Approximation Properties of Zonal Function Networks Using Scattered Data on the Sphere

H. N. Mhaskar*
Department of Mathematics, California State University
Los Angeles, California, 90032, U. S. A.

F. J. Narcowich* and J. D. Ward*
Department of Mathematics, Texas A&M University
College Station, Texas, 77843, U. S. A.

Abstract

A zonal function (ZF) network is a function of the form $\mathbf{x} \mapsto \sum_{k=1}^{n} c_k \phi(\mathbf{x} \cdot \mathbf{y}_k)$, where $\mathbf{x}$ and the $\mathbf{y}_k$'s are on the on the unit sphere in $q+1$ dimensional Euclidean space, and where the $\mathbf{y}_k$'s are scattered points. In this paper, we study the degree of approximation by ZF networks. In particular, we compare this degree of approximation with that obtained with the classical spherical harmonics. In many cases of interest, this is the best possible for a given amount of information regarding the target function. We also discuss the construction of ZF networks using scattered data. Our networks require no training in the traditional sense, and provide theoretically predictable rates of approximation.

*Research of the authors was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant numbers F49620-97-1-0211 and F49620-98-1-0204. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

1 Introduction

In many applications in geophysics and meterology, the data is collected over the surface of the earth by satellites or ground stations. One then seeks to find a functional model for the mechanism that generates this data. The data is typically noisy, and statistical techniques must be used to eliminate this noise. However, even with statistically pure data, it is a major problem to find a suitable model that makes no assumptions about the nature of the actual relationships within the data. The actual relations are likely to be too complicated to be understood precisely, or even totally unknown.

In recent years, neural networks have provided a versatile tool for modelling such phenomena. The classical methods from approximation theory also provide many important tools for such a modelling, but one advantage of the neural network model is that it allows a fast, parallel computation, which in addition, can be implemented in hardware. The *complexity problem* in this connection consists of determining the size of the network under minimal assumptions on the unknown target function. In the theoretical investigations, these assumptions are encoded by the statement that the target function belongs to some function space, such as the Sobolev space. A survey of some of the results in this direction can be found in [5], where the theory is described for *generalized translation networks,* a concept that encompasses both neural networks and radial basis function networks.

The need to assume that the target function belongs to some function space, but is unknown otherwise, imposes certain inherent limitations on how well one can approximate the function using *any* reasonable process depending on a fixed, finite number of parameters. In the case of many important function spaces, these limitations have been explored in the theory of n-widths in classical approximation theory (cf. [11], [1]). It is perhaps surprising that the classical polynomial or linear approximation methods provide a degree of approximation that already achieves the order of magnitude of these lower bounds in the case of many function spaces. Thus, the possibility of nonlinear approximation using neural networks *does not* automatically yield a better approximation; whether it does or does not do so depends upon the a priori assumptions on the target function. One object of this paper is to construct neural (and radial basis function) networks that can compete with the known optimal classes of approximants with respect to the degree of approximation that can be achieved.

In this paper, we study the approximation of a function on the (surface of the) unit sphere in a Euclidean space of dimension $q+1$, where $q \geq 1$ is an integer that will be fixed throughout the rest of this paper. We note that if φ is a univariate function, and $\mathbf{x}$, $\mathbf{t}$ are on the sphere, then

$$\varphi(\|\mathbf{x}-\mathbf{t}\|^2) = \varphi(2-2\mathbf{x}\cdot\mathbf{t}).$$

Therefore, restricted to the sphere, RBF networks (with centers on the sphere) and neural networks (with weights on the sphere) have the same form. Accordingly, we define a *zonal function network* (ZF network) to be a finite linear combination of functions of the form $\mathbf{x} \mapsto \phi(\mathbf{x}\cdot\mathbf{y})$. In the same vein as in [6], we will compare the degree of approximation by ZF networks with the degree of approximation provided by spherical harmonics. We will obtain very general estimates, valid for essentially arbitrary target functions, and all ϕ under certain minimal conditions. For certain natural function classes to which the target function may be assumed to belong, and ϕ satisfying additional conditions, our results will be close to optimal. We will give explicit constructions of linear operators which accomplish the approximation.

Although our results are optimal for the approximation of functions from the same classes for which approximation by spherical harmonics is optimal, there are many reasons to work with ZF networks rather than spherical harmonics. In particular, computation of spherical harmonics is expensive, matrices whose entries consist of evaluations of spherical harmonics are known to be highly unstable, and finally, Gibbs-like phenomena in polynomial approximations provide ample reason to consider these alternative approximation schemes.

In the next section, we summarize certain central facts about polynomials on the sphere, including the Marcinkiewicz-Zygmund inequalities and quadrature formulas proved in [7]. In Section 3, we review the approximation properties of polynomials on the sphere. We also develop linear methods involving discretized delayed means operators to construct spherical harmonics, based on the values of the target function at scattered data, that give a near-best order of approximation to the target function. In Section 4, we will use the operators constructed in Section 4 to describe the approximation properties of ZF networks. Section 5 provides two examples illustrating the salient aspects of the theory.

2 Polynomials on the sphere

2.1 Spherical Harmonics

Let $q \geq 1$ be an integer which will be fixed throughout the rest of this paper, and let $\mathbb{S}^q$ be the (surface of the) unit sphere in the Euclidean space $\mathbb{R}^{q+1}$, with $d\mu_q$ being its usual volume element. We note that the volume element is invariant under arbitrary coordinate changes. The volume of $\mathbb{S}^q$ is

$$\omega_q := \int_{\mathbb{S}^q} d\mu_q = \frac{2\pi^{(q+1)/2}}{\Gamma((q+1)/2)}. \tag{2.1}$$

Corresponding to $d\mu_q$, we have the inner product and $L^p(\mathbb{S}^q)$ norms,

$$\langle f, g\rangle_{\mathbb{S}^q} := \int_{\mathbb{S}^q} f(\mathbf{x})\overline{g(\mathbf{x})}d\mu_q(\mathbf{x}) \tag{2.2}$$

$$\|f\|_{\mathbb{S}^q,p} := \begin{cases} \left\{\displaystyle\int_{\mathbb{S}^q} |f(\mathbf{x})|^p d\mu_q(\mathbf{x})\right\}^{1/p} & \text{if } 1 \leq p < \infty \\ \operatorname*{ess\,sup}_{\mathbf{t}\in\mathbb{S}} |f(\mathbf{t})| & \text{if } p = \infty \end{cases} \tag{2.3}$$

The class of all measurable functions $f : \mathbb{S}^q \to \mathbb{C}$ for which $\|f\|_{\mathbb{S}^q,p} < \infty$ will be denoted by $L^p(\mathbb{S}^q)$, with the usual understanding that functions that are equal almost everywhere are considered equal as elements of $L^p(\mathbb{S}^q)$. All continuous complex valued functions on $\mathbb{S}^q$ will be denoted by $C(\mathbb{S}^q)$.

For integer $\ell \geq 0$, the restriction to $\mathbb{S}^q$ of a homogeneous harmonic polynomial of degree ℓ is called a spherical harmonic of degree ℓ. Most of the following information is based on [9] and [13, §IV.2], although we use a different notation. The class of all spherical harmonics of degree ℓ will be denoted by $\mathbf{H}^q_\ell$, and the class of all spherical harmonics of degree $\ell \leq n$ will be denoted by Π^q_n. Of course, $\Pi^q_n = \bigoplus_{\ell=0}^n \mathbf{H}^q_\ell$, and it comprises the restriction to $\mathbb{S}^q$ of all algebraic polynomials in $q+1$ variables of total degree not exceeding n. The dimension of $\mathbf{H}^q_\ell$ is given by

$$d^{\,q}_\ell := \dim \mathbf{H}^q_\ell = \begin{cases} \dfrac{2\ell+q-1}{\ell+q-1}\dbinom{\ell+q-1}{\ell} & \text{if } \ell \geq 1, \\ 1 & \text{if } \ell = 0. \end{cases} \tag{2.4}$$

and that of Π^q_n is $\sum_{\ell=0}^n d^{\,q}_\ell$.

The spherical harmonics have an intrinsic characterization. To describe this, we first define the Laplace-Beltrami operator. If we introduce local coordinates $\{\theta_1, \ldots, \theta_q\}$ for a coordinate patch on $\mathbb{S}^q$, then the corresponding patch embedded in $\mathbb{R}^{q+1}$ will have the parametrization $x_i = f_i(\theta_1, \ldots, \theta_q)$, $i = 1, \ldots, q+1$. The metric $g_{j,k}$ on $\mathbb{S}^q$ is then induced via restricting $ds^2 = \sum_{i=1}^{q+1} dx_i^2$ to $\mathbb{S}^q$; that is,

$$\sum_{j,k=1}^{q} g_{j,k} d\theta_j d\theta_k \equiv \sum_{i=1}^{q+1} \Big(\sum_{k=1}^{q} \frac{\partial f_i}{\partial \theta_k} d\theta_k \Big)^2 .$$

Note that as a matrix (fix the θ_k's), $g_{j,k}$ is symmetric and positive definite. We follow standard conventions in letting $g^{j,k}$ be the matrix inverse of $g_{j,k}$ and $g = \det g_{j,k}$. The Laplace-Beltrami operator on $\mathbb{S}^q$ is given by

$$\Delta_{\mathbb{S}^q} := g^{-\frac{1}{2}} \sum_{j=1}^{m} \sum_{k=1}^{m} \frac{\partial}{\partial \theta_j} \left(g^{\frac{1}{2}} g^{ij} \frac{\partial}{\partial \theta_k} \right) . \tag{2.5}$$

The operator $\Delta_{\mathbb{S}^q}$ is an elliptic, (unbounded) selfadjoint operator on $L^2(\mathbb{S}^q)$, is invariant under arbitrary coordinate changes, and its spectrum comprises distinct eigenvalues $\lambda_\ell := \ell(\ell + q - 1)\, \ell = 0, 1, \ldots$, each having finite multiplicity, d_ℓ^q. The space $\mathbf{H}_\ell^q$ can be characterized intrinsically as the eigenspace corresponding to λ_ℓ.

Since the λ_ℓ's are distinct, and the operator is selfadjoint, the spaces $\mathbf{H}_\ell^q$'s are mutually orthogonal relative to (2.2); also, $L^2(\mathbb{S}^q) = \text{closure}\{ \bigoplus_\ell \mathbf{H}_\ell^q \}$. Hence, if we choose an orthonormal basis $\{Y_{\ell,k} : k = 1, \cdots, d_\ell^q\}$ for each $\mathbf{H}_\ell^q$, then the set $\{Y_{\ell,k} : \ell = 0, 1, \ldots \text{ and } k = 1, \cdots, d_\ell^q\}$ is an orthonormal basis for $L^2(\mathbb{S}^q)$. One has the well-known addition formula [9]:

$$\sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\mathbf{x}) \overline{Y_{\ell,k}(\mathbf{y})} = \frac{d_\ell^q}{\omega_q} \mathcal{P}_\ell(q+1; \mathbf{x} \cdot \mathbf{y}), \qquad \ell = 0, 1, \cdots, \tag{2.6}$$

where $\mathcal{P}_\ell(q+1; x)$ is the degree-ℓ Legendre polymonial in $q+1$-dimensions. (We note that Müller's ω_q and $N(q, \ell)$ are the same as our ω_{q+1} and d_ℓ^{q+1}.)

The Legendre polynomials are normalized so that $\mathcal{P}_\ell(q+1; 1) = 1$, and satisfy the orthogonality relations [9, Lemma 10]

$$\int_{-1}^{1} \mathcal{P}_\ell(q+1; x) \mathcal{P}_k(q+1; x) (1 - x^2)^{\frac{q}{2} - 1} dx = \frac{\omega_q}{\omega_{q-1} d_\ell^q} \delta_{\ell,k}. \tag{2.7}$$

They are related to the ultraspherical (Gegenbauer) polynomials $P_\ell^{(\frac{q-1}{2})}$ (cf. [14], [9, p. 33]), and the Jacobi polynomials, $P_\ell^{(\alpha,\beta)}$, with $\alpha = \beta = \frac{q}{2} - 1$, via

$$P_\ell^{(\frac{q-1}{2})}(x) = \binom{\ell+q-2}{\ell}\mathcal{P}_\ell(q+1;x) \quad (q \geq 2) \tag{2.8}$$

$$P_\ell^{(\frac{q}{2}-1,\frac{q}{2}-1)}(x) = \binom{\ell+\frac{q}{2}-1}{\ell}\mathcal{P}_\ell(q+1;x). \tag{2.9}$$

When $q = 1$, the Legendre polynomials $\mathcal{P}_\ell(2;x)$ coincide with the Chebyshev polynomials $T_\ell(x)$; the ultraspherical polynomials $P_\ell^{(0)}(x) = (2/\ell)T_\ell(x)$, if $\ell \geq 1$. For $\ell = 0$, $P_0^{(0)}(x) = 1$. From the fact that $\Pi_n^q = \bigoplus_{\ell=0}^n \mathbf{H}_\ell^q$, and the addition formula (2.6), we obtain that for any $P \in \Pi_n^q$ and $\mathbf{x} \in \mathbb{S}^q$,

$$P(\mathbf{x}) = \sum_{\ell=0}^{n} \frac{d_\ell^q}{\omega_q} \int_{\mathbb{S}^q} P(\mathbf{y})\mathcal{P}_\ell(q+1;\mathbf{x}\cdot\mathbf{y})d\mu_q(\mathbf{y}). \tag{2.10}$$

In addition to the inner product and norms defined above on $\mathbb{S}^q$, we will need the following related inner product and norms for $[-1,1]$, with weight function $w_q(x) := (1-x^2)^{\frac{q}{2}-1}$:

$$\langle f,g\rangle_{w_q} := \int_{-1}^{1} f(x)\overline{g(x)}w_q(x)dx, \ w_q(x) := (1-x^2)^{\frac{q}{2}-1} \tag{2.11}$$

$$\|f\|_{w_q,p} := \begin{cases} \left\{\int_{-1}^{1} |f(x)|^p w_q(x)dx\right\}^{1/p}, & \text{if } 1 \leq p < \infty, \\ \operatorname*{ess\,sup}_{x\in[-1,1]} |f(x)|, & \text{if } p = \infty. \end{cases} \tag{2.12}$$

Finally, we note that the Funk-Hecke formula [9, Theorem 6] implies the following useful connection between integrals over $\mathbb{S}^q$ and integrals over $[-1,1]$ with respect to the weight function w_q. For any $\phi \in L^1_{w_q}[-1,1]$, $\mathbf{y} \in \mathbb{S}^q$, and any $Y_\ell \in \mathbf{H}_\ell^q$, we have

$$\int_{\mathbb{S}^q} \phi(\mathbf{x}\cdot\mathbf{z})Y_\ell(\mathbf{z})d\mu_q(\mathbf{z}) = \omega_{q-1}Y_\ell(\mathbf{x})\int_{-1}^{1}\phi(t)\mathcal{P}_\ell(q+1;t)w_q(t)dx \tag{2.13}$$

$$=: \frac{\omega_q}{d_\ell^q}\hat{\phi}(\ell)Y_\ell(\mathbf{x}). \tag{2.14}$$

We remark that our definition of $\hat{\phi}$ in (2.14) is chosen so that the Legendre expansion of ϕ has the form $\sum \hat{\phi}(\ell)\mathcal{P}_\ell(q+1;\cdot)$.

2.2 Inequalities

In this subsection, we review certain facts about polynomials which will be utilized often in the sequel.

Let $\mathcal{C}$ be a finite set of distinct points on $\mathbb{S}^q$. The mesh norm of $\mathcal{C}$ is defined to be

$$\delta_{\mathcal{C}} := \sup_{x\in\mathbb{S}^q} \mathrm{dist}(x, \mathcal{C}). \tag{2.15}$$

The following theorem summarizes the Marcinkiewicz-Zygmund inequalities ((2.18) below) and the quadrature formula ((2.17), (2.16) below) given in [7]. In the sequel, we adopt the following convention regarding constants. The letters $c, c_1, \cdots$ will denote positive constants depending only the dimension q, and the different norms involved in the formula. Their value will be different at different ocurrences, even within the same formula. The symbol $A \sim B$ will mean $cA \le B \le c_1 A$.

Theorem 2.1 *There exist constants α_q and N_q with the following property. Let $1 \le p \le \infty$, $\mathcal{C}$ be a finite set of distinct points on $\mathbb{S}^q$, and n be an integer with $N_q \le n \le \alpha_q \delta_{\mathcal{C}}^{-1}$. Then there exist nonnegative weights $\{A_\xi\}_{\xi\in\mathcal{C}}$ and $\{a_\xi\}_{\xi\in\mathcal{C}}$, with*

$$\sum_{\xi\in\mathcal{C}} \frac{a_\xi}{A_\xi} \le c, \tag{2.16}$$

such that for every $P \in \Pi_n^q$,

$$\frac{1}{\omega_q}\int_{\mathbb{S}^q} P(\mathbf{x})d\mu_q(\mathbf{x}) = \sum_{\xi\in\mathcal{C}} a_\xi P(\xi), \tag{2.17}$$

and

$$\|P\|_{\mathcal{C},p} \sim \|P\|_{\mathbb{S}^q,p}, \tag{2.18}$$

where

$$\|P\|_{\mathcal{C},p} := \begin{cases} \left(\displaystyle\sum_{\xi\in\mathcal{C}} |P(\xi)|^p A_\xi\right)^{1/p} & \text{if } 1 \le p < \infty, \\ \displaystyle\sup_{\xi\in\mathcal{C}}\{|P(\xi)|\} & \text{if } p = \infty. \end{cases} \tag{2.19}$$

Further,

$$|\{\xi \ : \ a_\xi \ne 0\}| \sim n^q \sim \dim(\Pi_n^q). \tag{2.20}$$

In [7], we have defined a partition of $\mathbb{S}^q$, such that each element R of this partition contains at least one point of $\mathcal{C}$, and satisfies some other technical properties that do not concern us in this paper. As pointed out in [7], we may then choose a subset $\widetilde{\mathcal{C}}$ of $\mathcal{C}$, such that each element of the partition contains exactly one point of $\widetilde{\mathcal{C}}$. We have then defined A_ξ as the surface area of the element containing the point ξ if this point is in $\widetilde{\mathcal{C}}$. For points of $\mathcal{C}\setminus\widetilde{\mathcal{C}}$, A_ξ (and a_ξ) are chosen to be equal to 0, with the convention that the corresponding quantitiy a_ξ/A_ξ in (2.16) is also equal to zero. In [7], we have discussed algorithms to compute a_ξ.

In addition to the quadrature formula, we will need these Nikolskii inequalities.

Proposition 2.1 *Let $1\le p<r\le\infty$, $n\ge 1$ be an integer, and $P\in\Pi_n^q$. Then*

$$\|P\|_{\mathbb{S}^q,p}\le \omega_q^{1/p-1/r}\|P\|_{\mathbb{S}^q,r}\le cn^{q/p-q/r}\|P\|_{\mathbb{S}^q,p}, \tag{2.21}$$

where the constant c depends only on q.

PROOF. The first estimate in (2.21) follows from Hölder inequality. To prove the second estimate, let $1\le s<\infty$, $|P(\mathbf{x})|=\|P\|_{\mathbb{S}^q,\infty}$, and $0<\eta<1$. Let $C_{\eta,n}$ be the spherical cap $\{\mathbf{y}\in\mathbb{S}^q\ :\ d(\mathbf{y},\mathbf{x})\le\eta/n\}$. We note that $\mu_q(C_{\eta,n})\sim(\eta/n)^q$. In view of a Bernstein inequality (cf. [3], eqn. (7)), we have for $\mathbf{y}\in\mathcal{C}_{\eta,n}$,

$$|P(\mathbf{x})-P(\mathbf{y})|\le nd(\mathbf{y},\mathbf{x})\|P\|_{\mathbb{S}^q,\infty}\le\eta|P(\mathbf{x})|.$$

Hence, $|P(\mathbf{y})|\ge(1-\eta)|P(\mathbf{x})|=(1-\eta)\|P\|_{\mathbb{S}^q,\infty}$ for $\mathbf{y}\in\mathcal{C}_{\eta,n}$, and

$$\int_{\mathbb{S}^q}|P(\mathbf{y})|^s d\mu_q(\mathbf{y})\ge\int_{C_{\eta,n}}|P(\mathbf{y})|^s d\mu_q(\mathbf{y})\ge(1-\eta)^s\mu_q(C_{\eta,n})\|P\|_{\mathbb{S}^q,\infty}^s.$$

Hence, for $1\le s<\infty$,

$$\|P\|_{\mathbb{S}^q,\infty}\le(1-\eta)^{-1}\mu_q(C_{\eta,n})^{-1/s}\|P\|_{\mathbb{S}^q,s}. \tag{2.22}$$

Taking $s=p$, this proves the second estimate in (2.21) in the case $r=\infty$. Next, if $p<r<\infty$, the estimate (2.22) (with $s=r$) yields

$$\begin{aligned}\int_{\mathbb{S}^q}|P(\mathbf{y})|^r d\mu_q(\mathbf{y}) &= \int_{\mathbb{S}^q}|P(\mathbf{y})|^{r-p}|P(\mathbf{y})|^p d\mu_q(\mathbf{y})\\ &\le \left(\frac{\|P\|_{\mathbb{S}^q,r}}{(1-\eta)\mu_q(C_{\eta,n})^{1/r}}\right)^{r-p}\int_{\mathbb{S}^q}|P(\mathbf{y})|^p d\mu_q(\mathbf{y}).\end{aligned}$$

This leads to the second estimate in (2.21) also when $r < \infty$. $\square$

3 Polynomial approximation on the sphere

3.1 Degree of approximation

In this section, we review (and comment upon) some results of Pawelke [10] regarding the approximation properties of functions by spherical harmonics. For $f \in L^p(\mathbb{S}^q)$ and integer $n \ge 0$, we write

$$E_{\mathbb{S}^q,n,p}(f) := \inf_{P \in \Pi_n^q} \|f - P\|_{\mathbb{S}^q,p}. \tag{3.1}$$

If $r \ge 1$ is an integer, the class of functions f for which $\Delta_{\mathbb{S}^q}^r f \in L^p(\mathbb{S}^q)$ is denoted by $W_r^p(\mathbb{S}^q)$. Pawelke has proved that for $1 \le p \le \infty$, integer $n \ge 1$, and $f \in W_r^p(\mathbb{S}^q)$,

$$E_{\mathbb{S}^q,n,p}(f) \le cn^{-2r}\|\Delta_{\mathbb{S}^q}^r f\|_{\mathbb{S}^q,p}, \tag{3.2}$$

where c is a positive constant depending only on r and q. We observe that the number m of complex parameters involved in expressing a polynomial in Π_n^q is $\mathcal{O}(n^q)$. Hence, the degree of approximation is $\mathcal{O}(m^{-(2r)/q})$ in terms of the number of parameters in the approximating class.

To examine how good this degree of approximation is, we recall the notion of nonlinear m-widths, introduced in [1]. Let X be a Banach space, and K be a compact subset of X. Any approximation process for elements of K depending on m parameters may be thought of as a composition of two mappings, the function $\pi : K \to \mathbb{R}^m$ selects the parameters of the process and the function $\mathcal{M} : \mathbb{R}^m \to X$ reconstructs the approximation with these parameters. The approximation to an element f is then $\mathcal{M}(\pi(f))$. The nonlinear m-width of K is defined by

$$\delta_m(K) := \inf_{\mathcal{M},\pi} \sup_{f \in K} \|f - \mathcal{M}(\pi(f))\|, \tag{3.3}$$

where $\|\cdot\|$ is the norm of X, and in taking the infimum, the function π is assumed to be continuous. Another associated, but somewhat technical, concept is that of the Bernstein m-width which is defined as follows. Let

$\mathcal{X}_{m+1}$ be the set of all linear subspaces of X having dimension not exceeding $m+1$. The Bernstein m-width is given by the expression

$$\beta_m(K) := \sup_{Y \in \mathcal{X}_{m+1}} \{\rho \ : \ \rho f/\|f\| \in K \text{ for all } f \in Y\}. \tag{3.4}$$

According to a result of DeVore, Howard and Micchelli ([1]),

$$\delta_m(K) \geq \beta_m(K), \qquad m = 1, 2, \cdots. \tag{3.5}$$

Coming back to the topic of spherical harmonics, let $K := K_{q;r,p}$ be the class of all $f \in W_r^p(\mathbb{S}^q)$ for which $\|\Delta_{\mathbb{S}^q}^r f\|_{\mathbb{S}^q,p} \leq 1$. The estimate (3.2) shows that $\delta_m(K) \leq cm^{-2r/q}$. To see that the reverse inequality holds, we point out that from [10] we have

$$\|\Delta_{\mathbb{S}^q}^r P\|_{\mathbb{S}^q,p} \leq cn^{2r}\|P\|_{\mathbb{S}^q,p}, \qquad P \in \Pi_n^q, \ n = 1, 2, \cdots. \tag{3.6}$$

Considering that the dimension of Π_n^q is $\mathcal{O}(n^q)$, we obtain from (3.6), (3.4), (3.5) that $\delta_m(K) \geq cm^{-2r/q}$. In particular, we conclude that the class Π_n^q is the "best class of approximants" for K.

Finally, we remark that the estimate (3.6) implies "converse theorems" for polynomial approximation on the sphere in terms of properly defined K-functionals (cf. [2]).

3.2 Delayed mean operators

In classical trigonometric approximation theory, the de la Vallée Poussin operators play an important role. These are uniformly bounded operators v_n, such that for each continuous 2π-periodic function f, $v_n(f)$ is a trigonometric polynomial of order at most $2n-1$, and $v_n(T) = T$ for each trigonometric polynomial of order at most n. Consequently, these are near-best approximating trigonometric polynomials (cf. [4]). These operators played a central role also in the proofs and constructions in [6]. In [12], Stein has developed similar operators for the ultraspherical polynomial expansions, using higher order Cesàro means of these expansions. We summarize these results in the following theorem; the details are explained in [7].

Theorem 3.1 *Let $q \geq 1$ be an integer, and k_q be the smallest integer greater than $(q-1)/2$. There exists a sequence of univariate polynomials $\{\tau_n^q\}$, and operators $\{T_n^q\}$ defined by*

$$T_n^q(f, \mathbf{x}) := \int_{\mathbb{S}^q} f(\mathbf{y})\tau_n^q(\mathbf{x}\cdot\mathbf{y})d\mu_q(\mathbf{y}), \qquad f \in L^1(\mathbb{S}^q), \ \mathbf{x} \in \mathbb{S}^q, \tag{3.7}$$

with the following properties.
(a) *For each integer $n \ge 1$, the degree of τ_n^q is at most $(k_q+1)n$.*
(b) *We have*

$$\sup_{n\ge 1} \|\tau_n^q\|_{w_q,1} \le c. \tag{3.8}$$

(c) *For each $P \in \Pi_n^q$, $T_n^q(P) = P$.*
(d) *For $1 \le p \le \infty$ and $f \in L^p(\mathbb{S}^q)$, $T_n^q(f) \in \Pi^q_{(k_q+1)n}$, and*

$$\sup_{n\ge 1} \|T_n^q(f)\|_{\mathbb{S}^q,p} \le c\|f\|_{\mathbb{S}^q,p}. \tag{3.9}$$

Consequently, for each integer $n \ge 1$,

$$E_{\mathbb{S}^q,(k_q+1)n,p}(f) \le \|f - T_n^q(f)\|_{\mathbb{S}^q,p} \le cE_{\mathbb{S}^q,n,p}(f). \tag{3.10}$$

We now define a discretized version of the operators T_n^q which can be computed using the samples of the target function at a scattered data set. Let $\mathcal{C}$ be a set of distinct points on $\mathbb{S}^q$, n be an integer with $N_q \le n \le (k_q+2)n \le \alpha_q\delta_{\mathcal{C}}^{-1}$, and the weights $\{a_\xi\}_{\xi\in\mathcal{C}}$ be constructed as in Theorem 2.1. In the sequel, we denote by $\nu_{\mathcal{C}}$ the measure that associates the mass a_ξ with the point $\xi \in \mathcal{C}$. $\nu_{\mathcal{C}}$ may be viewed either as a measure on $\mathbb{S}^q$ that restricts to $\mathcal{C}$ or as a measure on $\mathcal{C}$ that extends to $\mathbb{S}^q$; we will always write it as being over $\mathbb{S}^q$. The norms $\|\cdot\|_{\mathbb{S}^q,p,\nu_C}$ are defined for $f : \mathcal{C} \to \mathbb{R}$ by

$$\|f\|_{\mathbb{S}^q,p,\nu_C} := \begin{cases} \left(\displaystyle\sum_{\xi\in\mathcal{C}} |f(\xi)|^p a_\xi\right)^{1/p} & \text{if } 1 \le p < \infty, \\ \displaystyle\max_{\xi\in\mathcal{C}}\{|f(\xi)|\} & \text{if } p = \infty. \end{cases}$$

The discrete operators are defined for $f : \mathcal{C} \to \mathbb{R}$ by

$$T_n^{\mathcal{C}}(f,\mathbf{x}) := \int_{\mathbb{S}^q} f(\mathbf{y})\tau_n^q(\mathbf{x}\cdot\mathbf{y})d\nu_{\mathcal{C}}(\mathbf{y}), \qquad \mathbf{x} \in \mathbb{S}^q. \tag{3.11}$$

Our conditions on n imply that the quadrature formula (2.17), the estimates (2.16) and the Marcinkiewicz-Zygmund inequalities (2.18) hold for all polynomials in $\Pi^q_{(k_q+2)n}$. In particular, since $a_\xi \le cA_\xi$, $\xi \in \mathcal{C}$, the estimates (2.18) imply that for $1 \le p \le \infty$, and $P \in \Pi^q_{(k_q+2)n}$, we have

$$\|P\|_{\mathbb{S}^q,p,\nu_{\mathcal{C}}} \le c\|P\|_{\mathbb{S}^q,p}. \tag{3.12}$$

Therefore, Theorem 2(b) of [8] and Theorem 3.1 imply the following theorem regarding the discretized operators $T_n^{\mathcal{C}}$. For the convenience of the reader, we will give a slightly different proof.

Theorem 3.2 *Assume the notations and conditions just described. For each $P \in \Pi_n^q$, $T_n^{\mathcal{C}}(P) = P$. For each $f : \mathcal{C} \to \mathbb{R}$, $T_n^{\mathcal{C}}(f) \in \Pi_{(k_q+2)n}^q$, and for $1 \le p \le \infty$,*

$$\|T_n^{\mathcal{C}}(f)\|_{\mathbb{S}^q,p} \le c\|f\|_{\mathbb{S}^q,p,\nu_{\mathcal{C}}}\ , \tag{3.13}$$

where the constant c is independent of $\mathcal{C}$ and n. In particular, if $f \in C(\mathbb{S}^q)$ then

$$E_{\mathbb{S}^q,(k_q+2)n,\infty}(f) \le \|f - T_n^{\mathcal{C}}(f)\|_{\mathbb{S}^q,\infty} \le cE_{\mathbb{S}^q,n,\infty}(f). \tag{3.14}$$

Proof. In this proof only, let

$$M_n := \max_{\mathbf{x}\in\mathbb{S}^q} \int_{\mathbb{S}^q} |\tau_n^q(\mathbf{x}\cdot\mathbf{y})| d\mu_q(\mathbf{y}) = \max_{\mathbf{y}\in\mathbb{S}^q} \int_{\mathbb{S}^q} |\tau_n^q(\mathbf{x}\cdot\mathbf{y})| d\mu_q(\mathbf{x}).$$

The estimate (3.9) (with $p = \infty$) implies that $M_n \le c$ for all $n \ge 1$. We observe that for every $\mathbf{x} \in \mathbb{S}^q$, $\tau_n^q(\mathbf{x}\cdot\mathbf{y}) \in \Pi_{(k_q+2)n}^q$ as a function of $\mathbf{y}$. Hence, (3.12) implies that

$$\max_{\mathbf{x}\in\mathbb{S}^q} \int_{\mathbb{S}^q} |\tau_n^q(\mathbf{x}\cdot\mathbf{y})| d\nu_{\mathcal{C}}(\mathbf{y}) \le cM_n \le c.$$

Therefore,

$$\begin{aligned} \|T_n^{\mathcal{C}}(f)\|_{\mathbb{S}^q,\infty} &\le \max_{\mathbf{x}\in\mathbb{S}^q} \int_{\mathbb{S}^q} |\tau_n^q(\mathbf{x}\cdot\mathbf{y})||f(\mathbf{y})| d\nu_{\mathcal{C}}(\mathbf{y}) \\ &\le cM_n\|f\|_{\mathbb{S}^q,\nu_{\mathcal{C}},\infty} \le c\|f\|_{\mathbb{S}^q,\nu_{\mathcal{C}},\infty}. \end{aligned} \tag{3.15}$$

Next, using Fubini's theorem, we get

$$\begin{aligned} &\|T_n^{\mathcal{C}}(f)\|_{\mathbb{S}^q,1} \le \int_{\mathbb{S}^q}\int_{\mathbb{S}^q} |\tau_n^q(\mathbf{x}\cdot\mathbf{y})||f(\mathbf{y})| d\nu_{\mathcal{C}}(\mathbf{y}) d\mu_q(\mathbf{x}) \\ &\le \left\{ \max_{\mathbf{y}\in\mathbb{S}^q} \int_{\mathbb{S}^q} |\tau_n^q(\mathbf{x}\cdot\mathbf{y})| d\mu_q(\mathbf{x}) \right\} \|f\|_{\mathbb{S}^q,1,\nu_{\mathcal{C}}} = M_n\|f\|_{\mathbb{S}^q,1,\nu_{\mathcal{C}}} \\ &\le c\|f\|_{\mathbb{S}^q,1,\nu_{\mathcal{C}}}. \end{aligned} \tag{3.16}$$

The estimate (3.13) follows from (3.15), (3.16) and the Riesz-Thorin interpolation theorem. Since $T_n^{\mathcal{C}}(P) = P$ for every $P \in \Pi_n^q$, the estimate (3.14) follows easily from the case $p = \infty$ of (3.13). $\square$

4 Approximation by ZF networks

Following the ideas of [6], we first estimate the degree of approximation of a polynomial by ZF networks. The results in Section 3.2 will then enable us to obtain a close connection between the approximation of an "arbitrary" function by ZF networks and polynomials.

Throughout this section, we will assume that $\phi : [-1, 1] \to \mathbb{R}$ is a continuous function, which satisfies the additional condition

$$\hat{\phi}(\ell) \neq 0, \qquad \ell = 0, 1, 2, \cdots, \tag{4.1}$$

where $\hat{\phi}(\ell)$ is defined in (2.14). In view of the Funk-Hecke formula (2.13), the class of all ZF networks is not dense in $C(\mathbb{S}^q)$ without this condition; in fact, if $\hat{\phi}(\ell) = 0$ for some ℓ then this class is orthogonal to $\mathbf{H}_\ell^q$. For integers $n = 0, 1, \cdots$, we write

$$E_{n,p}(\phi) := \min \|\phi - P\|_{w_q,p}, \tag{4.2}$$

where the minimum is taken over all univariate polynomials P of degree not exceeding n.

Let $\mathcal{C}$ be a set of distinct points, and $\nu_{\mathcal{C}}$ the corresponding measure as in Section 3. For integer $N \geq 1$, we define the univariate polynomial of degree N by

$$\phi_N^+ := \sum_{\ell=0}^{N} \frac{(d_\ell^q)^2}{\omega_q^2\, \hat{\phi}(\ell)} \mathcal{P}_\ell(q+1; \cdot). \tag{4.3}$$

Next, we define the operators

$$\begin{aligned} \Phi f(\mathbf{x}) &:= \int_{\mathbb{S}^q} \phi(\mathbf{x} \cdot \mathbf{y}) f(\mathbf{y}) d\mu_q(\mathbf{y}), \qquad f \in L^1(\mathbb{S}^q),\ \mathbf{x} \in \mathbb{S}^q, \quad (4.4)\\ \Phi_N^+ f(\mathbf{x}) &:= \int_{\mathbb{S}^q} \phi_N^+(\mathbf{x} \cdot \mathbf{y}) f(\mathbf{y}) d\mu_q(\mathbf{y}), \qquad f \in L^1(\mathbb{S}^q),\ \mathbf{x} \in \mathbb{S}^q, \quad (4.5) \end{aligned}$$

and for $f : \mathcal{C} \to \mathbb{R}$,

$$\Phi^{\mathcal{C}} f(\mathbf{x}) := \int_{\mathbb{S}^q} \phi(\mathbf{x} \cdot \mathbf{y}) f(\mathbf{y}) d\nu_{\mathcal{C}}(\mathbf{y}), \qquad \mathbf{x} \in \mathbb{S}^q. \tag{4.6}$$

We observe that $\Phi^{\mathcal{C}}$ is a ZF network, obtained by a discretization of the operator Φ. Further, the relation (2.20) implies that the number of evaluations of ϕ involved in the computation of $\Phi^{\mathcal{C}} f$ is of the order of magnitude $\delta_{\mathcal{C}}^{-q}$.

In view of (4.3) and the Funk-Hecke formula (2.13) with $\mathbf{z} \mapsto \mathcal{P}_\ell(q+1;\mathbf{y}\cdot\mathbf{z})$ in place of Y_ℓ, we see that

$$\begin{aligned}
\int_{\mathbb{S}^q} \phi(\mathbf{x}\cdot\mathbf{z})\phi_N^+(\mathbf{y}\cdot\mathbf{z})d\mu_q(\mathbf{z}) &= \sum_{\ell=0}^N \frac{(d_\ell^q)^2}{\omega_q^2\,\hat{\phi}(\ell)} \int_{\mathbb{S}^q} \phi(\mathbf{x}\cdot\mathbf{z})\mathcal{P}_\ell(q+1;\mathbf{y}\cdot\mathbf{z})d\mu_q(\mathbf{z}) \\
&= \sum_{\ell=0}^N \frac{d_\ell^q}{\omega_q}\mathcal{P}_\ell(q+1;\mathbf{x}\cdot\mathbf{y}).
\end{aligned}$$

Hence, for any $P \in \Pi_N^q$, the formula (2.10) implies that

$$\Phi\Phi_N^+ P = P. \tag{4.7}$$

It is now reasonable to expect that if $N_q \le N \le \alpha_q\delta_{\mathcal{C}}^{-1}$, then for any polynomial $P \in \Pi_N^q$, $P = \Phi\Phi_N^+ P$ and $\Phi^{\mathcal{C}}\Phi_N^+ P$ should be close. The following theorem gives a quantitative description of this fact.

Theorem 4.1 *Let $1 \le p \le \infty$, $\mathcal{C}$ be a set of distinct points on $\mathbb{S}^q$, M, N be integers that satisfy $N \ge N_q$ and $M + N \le \alpha_q\delta_{\mathcal{C}}^{-1}$, $\phi \in C[-1,1]$ satisfy (4.1). We write*

$$\beta := \beta(p) := \max\left(0, \frac{1}{p} - \frac{1}{2}\right), \qquad m_N := \min_{0\le\ell\le N} \frac{|\hat{\phi}(\ell)|}{d_\ell^q}. \tag{4.8}$$

Then for any $P \in \Pi_N^q$, we have

$$\|P - \Phi^{\mathcal{C}}\Phi_N^+ P\|_{\mathbb{S}^q,p} \le c\frac{N^\beta}{m_N} E_{M,p}(\phi)\|P\|_{\mathbb{S}^q,p}. \tag{4.9}$$

In order to prove the theorem, we first prove a lemma.

Lemma 4.1 *Let $1 \le p \le \infty$, σ be a signed measure on $\mathbb{S}^q$, and for a univariate function f and a function g on $\mathbb{S}^q$,*

$$f *_\sigma g(\mathbf{x}) := \int_{\mathbb{S}^q} f(\mathbf{x}\cdot\mathbf{y})g(\mathbf{y})d\sigma(\mathbf{y}), \qquad \mathbf{x} \in \mathbb{S}^q,$$

whenever the integral is well defined. Then

$$\|f *_\sigma g\|_{\mathbb{S}^q,p} \le \omega_q^{1/p}\|f\|_{w_q,p} \int_{\mathbb{S}^q} |g(\mathbf{y})|d|\sigma|(\mathbf{y}). \tag{4.10}$$

PROOF. The estimate (4.10) is clear for the case $p = \infty$, and is easy to deduce in the case $p = 1$ using the rotation invariance of the volume measure (cf. (2.13) with $\ell = 0$). The general case follows from the Riesz-Thorin interpolation theorem. $\square$

PROOF. In this proof only, let $R := \Phi_N^+ P \in \Pi_N^q$. If Q is any polynomial of degree at most M and $\mathbf{x} \in \mathbb{S}^q$ then $Q(\mathbf{x}\cdot\xi)R(\xi)$ is in Π_{M+N}^q as a function of ξ. Consequenty, in view of (2.17),

$$\int_{\mathbb{S}^q} Q(\mathbf{x}\cdot\xi)R(\xi)d\mu_q(\xi) = \int_{\mathbb{S}^q} Q(\mathbf{x}\cdot\xi)R(\xi)d\nu_{\mathcal{C}}(\xi).$$

Hence,

$$\begin{aligned}\left(\Phi R - \Phi^{\mathcal{C}} R\right)(\mathbf{x}) &= \textstyle\int_{\mathbb{S}^q} \phi(\mathbf{x}\cdot\xi)R(\xi)(d\mu_q(\xi) - d\nu_{\mathcal{C}}(\xi))\\ &= \textstyle\int_{\mathbb{S}^q} (\phi(\mathbf{x}\cdot\xi) - Q(\mathbf{x}\cdot\xi))\, R(\xi)\,(d\mu_q(\xi) - d\nu_{\mathcal{C}}(\xi))\\ &\quad + \textstyle\int_{\mathbb{S}^q} Q(\mathbf{x}\cdot\xi)R(\xi)\,(d\mu_q(\xi) - d\nu_{\mathcal{C}}(\xi))\\ &= \textstyle\int_{\mathbb{S}^q} (\phi(\mathbf{x}\cdot\xi) - Q(\mathbf{x}\cdot\xi))\, R(\xi)\,(d\mu_q(\xi) - d\nu_{\mathcal{C}}(\xi))\,.\end{aligned}$$

Hence, using Lemma 4.1 with $\phi - Q$ in place of f, R in place of g and $\mu_q - \nu_{\mathcal{C}}$ in place of σ, we obtain

$$\begin{aligned}\|P - \Phi^{\mathcal{C}}\Phi_N^+ P\|_{\mathbb{S}^q,p} &= \|\Phi R - \Phi^{\mathcal{C}} R\|_{\mathbb{S}^q,p}\\ &\leq \|\phi - Q\|_{w_q,p}\left(\textstyle\int_{\mathbb{S}^q} |R(\xi)|d\mu_q(\xi) + \int_{\mathbb{S}^q} |R(\xi)|d\nu_{\mathcal{C}}(\xi)\right).\end{aligned}$$

Since this estimate is true for all polynomials Q of degree at most M, and our assumptions imply that the estimate (3.12) holds for R, we have proved that

$$\|P - \Phi^{\mathcal{C}}\Phi_N^+ P\|_{\mathbb{S}^q,p} \leq cE_{M,p}(\phi)\|\Phi_N^+ P\|_{\mathbb{S}^q,1}. \tag{4.11}$$

Next, we denote by P_ℓ the projection operator onto $\mathbf{H}_\ell^q$. Then, using the Nikolskii inequalities (2.21), we obtain

$$\begin{aligned}\|\Phi_N^+ P\|_{\mathbb{S}^q,1}^2 &\leq c\|\Phi_N^+ P\|_{\mathbb{S}^q,2}^2 = c\sum_{\ell=0}^{N}\left(\frac{d_\ell^q}{\omega_q\,\hat{\phi}(\ell)}\right)^2 \|\mathsf{P}_\ell P\|_{\mathbb{S}^q,2}^2\\ &\leq \left(\frac{c}{m_N}\right)^2 \sum_{\ell=0}^{N} \|\mathsf{P}_\ell P\|_{\mathbb{S}^q,2}^2\\ &= \left(\frac{c}{m_N}\right)^2 \|P\|_{\mathbb{S}^q,2}^2\\ &\leq \left(\frac{cN^\beta}{m_N}\right)^2 \|P\|_{\mathbb{S}^q,p}^2.\end{aligned}$$

Along with (4.11), this completes the proof. □

We are now in a position to state our approximation theorems for the case of "arbitrary" functions.

Theorem 4.2 *Let n, M be integers, $n \ge N_q$, and $1 \le p \le \infty$. Let $\mathcal{C}$ be a set of distinct points on $\mathbb{S}^q$ such that $(k_q+2)n+M \le \alpha_q\delta_{\mathcal{C}}^{-1}$. Let $N = (k_q+1)n-1$.*
(a) *If $f \in L^p(\mathbb{S}^q)$, then*

$$\|f - \Phi^{\mathcal{C}}\Phi_N^+ T_n^q f\|_{\mathbb{S}^q,p} \le c\Big(E_{\mathbb{S}^q,n,p}(f) + \frac{E_{M,\infty}(\phi)n^\beta}{m_N}\|f\|_{\mathbb{S}^q,p}\Big). \tag{4.12}$$

(b) *If $f \in C(\mathbb{S})$, then*

$$\|f - \Phi^{\mathcal{C}}\Phi_N^+ \mathsf{T}_n^{\mathcal{C}} f\|_{\mathbb{S}^q,\infty} \le c\Big(E_{\mathbb{S}^q,n,\infty}(f) + \frac{E_{M,\infty}(\phi)}{m_N}\|f\|_{\mathbb{S}^q,\infty}\Big). \tag{4.13}$$

PROOF. We obtain (4.12) from (3.10) and (4.9) with $T_n^q f$ in place of P. The estimate (4.13) follows similarly from (3.14) and (4.9) with $T_n^{\mathcal{C}} f$ in place of P. We observe that $\beta = 0$ in this case. □

Theorem 4.1 leads to the following "converse theorem", relating the degree of approximation by polynomials to that by our operators.

Theorem 4.3 *We continue the notations and conditions of Theorem 4.2. In addition, let $R > 0$, $0 < \gamma \le R$ and*

$$\frac{E_{M,\infty}(\phi)n^\beta}{m_N} \le cn^{-R}. \tag{4.14}$$

(a) *If $\|f - \Phi^{\mathcal{C}}\Phi_N^+ T_n^q f\|_{\mathbb{S}^q,p} \le cn^{-\gamma}$ then $E_{\mathbb{S}^q,n,p}(f) \le cn^{-\gamma}$.*
(b) *If $f \in C(\mathbb{S}^q)$, and $\|f - \Phi^{\mathcal{C}}\Phi_N^+ T_n^{\mathcal{C}} f\|_{\mathbb{S}^q,\infty} \le cn^{-\gamma}$ then $E_{\mathbb{S}^q,n,\infty}(f) \le cn^{-\gamma}$.*

PROOF. In view of (4.14), Theorem 4.1 with $T_n^q f$ in place of P implies that

$$\|T_n^q f - \Phi^{\mathcal{C}}\Phi_N^+ T_n^q f\|_{\mathbb{S}^q,p} \le cn^{-R}.$$

Consequently, our assumption in part (a) implies that $\|f - T_n^q f\|_{\mathbb{S}^q,p} \le cn^{-\gamma}$. The conclusion of part (a) now follows from (3.10). For the proof of part (b), we use Theorem 4.1 with $T_n^{\mathcal{C}} f$ in place of P, and (3.14) instead of (3.10). □

5 Examples

Let $r \geq 1$ be an integer. In this section, we are interested in the class $B_{q,r}$ of functions $f : \mathbb{S}^q \to \mathbb{R}$ for which $\|\Delta^r_{\mathbb{S}^q} f\|_{\mathbb{S}^q,\infty} \leq 1$. We recall from Section 3 that for $f \in B_{q,r}$, we have $E_{\mathbb{S}^q,n,\infty}(f) \leq cn^{-2r}$, and that this is the best order of approximation (up to constant factors) that is possible for the whole class $B_{q,r}$ by any process depending on $\mathcal{O}(n^q)$ parameters. We will assume that $|\mathcal{C}| = \mathcal{O}(n^q)$ and examine if the network in Theorem 4.2 yields this optimal order of approximation for the class $B_{q,r}$.

EXAMPLE 5.1. Let $0 < \rho < 1$ be a fixed number. If $q \geq 2$, set

$$\phi^G_\rho(x) := (1 - 2\rho x + \rho^2)^{-(q-1)/2}, \qquad x \in [-1, 1]. \tag{5.1}$$

From [9, Lemma 18], we have the expansion

$$\phi^G_\rho(x) = \sum_{\ell=0}^{\infty} \rho^\ell \binom{\ell + q - 2}{\ell} \mathcal{P}_\ell(q + 1; x).$$

Consequently, $\hat{\phi}^G_\rho(\ell) = \rho^\ell \binom{\ell+q-2}{\ell}$. From the expression for $d^{\,q}_\ell$ given in (2.4), we see that

$$\frac{\hat{\phi}^G_\rho(\ell)}{d^{\,q}_\ell} = \rho^\ell \frac{q - 1}{2\ell + q - 1} \tag{5.2}$$

and so, in this case,

$$m_N \sim \rho^N N^{-1}. \tag{5.3}$$

We observe that the function ϕ^G_ρ has an analytic extension to the ellipse $|z + \sqrt{z^2 - 1}| < \rho$, where the principal branch of the square root is chosen. Hence, a well known theorem of Bernstein [4] implies that for any $\rho_1 > \rho$,

$$E_{M,\infty}(\phi^G_\rho) \leq c\rho_1^M, \qquad M = 0, 1, 2, \cdots.$$

Choosing $\rho_1 = \sqrt{\rho}$ and $M = 4N$, and taking into account (5.3), we obtain

$$\frac{E_{M,\infty}(\phi^G_\rho)}{m_N} \leq cN\rho^N.$$

The estimate (4.13) now takes the form

$$\|f - \Phi^{\mathcal{C}}\Phi^+_N \mathsf{T}^{\mathcal{C}}_n f\|_{\mathbb{S}^q,\infty} \leq c\Big(E_{\mathbb{S}^q,n,\infty}(f) + N\rho^N \|f\|_{\mathbb{S}^q,\infty}\Big), \tag{5.4}$$

if $\delta_{\mathcal{C}}$ is sufficiently small. In particular, for all $f \in B_{q,r}$, and $\delta_{\mathcal{C}} \leq cn^{-1}$ for a sufficiently small constant c,

$$\|f - \Phi^{\mathcal{C}}\Phi_N^+\mathsf{T}_n^{\mathcal{C}}f\|_{\mathbb{S}^q,\infty} \leq c(\rho)n^{-2r}. \tag{5.5}$$

□

EXAMPLE 5.2. This example treats the important Gaussian function,

$$\exp(-\|\mathbf{x}-\mathbf{y}\|^2/(2\sigma^2)) = \exp(-\sigma^{-2})\exp(\mathbf{x}\cdot\mathbf{y}/\sigma^2), \qquad \mathbf{x},\mathbf{y} \in \mathbb{S}^q,$$

where $\sigma > 0$ is a fixed real number. Thus, with $\rho = \sigma^{-2}$, we consider the function

$$\phi_\rho^E(x) := \exp(\rho x), \qquad x \in [-1,1]. \tag{5.6}$$

To underline the fact that, in contrast to the theory of interpolation of functions, we do not require ϕ to be positive definite, we will allow in this section ρ to be any nonzero real number, positive or negative.

Lemma 5.1 *For any integer $\ell \geq 1$,*

$$\begin{aligned} \hat{\phi_\rho^E}(\ell) &= d_\ell^{\,q}\left(\frac{\rho}{2}\right)^{-\frac{q-1}{2}}\Gamma\left(\frac{q+1}{2}\right) I_{\ell+\frac{q-1}{2}}(\rho) &(5.7)\\ &= \frac{d_\ell^{\,q}\,\rho^\ell}{2^\ell\,\Gamma(\ell+\frac{q+1}{2})}\Big(1+\mathcal{O}(1/\ell)\Big), &(5.8)\end{aligned}$$

where $I_n(\cdot)$ is the order nth order modified Bessel function of the first kind.

PROOF. From [15, Chapter XVII, problem 37], we have for $\nu > 0$,

$$e^{\rho x} = 2^{\nu-1}\Gamma(\nu)\sum_{\ell=0}^\infty(\nu+\ell)\rho^{-\nu}I_{\ell+\nu}(\rho)P^{(\nu)}(x).$$

If we set $\nu = \frac{q-1}{2}$, with $q > 1$, then, from the expression for $d_\ell^{\,q}$ in (2.4) and the connection between the Legendre polynomials and the ultraspherical polynomials in (cf. [14], [9, p. 33]), we have

$$\phi_\rho^E(x) = e^{\rho x} = \Gamma(\frac{q+1}{2})\left(\frac{\rho}{2}\right)^{-\frac{q-1}{2}}\sum_{\ell=0}^\infty d_\ell^{\,q}\,I_{\ell+\frac{q-1}{2}}(\rho)\mathcal{P}_\ell(q+1;x)$$

This establishes (5.7) for $q \geq 2$. The $q = 1$ case can be obtained by direct computation or by using a limiting argument. Equation (5.8) follows from (5.7) and the standard series expansion for $I_{\ell+\frac{q-1}{2}}(\rho)$, with ρ fixed. $\square$

Thus, for ϕ_ρ^E, we have

$$m_N \sim \frac{|\rho|^N}{2^N \Gamma(N + \frac{q+1}{2})}.$$

It is straightforward to check, using the power series for $e^{\rho x}$ that

$$E_{M,\infty}(\phi_\rho^E) \leq c\frac{|\rho|^M}{M!}.$$

Let $t > 0$ be a fixed number, and take $M = \lfloor N(1+t) \rfloor$; clearly, $M! = \Gamma(M+1) \geq \Gamma\Big((1+t)N\Big)$. Using Stirling's approximation and some standard manipulations, we arrive at

$$\frac{E_{M,\infty}(\phi_\rho^E)}{m_N} \leq c(\rho) N^{-tN/2}.$$

Thus, estimate (4.13) now takes the form

$$\|f - \Phi^{\mathcal{C}} \Phi_N^+ \mathsf{T}_n^{\mathcal{C}} f\|_{\mathbb{S}^q,\infty} \leq c\Big(E_{\mathbb{S}^q,n,\infty}(f) + c(\rho) N^{-tN/2} \|f\|_{\mathbb{S}^q,\infty}\Big). \tag{5.9}$$

if $\delta_{\mathcal{C}}$ is sufficiently small. In particular, for all $f \in B_{q,r}$, and $\delta_{\mathcal{C}} \leq cn^{-1}$ for a sufficiently small constant c,

$$\|f - \Phi^{\mathcal{C}} \Phi_N^+ \mathsf{T}_n^{\mathcal{C}} f\|_{\mathbb{S}^q,\infty} \leq c(\rho) n^{-2r}. \tag{5.10}$$

$\square$

References

[1] R. DeVore, R. Howard and C. A. Micchelli, *Optimal nonlinear approximation,* Manuscripta Mathematica, **63** (1989), 469-478.

[2] R. DeVore and G. G. Lorentz, "Constructive Approximation", Springer Verlag, 1994.

[3] K. Jetter, J. Stöckler, and J. D. Ward, *Error estimates for scattered data interpolation,* Math. Comp., to appear.

[4] G. G. Lorentz, "Approximation of Functions", Holt, Rinehart and Winston, New York, 1966.

[5] H. N. Mhaskar, *Approximation of smooth functions by neural networks,* in "Dealing with complexity: A neural network approach", (K. Warwick et. al. eds), "Perspectives in Neural Computing", Springer Verlag, London, 1998, pp.189–204.

[6] H. N. Mhaskar and C. A. Micchelli, *Degree of approximation by neural and translation networks with a single hidden layer,* Advances in Applied Mathematics, **16** (1995), 151-183.

[7] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, *Quadrature Formulas on Spheres Using Scattered Data,* Center for Approximation Theory Report # 393, Department of Mathematics, Texas A & M University, 1998.

[8] H. N. Mhaskar and J. Prestin, *Marcinkiewicz-Zygmund-Type Inequalities,* To appear in "Approximation theory: in memory of A. K. Varma", (N. K. Govil, R. N. Mohapatra, Z. Nashed, A. Sharma, and J. Szabados Eds.), Marcel Dekker.

[9] C. Müller, "Spherical Harmonics", Lecture Notes in Mathematics, Vol. 17, Springer Verlag, Berlin, 1966.

[10] S. Pawelke, *Über die Approximationsordnung bei Kugelfunktionen und algebraischen Polynomen,* Tôhoku Math. Journ., **24** (1972), 473–486.

[11] A. Pinkus, "n-widths in approximation theory", Springer Verlag, New York, 1985.

[12] E. M. Stein, *Interpolation in polynomial classes and Markoff's inequality,* Duke Math. J., **24** (1957), 467–476.

[13] E. M. Stein and G. Weiss, "Fourier analysis on Euclidean spaces", Princeton University Press, Princeton, New Jersey, 1971.

[14] G. Szegö, "Orthogonal polynomials", Amer. Math. Soc. Colloq. Publ. **23**, Amer. Math. Soc., Providence, 1975.

[15] E. T. Whittaker and G. N. Watson, “A course of Modern Analysis”, Cambridge University Press, Cambridge, 1965.