$. Then, for $\alpha \geq q(1/p - 1/2)_+ + \beta$ and $\partial_{\beta} f \in L^p(S^q)$ there holds*

$$K\left(f, \frac{1}{2^{\alpha m}}\right)_{\mathscr{H}^{\wedge}_{K(\phi)}} = O\left(\frac{1}{2^{m\beta}}\right), \tag{2.14}$$

*where* $(a)_+ = \max(a, 0)$.

## 3. Some Lemmas

To prove Theorems 2.1 and 2.3, we need some lemmas. The first one is about the Gauss integral formula and Marcinkiewicz inequalities.

**Lemma 3.1** (see [47–50]). *There exist constants $\gamma > 0$ depending only on $q$ such that for any positive integer $n$ and any $(\delta/n, a)$-covering $\wedge$ of $S^q$ satisfying $0 < \delta < \gamma/a$, there exists a set of real numbers $\lambda_\omega \sim n^{-q}$, $(\omega \in \wedge)$ such that*

$$\int_{S^q} f(y) d\mu_q(y) = \sum_{\omega \in \wedge} \lambda_\omega f(\omega) \tag{3.1}$$

*for any $f \in \Pi^q_{3n}$, and for $f \in \Pi^q_n$*

$$\|f\|_{p,S^q} \sim \begin{cases} \left(\sum_{\omega \in \wedge} \lambda_\omega |f(\omega)|^p\right)^{1/p}, & 0 < p < +\infty, \\ \max_{\omega \in \wedge} |f(\omega)|, & p = +\infty, \end{cases} \tag{3.2}$$

*where $\sharp\wedge \sim n^q \sim \dim \Pi^q_n$, the constants of equivalence depending only on $q$, $a$, $\delta$, and $p$ when $p$ is small. Here one employs the slight abuse of notation that $0^0 = 1$.*

The second lemma we shall use is the Nikolskii inequality for the spherical harmonics.

**Lemma 3.2** (see [38, 45, 49, 51, 52]). *If $0 < p < r \leq +\infty$, $p \in \Pi^q_n$, then one has the following Nikolskii inequality:*

$$\|p\|_{p,S^q} \leq C(q) \|p\|_{r,S^q} \leq C(q) n^{q(1/p - 1/r)} \|p\|_{p,S^q}, \tag{3.3}$$

*where the constant $C(q)$ depends only on $q$.*

We now restate the general approximation frame of the Cesàro means and de la Vallée Poussin means provided by Dai and Ditzian (see [53]).

**Lemma 3.3.** *Let $\mu$ be a positive measure on X. $\{H_k\}_{k=0}^{+\infty}$ is a sequence of finite-dimensional spaces satisfying the following:*

(I) $H_k \subset L^p(d\mu)$.

(II) $H_k$ *is orthogonal to* $H_l$*(in* $L^p(d\mu)$*) when* $l \neq k$.

(III) $\operatorname{span}(\bigcup)_{k=0}^{\infty} H_k)$ *is dense in* $L^p(d\mu)$ *for all* $1 \leq p \leq +\infty$.

(IV) $H_0$ *is the collection of the constants.*

*The Cesàro means $C(\delta)$ of $f(x)$ is given by*

$$\sigma_N^{\delta}(f,x) = \frac{1}{A_N^{\delta}} \sum_{k=0}^{N} A_{N-k}^{\delta} P_k(f,x), \qquad A_N^{\delta} = \frac{\Gamma(k+\delta+1)}{\Gamma(k+1)\Gamma(\delta+1)} \tag{3.4}$$

*for $N = 1, 2, \ldots$, where*

$$P_k(f,x) = \int_X f(y) \left\{ \sum_{l=1}^{d_k} \varphi_{k,i}(x) \varphi_{k,i}(y) \right\} d\mu(y), \quad x \in X, \tag{3.5}$$

*and $\{\varphi_{k,i}(x)\}_{i=1}^{d_k}$ is an orthogonal base of $H_k$ in $L^2(d\mu)$. One sets,for a given $\alpha > 0$, $\partial_\alpha f(x) \sim \sum_{k=1}^{\infty} k^\alpha P_k(f,x)$ and $\partial_\alpha f \in L^p(d\mu)$ if there exists $\varphi \in L^p(d\mu)$ such that $P_k(\varphi,x) = k^\alpha P_k(f,x)$.*

*Let $\eta$ be defined as $\eta(u) \in C^\infty, \eta(u) = 1$ for $u \in [0,1/2]$ and $\eta(u) = 0$ for $u > 1$ and is a nonegative and nonincrease function. $\rho_N(f,x)$ are the de la Vallée Poussin means defined as*

$$\rho_N(f,x) = \sum_{k=0}^{+\infty} \eta\left(\frac{k}{2N}\right) P_k(f,x), \quad x \in X, \ f \in L^1(d\mu). \tag{3.6}$$

*Then, $\rho_N(f,x) = f(x)$, $f \in \bigcup_{k \leq N} H_k$. If for some $\delta > 0$, $\|\sigma_N^\delta(f)\|_{p,d\mu} \leq C\|f\|_{p,d\mu}$, $1 \leq p \leq +\infty$, then, $\|\rho_N(f)\|_{p,d\mu} \leq C\|f\|_{p,d\mu}$ and*

$$\|\rho_N(f) - f\|_{p,d\mu} = O\left(\frac{\|\partial_\alpha f\|_{p,d\mu}}{N^\alpha}\right), \quad \partial_\alpha f \in L^p(d\mu). \tag{3.7}$$

Lemma 3.3 makes the following Lemma 3.4.

**Lemma 3.4.** *Let $\eta$ be the function defined as in Lemma 3.3. Define two kinds of operators, respectively, by*

$$\begin{aligned} V_N(f,x) &= \sum_{k=0}^{+\infty} \eta\left(\frac{k}{2N}\right) Y_k(f,x), \quad x \in S^q, \ f \in L^1(S^q), \\ V_N^*(f,x) &= \sum_{k=0}^{+\infty} \eta\left(\frac{k}{2N}\right) a_l(f) \frac{d_l^q}{\omega_q} p_l^{q+1}(x), \quad x \in [-1,1], \ f \in L^1_{W_q}. \end{aligned} \tag{3.8}$$

*Then,* $V_N(f,x) = f(x)$ *for any* $f \in \Pi_N^q$ *and* $V_N^*(f,x) = f(x)$ *for any* $f \in P_N$. *Moreover,*

$$\|V_N(f) - f\|_{p,S^q} = O\left(\frac{\|\partial_\alpha f\|_{p,S^q}}{N^\alpha}\right), \quad \partial_\alpha f \in L^p(S^q), \tag{3.9}$$

$$\|V_N^*(f) - f\|_{p,W_q} = O\left(\frac{\|\partial_\alpha f\|_{p,W_q}}{N^\alpha}\right), \quad \partial_\alpha f \in L^p_{W_q}, \tag{3.10}$$

*where for* $f \in L^p_{W_q}$ *one defines*

$$\partial_\alpha f(x) \sim \sum_{k=1}^{\infty} k^\alpha a_k(f) \frac{d_l^q}{\omega_q} p_l^{q+1}(x) \quad x \in [-1,1]. \tag{3.11}$$

*Proof.* By [54, Lemma 2.2] we know $\|V_N(f)\|_{p,S^q} \le C\|f\|_{p,S^q}$ for some $C > 0$. Hence, (3.9) holds by (3.7). By [19, Theorem A] we know $\|\sigma_N^\delta(f)\|_{p,W_q} \le C\|f\|_{p,W_q}$ for $\delta > q/2-1/2$. Hence, (3.10) holds by (3.7).

Let $\wedge \subset S^q$ be a finite set. Then we call $\wedge$ an M-Z quadrature measure of order $n$ if (3.1) and (3.2) hold for $f \in \Pi_n^q$. By this definition one knows the finite set $\wedge$ in Lemma 3.1 is an M-Z quadrature measure of order $n$.

Define an operator as

$$\sigma_y(\wedge, \eta, f, x) = \sum_{l=0}^{+\infty} \eta\left(\frac{l}{y}\right) \frac{d_l^q}{\omega_q} \sum_{\omega \in \wedge} \lambda_\omega f(\omega) p_l^{q+1}(\omega x), \quad x \in S^q. \tag{3.12}$$

Then, we have the following results. □

**Lemma 3.5** (see [39]). *For a given integer* $n \ge 1$, *let* $\wedge$ *be an M-Z quadrature measure of order* $6(2^n)$, $\phi \in L^1_{W_q}$, $S > q$ *an integer,* $1 \le p \le +\infty$, $\beta > q/p'$, *where* $p'$ *satisfies* $1/p + 1/p' = 1$ *which satisfies* $p' = 1$ *if* $p = +\infty$ *and* $p' = +\infty$ *if* $p = 1$. $\eta(t)$ *defined in Lemma 3.3 is a nonnegative and non-increasing function. Let* $\phi \in L^1_{W_q}$ *satisfy* $\{(l+1)^{-\beta} a_l^{-1}(\phi)\}_{l=0}^{+\infty} \in \mathcal{B}_S^*$. *Then, for* $f \in H_{\gamma,p}$, $0 < \gamma \le \beta$, *where* $H_{\gamma,p}$ *consists of* $f \in L^p(S^q)$ *for which the derivative of order* $\gamma$; *that is,* $\partial\gamma f$, *belongs to* $L^p(S^q)$. *Then, there is an operator* $\mathfrak{D}_\phi : H_{\gamma,p} \to L^P(S^q)$ *such that*

(i) *(see [39, Proposition 3.1, (b)]).* $\|\mathfrak{D}_\phi f\|_{p,S^q} \le c\|\partial\gamma f\|_{p,S^q}$ *for* $f \in H_{\beta,p}$

$$f(x) = \int_{S^q} \phi(xy) \mathfrak{D}_\phi f(y) d\mu_q(y), \quad x \in S^q, \tag{3.13}$$

$$\widehat{f}(l,k) = a_l(\phi) \widehat{\mathfrak{D}_\phi f}(l,k), \tag{3.14}$$

*where* $\widehat{f}(l,k) = \int_{S^q} f(x) Y_{l,k}(x) d\mu_q(x)$.

(ii) *(see [39, Theorem 3.1]). Moreover, if one adds an assumption that* $\{(l+1)^{\beta} a_l(\phi)\}_{l=0}^{+\infty} \in \mathcal{B}_S$, *then, there are constants* $C > 0$ *and* $C_1 > 0$ *such that*

$$\left\| f(\cdot) - \sum_{\omega\in\wedge} \lambda_\omega \mathfrak{D}_\phi \sigma_{2^n}(\wedge,\eta,f,\omega)\phi(\cdot\omega) \right\|_{p,S^q} \le C 2^{-n\gamma} \|\partial\gamma f\|_{p,S^q} \tag{3.15}$$

*and for* $0 \le \gamma \le \beta$

$$\|\partial_\gamma f - \partial_\gamma \sigma_n(\wedge,\eta,f)\|_{p,S^q} \le C_1 n^{\gamma-\beta} \|\partial_\beta f\|_{p,S^q}. \tag{3.16}$$

**Lemma 3.6** (see e.g., [29, Page 230]). *Let* $f \in L^2(S^q)$. *Then,*

$$\|f\|_{2,S^q} = \left( \sum_{k=0}^{+\infty} \|Y_k f\|_{2,S^q}^2 \right)^{1/2}. \tag{3.17}$$

Following Lemma 3.7 deals with the orthogonality of the Legendre polynomials $p_l(xy)$.

**Lemma 3.7.** *For the generalized Legendre polynomials* $p_l^{q+1}$ *one has*

$$\frac{d_l^q}{\omega_q} \int_{S^q} p_k^{q+1}(uy) p_l^{q+1}(xu)\, d\mu_q(u) = p_k^{q+1}(xy), \quad x,y \in S^q. \tag{3.18}$$

*Proof.* It may be obtained by (2.2). □

**Lemma 3.8.** *Let* $\phi \in L^\infty_{W_q}$ *satisfy* (2.7) *for* $p = +\infty$ *and* $\partial_q \phi \in L^\infty_{W_q}$. $\wedge \subset S^q$ *is a finite set satisfying the conditions of Theorem 2.1. Then, there is a constant* $C_\phi > 0$ *depending only on* $\phi$ *such that*

$$\left( \sum_{\omega,\omega'\in\wedge} \lambda_\omega \lambda_{\omega'} |K(\phi,\omega',\omega)|^2 \right)^{1/2} \le C_\phi < +\infty. \tag{3.19}$$

*Proof.* Define a matrix by $K_\wedge^{(\sqrt{\lambda})} = (\sqrt{\lambda_\omega} K_{(\wedge)}(\omega',\omega) \sqrt{\lambda_{\omega'}})_{\omega,\omega'\in\wedge}$, where $K_{(\wedge)}(\omega',\omega) = \sum_{l=0}^{\sharp\wedge} \lambda_l \times (d_l^q/\omega_q) p_l^{q+1}(\omega'\omega)$ with $\lambda_l \ge 0$ and $l_\wedge^2 = \{a = (a_\omega)_{\omega\in\wedge} : (\sum_{\omega\in\wedge} a_\omega^2)^{1/2} < +\infty\}$. Then,

$$\left( \sum_{\omega,\omega'\in\wedge} \lambda_\omega \lambda_{\omega'} |K_{(\wedge)}(\omega',\omega)|^2 \right)^{1/2} = \left\| K_\wedge^{(\sqrt{\lambda})} \right\|_{l_\wedge^2} = \sup_{a\neq 0,\, a\in l_\wedge^2} \frac{a^\top K_\wedge^{(\sqrt{\lambda})} a}{a^\top a}. \tag{3.20}$$

By the Parseval equality we have

$$
\begin{aligned}
a^{\top} K_{\wedge}^{(\sqrt{\lambda})} a &= \sum_{\omega,\omega'\in\wedge} a_{\omega}\sqrt{\lambda_{\omega}} K_{(\wedge)}(\omega',\omega) a_{\omega'}\sqrt{\lambda_{\omega'}} \\
&= \sum_{\omega,\omega'\in\wedge} a_{\omega}\sqrt{\lambda_{\omega}} \left( \sum_{l=0}^{\sharp\wedge} \lambda_l \times \frac{d_l^q}{\omega_q} p_l^{q+1}(\omega'\omega) \right) a_{\omega'}\sqrt{\lambda_{\omega'}} \\
&= \sum_{l=0}^{\sharp\wedge} \lambda_l \sum_{\alpha=0}^{d_l^q} \left| \sum_{\omega\in\wedge} a_{\omega}\sqrt{\lambda_{\omega}} Y_{l,\alpha}(\omega) \right|^2 \\
&\leq \max_{0\leq l\leq\sharp\wedge} \lambda_l \int_{S^q} \left| \sum_{l=0}^{\sharp\wedge} \sum_{\alpha=0}^{d_l^q} \left( \sum_{\omega\in\wedge} a_{\omega}\sqrt{\lambda_{\omega}} Y_{l,\alpha}(\omega) \right) Y_{l,\alpha}(x) \right|^2 d\mu_q(x) \\
&= \max_{0\leq l\leq\sharp\wedge} \lambda_l \int_{S^q} \left| \sum_{\omega\in\wedge} a_{\omega}\sqrt{\lambda_{\omega}} \sum_{l=0}^{\sharp\wedge} \frac{d_l^q}{\omega_q} p_l^{q+1}(x\omega) \right|^2 d\mu_q(x).
\end{aligned} \tag{3.21}
$$

Let $L_{\wedge}(x) \in \Pi_n^q$ satisfy $L_{\wedge}(\omega) = a_{\omega}/\sqrt{\lambda_{\omega}}$, $\omega \in \wedge$. Then, by (3.1)

$$
\begin{aligned}
a^{\top} K_{\wedge}^{(\sqrt{\lambda})} a &= \sum_{\omega,\omega'\in\wedge} a_{\omega}\sqrt{\lambda_{\omega}} K_{(\wedge)}(\omega',\omega) a_{\omega'}\sqrt{\lambda_{\omega'}} \\
&\leq \max_{0\leq l\leq\sharp\wedge} \lambda_l \int_{S^q} \left| \sum_{l=0}^{\sharp\wedge} \frac{d_l^q}{\omega_q} \int_{S^q} L_{\wedge}(\omega) p_l^{q+1}(x\omega) d\mu_q(\omega) \right|^2 d\mu_q(x) \\
&= \max_{0\leq l\leq\sharp\wedge} \lambda_l \int_{S^q} |L_{\wedge}(x)|^2 d\mu_q(x) \\
&= \max_{0\leq l\leq\sharp\wedge} \lambda_l \sum_{\omega\in\wedge} \lambda_{\omega} |L_{\wedge}(x)|^2 = \max_{0\leq l\leq\sharp\wedge} \lambda_l a^{\top} a.
\end{aligned} \tag{3.22}
$$

Hence, $a^{\top} K_{\wedge}^{(\sqrt{\lambda})} a \leq \max_{0\leq l\leq\sharp\wedge} \lambda_l a^{\top} a$. On the other hand, since $\|K(\phi,x,\cdot)\|_{\infty,S^q} \leq C\|\phi\|_{\infty,W_q}$, $x \in S^q$, we have for any $x \in S^q$ that

$$
\left| K(\phi,x,y) - K\left(V_{\sharp\wedge}^{*}(\phi),x,y\right) \right| = \left| K\left(\phi - V_{\sharp\wedge}^{*}(\phi),x,y\right) \right| \leq C \left\| \phi - V_{\sharp\wedge}^{*}(\phi) \right\|_{\infty,W_q}. \tag{3.23}
$$

It follows for $x, y \in S^q$ that

$$
K\left(V_{\sharp\wedge}^{*}(\phi),x,y\right) - C\left\|\phi - V_{\sharp\wedge}^{*}(\phi)\right\|_{\infty,W_q} \leq K(\phi,x,y) \leq K\left(V_{\sharp\wedge}^{*}(\phi),x,y\right) + C\left\|\phi - V_{\sharp\wedge}^{*}(\phi)\right\|_{\infty,W_q}. \tag{3.24}
$$

Define $K_{\wedge}^{(\sqrt{\lambda})}(\phi) = (\sqrt{\lambda_\omega} K(\phi, \omega', \omega)\sqrt{\lambda_{\omega'}})_{\omega,\omega' \in \wedge}$. Then, (3.24), (3.10), the Cauchy inequality, and the fact $\sharp\wedge \sim n^q$ make

$$\left| a^\top K_{\wedge}^{(\sqrt{\lambda})}(\phi) a - a^\top K_{\wedge}^{(\sqrt{\lambda})}\left(V_{\sharp\wedge}^*(\phi)\right) a \right| = O\left(\sharp \wedge \left\| \phi - V_{\sharp\wedge}^*(\phi) \right\|_{\infty, W_q}\right) \times a^\top a = O(1) a^\top a. \quad (3.25)$$

It follows that

$$\begin{aligned} a^\top K_{\wedge}^{(\sqrt{\lambda})}(\phi) a &\leq a^\top K_{\wedge}^{(\sqrt{\lambda})}\left(V_{\sharp\wedge}^*(\phi)\right) a + \left| a^\top K_{\wedge}^{(\sqrt{\lambda})}(\phi) a - a^\top K_{\wedge}^{(\sqrt{\lambda})}\left(V_{\sharp\wedge}^*(\phi)\right) a \right| \\ &\leq \max_{0 \leq l \leq \sharp\wedge} \eta\left(\frac{l}{2\sharp\wedge}\right) a^\top a + O(1) a^\top a. \end{aligned} \quad (3.26)$$

Equation (3.2) thus holds. □

## 4. Proof of the Main Results

We now show Theorems 2.1 and 2.3, respectively.

*Proof of Theorem 2.1.* Lemma 4.3 in [39] gave the following results.

*Let* $1 \leq p \leq +\infty$, $\gamma' > q/p'$, $S > q$ be an integer, and a sequence of real numbers such $a^{\gamma'} := \{(l+1)^{\gamma'} a_l\} \in \mathcal{B}_S$. Then, there exists $\phi \in L_{W_q}^p$ such that $a_l(\phi) \in a_l$, $l = 0, 1, 2, \ldots$.

Since $(1+l)^\beta = (1+l)^{\gamma'}(1+l)^{\beta-\gamma'}$ and $\beta - \gamma' > q$, we have a $\phi^* \in L_{W_q}^\infty$ such that $a_l(\phi^*) = (1+l)^{\gamma'} a_l(\phi)$. Hence, $\partial_{\gamma'}\phi \in L_{W_q}^\infty$ and

$$K(V_{2^{m+1}}^*(\phi), x, y) = \sum_{l=0}^{+\infty} a_l(V_{2^{m+1}}^*(\phi)) \frac{d_l^q}{\omega_q} p_l^{q+1}(xy), \quad x, y \in S^q, \quad (4.1)$$

and for $0 \leq k \leq 2^{m+1}$ there holds for $x, y \in S^q$ that

$$\frac{d_k^q}{\omega_q} \int_{S^q} K(V_{2^{m+1}}^*(\phi), x, u) p_1^{q+1}(uy) d\mu_q(u) = a_k(V_{2^{m+1}}^*(\phi)) \frac{d_k^q}{\omega_q} p_k^{q+1}(xy) = a_k(\phi) \frac{d_k^q}{\omega_q} p_k^{q+1}(xy). \quad (4.2)$$

It follows for $x, y \in S^q$ that

$$p_k^{q+1}(xy) = \frac{\int_{S^q} K\left(V_{2^{m+1}}^*(\phi), x, u\right) p_l^{q+1}(uy) d\mu_q(u)}{a_k(\phi)} \quad (4.3)$$

On the other hand, since

$$V_{2^m}(f, x) = Y_0(V_{2^m}(f), x) + \sum_{k=1}^{2^{m+1}} Y_k(V_{2^m}(f), x), \quad x \in S^q, \quad (4.4)$$

where for $1 \le k \le 2^{m+1}$, we have by (4.3)

$$\Upsilon_k(V_{2^m}(f), x) = \frac{d_k^q}{\omega_q a_k(\phi)} \int_{S^q} \left( K(V^*_{2^{m+1}}(\phi), x, u) p_l^{q+1}(uy) d\mu_q(u) \right) \times V_{2^m}(f, y) d\mu_q(y). \quad (4.5)$$

Hence, above equation and (3.1)-(3.2) makes

$$\begin{aligned} V_{2^m}(f, x) &= \Upsilon_0(V_{2^m}(f), x) \\ &\quad + \sum_{k=1}^{2^{m+1}} \frac{d_k^q}{\omega_q a_k(\phi)} \int_{S^q} \left( \int_{S^q} K(V^*_{2^{m+1}}(\phi), x, u) p_k^{q+1}(uy) d\mu_q(u) \right) \\ &\quad \times V_{2^m}(f, y) d\mu_q(y) \\ &= C_0 + \sum_{\omega \in \wedge} \lambda_\omega C_{\omega,\phi} K(V^*_{2^{m+1}}(\phi), x, \omega), \end{aligned} \quad (4.6)$$

where $C_0 = \Upsilon_0(V_{2^m}(f), x)$, $C_{\omega,\phi} = \sum_{k=1}^{2^{m+1}} (d_k^q/\omega_q)(\Upsilon_k(V_{2^m}(f), \omega)/a_k(\phi))$. Define

$$g_{m,\phi}(x) = C_0 + g^*_{m,\phi}(x) = C_0 + \sum_{\omega \in \wedge} \lambda_\omega C_{\omega,\phi} K(\phi, x, \omega), \quad x \in S^q. \quad (4.7)$$

Then, we know $g^*_{m,\phi}(x) \in \mathfrak{R} + \mathscr{H}^{\wedge}_{K(\phi)}$ and by (3.9)

$$\begin{aligned} \left\| g_{m,\phi} - f \right\|_{p,S^q} &\le \left\| g_{m,\phi} - V_{2^m}(f) \right\|_{p,S^q} + \left\| V_{2^m}(f) - f \right\|_{p,S^q} \\ &= O\left( \frac{1}{2^{m\gamma}} \right) + \left\| g_{m,\phi}(f) - V_{2^m}(f) \right\|_{p,S^q}, \end{aligned} \quad (4.8)$$

where

$$\left| g_{m,\phi}(x) - V_{2^m}(f, x) \right| \le \sum_{\omega \in \wedge} \left| \lambda_\omega C_{\omega,\phi} \right| \left| K(V^*_{2^{m+1}}(\phi) - \phi, x, \omega) \right|. \quad (4.9)$$

It follows by (3.9) that

$$\begin{aligned} \left\| g_{m,\phi} - V_{2^m}(f) \right\|_{p,S^q} &\le \sum_{\omega \in \wedge} \left| \lambda_\omega C_{\omega,\phi} \right| \left\| K(V^*_{2^{m+1}}(\phi) - \phi, \cdot, \omega) \right\|_{p,S^q} \\ &= O\left( \frac{1}{2^{m\gamma'}} \right) \sum_{\omega \in \wedge} \left| \lambda_\omega C_{\omega,\phi} \right|. \end{aligned} \quad (4.10)$$

On the other hand, by the definition of $V_{2^m}(f)$ and (3.14) we have for $1 \le k \le 2^{m+1}$ that

$$\Upsilon_k(V_{2^m}(f), \omega) = \eta\left( \frac{k}{2^{m+1}} \right) a_k(\phi) \Upsilon_k(\mathfrak{D}_\phi f, \omega), \quad \omega \in S^q, \quad (4.11)$$

where $\mathfrak{D}_\phi$ denotes the operator $\mathfrak{D}_\phi$ of Lemma 3.5 for $\phi(xy) = K(\phi, xy)$. Hence,

$$C_{\omega,\phi} = \sum_{k=1}^{2^{m+1}} \frac{d_k^q}{\omega_q} \eta\left(\frac{k}{2^{m+1}}\right) Y_k(\mathfrak{D}_\phi f, \omega), \quad \omega \in S^q. \tag{4.12}$$

Equation (3.2) and the definition of $\eta(t)$ make

$$\begin{aligned} \sum_{\omega\in\wedge} \left|\lambda_\omega C_{\omega,\phi}\right| &\le \sum_{\omega\in\wedge} \lambda_\omega \left| \sum_{k=1}^{2^{m+1}} \frac{d_k^q}{\omega_q} \eta\left(\frac{k}{2^{m+1}}\right) Y_k(\mathfrak{D}_\phi f, \omega) \right| \\ &\le C_q \sum_{k=1}^{2^{m+1}} d_k^q \int_{S^q} \left|Y_k(\mathfrak{D}_\phi f, \omega)\right| d\mu_q(\omega). \end{aligned} \tag{4.13}$$

The Hölder inequality, the (i) of Lemma 3.5, and the fact that $|p_l^{q+1}(x)| \le 1$ make $|Y_k(\mathfrak{D}_\phi f, \omega)| \le C_q d_k^q \|\mathfrak{D}_\phi f)\|_{p,S^q} \le C_q d_k^q \|\partial_Y f\|_{p,S^q}$. Therefore,

$$\sum_{\omega\in\wedge} \left|\lambda_\omega C_{\omega,\phi}\right| \le C_q \sum_{k=1}^{2^{m+1}} \left(d_k^q\right)^2 \left\|\partial_Y f\right\|_{p,S^q}. \tag{4.14}$$

Take $g_{m,\phi}^*(x) = \sum_{\omega\in\wedge} \lambda_\omega C_{\omega,\phi} K(\phi, x, \omega)$, then

$$\begin{aligned} \left\|g_{m,\phi}^*\right\|_{\mathscr{H}_{K(\phi)}^\wedge}^2 &= \sum_{\omega,\omega'\in\wedge} \lambda_\omega \lambda_{\omega'} C_{\omega,\phi} C_{\omega',\phi} K(\phi, \omega', \omega) \\ &\le \left(\sum_{\omega\in\wedge} \lambda_\omega \left|C_{\omega,\phi}\right|^2\right) \left(\sum_{\omega,\omega'\in\wedge} \lambda_\omega \lambda_{\omega'} \left|K(\phi, \omega', \omega)\right|^2\right)^{1/2}. \end{aligned} \tag{4.15}$$

Equations (3.2), (3.17), (3.16), and the Cauchy inequality make

$$\begin{aligned} \sum_{\omega\in\wedge} \lambda_\omega \left|C_{\omega,\phi}\right|^2 &\le C \left\| \sum_{k=1}^{2^{m+1}} \frac{d_k^q}{\omega_q} \frac{Y_k(V_m(f), \omega)}{a_k(\phi)} \right\|_{2,S^q}^2 \\ &\le C \left( \sum_{k=1}^{2^{m+1}} \left(d_k^q\right)^2 \left\|\partial_Y f\right\|_{p,S^q} \right)^2. \end{aligned} \tag{4.16}$$

Let $\Gamma(x)$ be the Gamma function. Then, it is well known that $\Gamma(x)/\Gamma(x+a) \sim x^{-a}$, $x \to +\infty$. Therefore,

$$d_l^q = \frac{2l+q-1}{l+q-1} \binom{l+q-1}{q-1} = \frac{(2l+q-1)\Gamma(l+q-1)}{\Gamma(q)\Gamma(l+1)} \sim \frac{2l+q-1}{\Gamma(q)} l^{q-2} \sim l^{q-1}. \tag{4.17}$$

Hence,

$$\sum_{k=1}^{2^{m+1}}\left(d_k^q\right)^2 = O\left(\sum_{l=1}^{2^{m+1}} l^{2(q-1)}\right) = O\left(2^{(m+1)(2q-1)}\right), \quad m \longrightarrow +\infty. \tag{4.18}$$

Equations (4.14) and (4.4) make

$$\left\|g_{m,\phi} - V_{2^m}(f)\right\|_{p,S^q} = O\left(\frac{1}{2^{m(\gamma'-2q+1)}}\right), \tag{4.19}$$

and hence

$$\left\|f - g_{m,\phi}\right\|_{p,S^q} = O\left(\frac{1}{2^{m\gamma}} + \frac{1}{2^{m(\gamma'-2q+1)}}\right). \tag{4.20}$$

Since $\gamma' \geq 2q - 1 + \gamma$, we have (2.11) by (4.20). Equation (2.12) follows by (4.3), (4.4), and (3.19). □

*Proof of Corollary 2.2.* By (2.11)-(2.12) one has

$$\begin{aligned}\overline{K}\left(f, \frac{1}{2^{2m\gamma'}}\right)_{\overline{\mathscr{H}_{K(\phi)}}} &\leq \left\|f - g_{m,\phi}\right\|_{p,S^q} + \frac{1}{2^{2m\gamma'}}\left\|g^*_{m,\phi}\right\|^2_{\mathscr{H}_{K(\phi)}} \\ &= O\left(\frac{1}{2^{m\gamma}}\right) + \frac{1}{2^{2m\gamma'}} \times O\left(2^{2m(2q-1)}\right) = O\left(\frac{1}{2^{m\gamma}}\right).\end{aligned} \tag{4.21}$$

□

*Proof of Theorem 2.3.* Take the place of $\phi(x\omega)$ in Lemma 3.5 with $K(\phi, x\omega)$, denote still by $\mathfrak{D}_\phi$ the operator $\mathfrak{D}_\phi$ in Lemma 3.5 with $\phi(xy) = K(\phi, xy)$ and

$$g_{m,\phi}(x) = \sum_{\omega\in\wedge}\lambda_\omega \mathfrak{D}_\phi \sigma_{2^m}(\wedge, \eta, f, \omega) K(\phi, x\omega), \quad x \in S^q, \tag{4.22}$$

then, $g_{m,\phi} \in H^\wedge_{K(\phi)}$ and by (3.15) $\|f - g_{m,\phi}\|_{p,S^q} = O(1/2^{m\beta})$. In this case,

$$\begin{aligned}\left\|g_{m,\phi}\right\|^2_{H^\wedge_{K(\phi)}} &= \sum_{\omega,\omega'\in\wedge}\lambda_\omega\lambda_{\omega'}\mathfrak{D}_\phi\sigma_{2^m}(\wedge,\eta,f,\omega)\mathfrak{D}_\phi\sigma_{2^m}(\wedge,\eta,f,\omega')K(\phi,\omega\omega') \\ &\leq \left(\sum_{\omega\in\wedge}\lambda_\omega\left|\mathfrak{D}_\phi\sigma_{2^m}(\wedge,\eta,f,\omega)\right|^2\right)\left(\sum_{\omega,\omega'\in\wedge}\lambda_\omega\lambda_{\omega'}\left|K(\phi,\omega\omega')\right|^2\right)^{1/2}.\end{aligned} \tag{4.23}$$

Since $\sigma_{2^m}(\wedge,\eta,x)$ is a spherical harmonics of order $\le 2^m$, we know by (i) of Lemma 3.5 that $\mathfrak{D}_\phi\sigma_{2^m}(\wedge,\eta,f,x)$ are also spherical harmonics of order $\le 2^m$. Then, (3.2), (i) of Lemma 3.5, (3.3), and (3.16) make

$$\begin{aligned}\sum_{\omega\in\wedge}\lambda_\omega\left|\mathfrak{D}_\phi\sigma_{2^m}(\wedge,\eta,f,\omega)\right|^2 &\le C\int_{S^q}\left|\mathfrak{D}_\phi\sigma_{2^m}(\wedge,\eta,f,\omega)\right|^2 d\mu_q(\omega)\\ &\le C\left\|\partial_\beta\sigma_{2^m}(\wedge,\eta,f)\right\|_{2,S^q}^2\\ &\le C2^{2mq(1/p-1/2)_+}\left\|\partial_\beta\sigma_{2^m}(\wedge,\eta,f)\right\|_{p,S^q}^2\\ &\le C2^{2mq(1/p-1/2)_+}\left(\left\|\partial_\beta f-\partial_\beta\sigma_{2^m}(\wedge,\eta,f)\right\|_{p,S^q}^2+C\left\|\partial_\beta f\right\|_{p,S^q}^2\right).\end{aligned}\tag{4.24}$$

Hence, (3.19) and above equation make $\|g_{m,\phi}\|^2_{H^\wedge_{K(\phi)}} \le C2^{mq(1/p-1/2)_+}\|\partial_\beta f\|^2_{p,S^q}$. Equation (2.14) follows by (3.15). □

## Acknowledgments

This work is supported by the National NSF (10871226) of China. The authors thank the reviewers for giving very valuable suggestions.

## References

[1] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society*, vol. 39, no. 1, pp. 1–49, 2002.

[2] F. Cucker and D.-X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, Mass, USA, 2007.

[3] V. N. Vapnik, *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, New York, NY, USA, 1998.

[4] D. R. Chen, Q. Wu, Y. M. Ying, and D. X. Zhou, "Support vector machine soft margin classifier: error analysis," *Journal of Machine Learning and Research*, vol. 5, pp. 1143–1175, 2004.

[5] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Advances in Computational Mathematics*, vol. 13, no. 1, pp. 1–50, 2000.

[6] Y. Li, Y. Liu, and J. Zhu, "Quantile regression in reproducing kernel Hilbert spaces," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 255–268, 2007.

[7] H. Tong, D.-R. Chen, and L. Peng, "Analysis of support vector machines regression," *Foundations of Computational Mathematics*, vol. 9, no. 2, pp. 243–257, 2009.

[8] H. Tong, D.-R. Chen, and L. Peng, "Learning rates for regularized classifiers using multivariate polynomial kernels," *Journal of Complexity*, vol. 24, no. 5-6, pp. 619–631, 2008.

[9] D.-X. Zhou and K. Jetter, "Approximation with polynomial kernels and SVM classifiers," *Advances in Computational Mathematics*, vol. 25, no. 1–3, pp. 323–344, 2006.

[10] D. Chen and D.-H. Xiang, "The consistency of multicategory support vector machines," *Advances in Computational Mathematics*, vol. 24, no. 1–4, pp. 155–169, 2006.

[11] E. De Vito, L. Rosasco, A. Caponnetto, M. Piana, and A. Verri, "Some properties of regularized kernel methods," *Journal of Machine Learning Research*, vol. 5, pp. 1363–1390, 2004.

[12] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.

[13] C. A. Micchelli, Y. Xu, and H. Zhang, "Universal kernels," *Journal of Machine Learning Research*, vol. 7, pp. 2651–2667, 2006.

[14] Q. Wu, Y. Ying, and D.-X. Zhou, “Multi-kernel regularized classifiers,” *Journal of Complexity*, vol. 23, no. 1, pp. 108–134, 2007.
[15] J. Bergh and J. Löfström, *Interpolation Spaces*, Springer, New York, NY, USA, 1976.
[16] H. Berens and G. G. Lorentz, “Inverse theorems for Bernstein polynomials,” *Indiana University Mathematics Journal*, vol. 21, no. 8, pp. 693–708, 1972.
[17] H. Berens and L. Q. Li, “The Peetre $K$-moduli and best approximation on the sphere,” *Acta Mathematica Sinica*, vol. 38, no. 5, pp. 589–599, 1995.
[18] H. Berens and Y. Xu, “$K$-moduli, moduli of smoothness, and Bernstein polynomials on a simplex,” *Indagationes Mathematicae*, vol. 2, no. 4, pp. 411–421, 1991.
[19] W. Chen and Z. Ditzian, “Best approximation and $K$-functionals,” *Acta Mathematica Hungarica*, vol. 75, no. 3, pp. 165–208, 1997.
[20] W. Chen and Z. Ditzian, “Best polynomial and Durrmeyer approximation in $L_p(S)$,” *Indagationes Mathematicae*, vol. 2, no. 4, pp. 437–452, 1991.
[21] F. Dai and Z. Ditzian, “Jackson inequality for Banach spaces on the sphere,” *Acta Mathematica Hungarica*, vol. 118, no. 1-2, pp. 171–195, 2008.
[22] Z. Ditzian and X. Zhou, “Optimal approximation class for multivariate Bernstein operators,” *Pacific Journal of Mathematics*, vol. 158, no. 1, pp. 93–120, 1993.
[23] Z. Ditzian and K. Runovskii, “Averages and $K$-functionals related to the Laplacian,” *Journal of Approximation Theory*, vol. 97, no. 1, pp. 113–139, 1999.
[24] Z. Ditzian, “A measure of smoothness related to the Laplacian,” *Transactions of the American Mathematical Society*, vol. 326, no. 1, pp. 407–422, 1991.
[25] Z. Ditzian and V. Totik, *Moduli of Smoothness*, vol. 9 of *Springer Series in Computational Mathematics*, Springer, New York, NY, USA, 1987.
[26] Z. Ditzian, “Approximation on Banach spaces of functions on the sphere,” *Journal of Approximation Theory*, vol. 140, no. 1, pp. 31–45, 2006.
[27] Z. Ditzian, “Fractional derivatives and best approximation,” *Acta Mathematica Hungarica*, vol. 81, no. 4, pp. 323–348, 1998.
[28] L. L. Schumaker, *Spline Functions: Basic Theory*, John Wiley & Sons, New York, NY, USA, 1981, Pure and Applied Mathematics.
[29] K. Y. Wang and L. Q. Li, *Harmonic Analysis and Approximation on the Unit Sphere*, Science Press, Beijing, China, 2000.
[30] Y. Xu, “Approximation by means of $h$-harmonic polynomials on the unit sphere,” *Advances in Computational Mathematics*, vol. 21, no. 1-2, pp. 37–58, 2004.
[31] S. Smale and D.-X. Zhou, “Estimating the approximation error in learning theory,” *Analysis and Applications*, vol. 1, no. 1, pp. 17–41, 2003.
[32] B. Sheng, “Estimates of the norm of the Mercer kernel matrices with discrete orthogonal transforms,” *Acta Mathematica Hungarica*, vol. 122, no. 4, pp. 339–355, 2009.
[33] S. Loustau, “Aggregation of SVM classifiers using Sobolev spaces,” *Journal of Machine Learning Research*, vol. 9, pp. 1559–1582, 2008.
[34] H. N. Mhaskar and C. A. Micchelli, “Degree of approximation by neural and translation networks with a single hidden layer,” *Advances in Applied Mathematics*, vol. 16, no. 2, pp. 151–183, 1995.
[35] B. H. Sheng, “Approximation of periodic functions by spherical translation networks,” *Acta Mathematica Sinica. Chinese Series*, vol. 50, no. 1, pp. 55–62, 2007.
[36] B. Sheng, “On the degree of approximation by spherical translations,” *Acta Mathematicae Applicatae Sinica. English Series*, vol. 22, no. 4, pp. 671–680, 2006.
[37] B. Sheng, J. Wang, and S. Zhou, “A way of constructing spherical zonal translation network operators with linear bounded operators,” *Taiwanese Journal of Mathematics*, vol. 12, no. 1, pp. 77–92, 2008.
[38] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, “Approximation properties of zonal function networks using scattered data on the sphere,” *Advances in Computational Mathematics*, vol. 11, no. 2-3, pp. 121–137, 1999.
[39] H. N. Mhaskar, “Weighted quadrature formulas and approximation by zonal function networks on the sphere,” *Journal of Complexity*, vol. 22, no. 3, pp. 348–370, 2006.
[40] H. Groemer, *Geometric Applications of Fourier Series and Spherical Harmonics*, vol. 61 of *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, Cambridge, Mass, USA, 1996.
[41] C. Müller, *Spherical Harmonics*, vol. 17 of *Lecture Notes in Mathematics*, Springer, Berlin, Germany, 1966.
[42] S. Z. Lu and K. Y. Wang, *Bochner-Riesz Means*, Beijing Normal University Press, Beijing, China, 1988.
[43] Y. Wang and F. Cao, “The direct and converse inequalities for jackson-type operators on spherical cap,” *Journal of Inequalities and Applications*, vol. 2009, Article ID 205298, 16 pages, 2009.

[44] H. Wendland, *Scattered Data Approximation*, vol. 17 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge, Mass, USA, 2005.

[45] H. N. Mhaskar, F. J. Narcowich, N. Sivakumar, and J. D. Ward, "Approximation with interpolatory constraints," *Proceedings of the American Mathematical Society*, vol. 130, no. 5, pp. 1355–1364, 2002.

[46] F. J. Narcowich and J. D. Ward, "Scattered data interpolation on spheres: error estimates and locally supported basis functions," *SIAM Journal on Mathematical Analysis*, vol. 33, no. 6, pp. 1393–1410, 2002.

[47] G. Brown and F. Dai, "Approximation of smooth functions on compact two-point homogeneous spaces," *Journal of Functional Analysis*, vol. 220, no. 2, pp. 401–423, 2005.

[48] G. Brown, D. Feng, and S. Y. Sheng, "Kolmogorov width of classes of smooth functions on the sphere $\mathbb{S}^{d-1}$," *Journal of Complexity*, vol. 18, no. 4, pp. 1001–1023, 2002.

[49] F. Dai, "Multivariate polynomial inequalities with respect to doubling weights and $A_\infty$ weights," *Journal of Functional Analysis*, vol. 235, no. 1, pp. 137–170, 2006.

[50] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, "Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature," *Mathematics of Computation*, vol. 70, no. 235, pp. 1113–1130, 2001.

[51] E. Belinsky, F. Dai, and Z. Ditzian, "Multivariate approximating averages," *Journal of Approximation Theory*, vol. 125, no. 1, pp. 85–105, 2003.

[52] A. I. Kamzolov, "Approximation of functions on the sphere $S^n$," *Serdica*, vol. 10, no. 1, pp. 3–10, 1984.

[53] F. Dai and Z. Ditzian, "Cesàro summability and Marchaud inequality," *Constructive Approximation*, vol. 25, no. 1, pp. 73–88, 2007.

[54] F. Dai, "Some equivalence theorems with $K$-functionals," *Journal of Approximation Theory*, vol. 121, no. 1, pp. 143–157, 2003.