Weighted quadrature formulas and approximation by zonal function networks on the sphere

H.N. Mhaskar

Department of Mathematics, California State University, Los Angeles, CA 90032, USA

Received 31 March 2005; accepted 5 October 2005
Available online 27 December 2005

Abstract

Let \( q \geq 1 \) be an integer, \( S^q \) be the unit sphere embedded in the Euclidean space \( \mathbb{R}^{q+1} \). A zonal function (ZF) network with an activation function \( \phi : [-1, 1] \to \mathbb{R} \) and \( n \) neurons is a function on \( S^q \) of the form \( x \mapsto \sum_{k=1}^{n} a_k \phi(x \cdot \zeta_k) \), where \( a_k \)'s are real numbers, \( \zeta_k \)'s are points on \( S^q \). We consider the activation functions \( \phi \) for which the coefficients \( \{ \hat{\phi}(\ell) \} \) in the appropriate ultraspherical polynomial expansion decay as a power of \( (\ell + 1)^{-1} \). We construct ZF networks to approximate functions in the Sobolev classes on the unit sphere embedded in a Euclidean space, yielding an optimal order of decay for the degree of approximation in terms of \( n \), compared with the nonlinear \( n \)-widths of these classes. Our networks do not require training in the traditional sense. Instead, the network approximating a function is given explicitly as the value of a linear operator at that function. In the case of uniform approximation, our construction utilizes values of the target function at scattered sites. The approximation bounds are used to obtain error bounds on a very general class of quadrature formulas that are exact for the integration of high degree polynomials with respect to a weighted integral. The bounds are better than those expected from a straightforward application of the Sobolev embeddings.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Approximation on the sphere; Quadrature formulas; Neural networks; Radial basis functions; Spherical basis functions

1. Introduction

In this paper, we wish to treat in a unified manner two apparently different problems. The first problem is to construct “zonal function (ZF) networks” on a Euclidean sphere which provide the best order of decay of the degree of approximation to functions in Sobolev classes on the sphere. The other problem is to estimate the error in approximating a weighted integral on the sphere.
using quadrature formulas exact for high degree polynomials. It turns out that in some cases, the two problems are two sides of the same coin. In Section 1.1, we describe the background for our paper from the perspective of approximation by ZF networks; in Section 1.2, we describe the background from the perspective of errors in quadrature formulas, and conclude the introduction in Section 1.3. In the remainder of this paper, \( q \geq 1 \) denotes a fixed integer, \( S^q \) denotes the unit sphere of the Euclidean space \( \mathbb{R}^{q+1} \); \( S^q := \{ (x_1, \ldots, x_{q+1}) : \sum_{k=1}^{q+1} x_k^2 = 1 \} \), and \( \mu_q^* \) denotes the volume (surface area) measure on \( S^q \).

1.1. Zonal function networks

In recent years, neural and radial basis function networks have been used extensively in many applications involving pattern recognition, artificial intelligence, density estimation, etc. One of the reasons for their effectiveness is their “universal approximation property”; i.e., their ability to approximate arbitrary continuous functions on arbitrary compact subsets of Euclidean spaces of arbitrary dimensions arbitrarily well. Moreover, these networks allow a fast, parallel computation. In the context of the sphere \( S^q \), we have introduced in [19] the notion of a ZF network. A ZF network with \( n \) neurons and an activation function \( \phi : [-1, 1] \to \mathbb{R} \) evaluates a function of the form \( \sum_{k=1}^{n} a_k \phi(\cdot - \xi_k) \), where \( a_k \)’s are real numbers, \( \xi_k \)’s are points on the sphere, and \( \cdot \) denotes the inner product on the Euclidean space. A ZF network is clearly a neural network (with all thresholds set to 0), but it can also be viewed upon as a radial basis function network, since \( x \cdot \xi = 1 - ||x - \xi||^2/2 \) for \( x, \xi \in S^q \).

In many applications, one has to approximate functions based on data collected at sites, without having a choice on the location of these sites. Such sites will be called scattered sites. The most natural approach is to interpolate the data using ZF networks. Indeed, one reason for the popularity of such networks is that if the Legendre coefficients (defined in (2.8)) satisfy \( \hat{\phi}(\ell) > 0 \) for each \( \ell = 0, 1, \ldots \), then interpolation is always possible at scattered sites with ZF networks with activation function \( \phi \). An activation function with this property has been called a spherical basis function, and many mathematicians have studied various aspects of interpolation by such networks, including the degree of approximation, and computational procedures. Recent surveys include [28,32].

There are several reasons for studying processes other than interpolation for approximation of functions. Firstly, in the case of experimental data, one expects the data to contain some noise or experimental errors. Therefore, one may not wish to reproduce such data exactly. Secondly, an interpolatory network will necessarily have as many neurons as the number of elements in the data set. Therefore, in the case of large data sets, it is natural to look for more parsimonious representations. Thirdly, the known estimates on the degree of approximation provided by interpolatory ZF networks are not asymptotically optimal. In fact, we are not aware of any estimates where the degree of approximation by interpolation in the uniform norm is estimated in a non-trivial way in the case when the smoothness is also assumed using the uniform norm. Fourthly, a necessary and sufficient condition for the ZF networks with activation function \( \phi \) to have the universal approximation property is that \( \hat{\phi}(\ell) \neq 0 \) for any integer \( \ell \geq 0 \); i.e., approximation is possible even in the case when interpolation might not be. Finally, there are such computational issues as the stability of the interpolation matrix, and the need to invert this matrix, that may make interpolation a costly procedure.

In [19], we have analyzed the approximation properties of ZF networks with activation function \( \phi \) with the only assumption that \( \hat{\phi}(\ell) \neq 0 \) for any integer \( \ell \geq 0 \). In the case when the
coefficients \( \hat{\phi}(f) \) decay exponentially, and the target function is in a Sobolev class, the degree of approximation of a target function by our networks characterizes the Sobolev class to which the target function belongs. In particular, the degree of approximation of such functions is optimal in the sense of nonlinear \( n \)-widths. We give explicit formulas for the networks that do not require any “training” in the customary sense, thus avoiding pitfalls of such techniques as backpropagation. Our formulas also do not require the solution of a matrix equation. An essential tool in our constructions is given by quadrature formulas, based on scattered sites, that integrate high degree spherical polynomials exactly. The existence of these formulas is proved in [18], as a consequence of Marcinkiewicz–Zygmund (M–Z) inequalities relating the continuous and discrete norms of spherical polynomials.

1.2. Quadrature formulas

Many applications related to earth sciences and differential equations on the sphere require the approximate calculation of an integral of the form \( \int_{S^2} f d\mu^s \) by a quadrature formula of the form \( Q_C(f) = \sum_{\xi \in C} w_\xi f(\xi) \), where \( C \) is a finite subset of \( S^2 \) and the weights \( w_\xi \) are real numbers. Typically, one requires that the quadrature formula be exact for spherical polynomials of a certain degree \( n \); i.e., \( Q_C(f) = \int_{S^2} f d\mu^s \) if \( f \) is a spherical polynomial of degree at most \( n \). Many quadrature formulas are known, based on specific choices of \( C \), for example, product Gaussian rules in the book [30] by Stroud, Driscoll–Healy formulas [6], and some newer formulas by Petrushev [26], Brown et al. [4], Potts et al. [27], etc. The existence of quadrature formulas based on scattered data was proved by Jetter et al. [13]. Their work was refined in [18] to prove the existence of quadrature formulas with positive weights.

The part of our paper dealing with quadrature formulas is motivated by the following theorem of Hesse and Sloan [11], where the notations will be explained precisely in Section 2.

**Theorem 1.1.** For each integer \( n \geq 1 \), let \( C_n \) be a set of points on \( S^2 \). Suppose that \( \{Q_{C_n}\} \) are quadrature formulas with the following properties. Each \( Q_{C_n} \) is exact for spherical polynomials of degree at most \( n \), and for any spherical cap \( S \), we have

\[
\sum_{\xi \in C_n \cap S} |w_\xi| \leq c_1 (\mu^s_2(S) + 1/n^2),
\]

where \( c_1 \) is a positive constant. Let \( \beta > 1 \) and \( f \in L^2(S^2) \) have a derivative of order \( \beta \), \( \partial^\beta f \in L^2(S^2) \). Then

\[
\left| \int_{S^2} f d\mu^s - Q_{C_n}(f) \right| \leq c_2 n^{-\beta} \| \partial^\beta f \|_2,
\]

where \( c_2 \) is a positive constant independent of \( n \) or \( f \).

We note that a straightforward application of Sobolev embeddings give only a rate of \( n^{1-\beta} \). Thus, the above estimate is a substantial improvement on the obvious one. Hesse and Sloan [12] have proved that this result cannot be improved for any quadrature formula, even if we do not require any conditions whatever on the quadrature formula, apart from the number of function evaluations involved. Brauchart and Hesse [2] have recently extended Theorem 1.1 to the case of \( L^2(S^2) \).
1.3. Objectives of the present paper

One purpose of this paper is to construct ZF networks with an activation function \( \phi \) for which the coefficients \( \{ \hat{\phi}(\ell) \} \) have only a polynomial decay. The bounds obtained for our networks in [19] are not optimal in this case. The constructions of the present paper yield an optimal error bound.

The other goal of this paper is to extend Theorem 1.1 in many different ways. We will consider the case of a general \( S^q \) and functions having derivatives in \( L^p(S^q), 1 \leq p \leq \infty \). We will also study the approximation of weighted integrals of the form \( \int_{S^q} f w \, d\mu_q^* \) for a general class of weight functions \( w \). Such integrals arise naturally if one wants to integrate functions defined on closed manifolds parameterized by the sphere; the weight \( w \) arising as the Jacobian of the parameterization. We will give a simple construction of quadrature formulas, starting from quadrature formulas for the unweighted integral.

During the course of our proofs, we will prove the equivalence of conditions of the form (1.1) with M–Z inequalities. We find this equivalence of independent interest, since it provides an intrinsic characterization of measures that satisfy M–Z inequalities.

The basic idea that unifies the construction of ZF networks and error bounds on quadrature formulas is the following observation. We have proved in [9] that under certain technical conditions on the coefficients \( \{ \hat{\phi}(\ell) \} \) (cf. Proposition 2.1 below), any function \( f \) in a suitable Sobolev class on the unit sphere \( S^q \) embedded in a Euclidean space \( \mathbb{R}^{q+1} \) admits a representation of the form

\[
f(x) = \int_{S^q} \phi(x \cdot y) D_{\phi} f(y) \, d\mu_q^*(y),
\]

where the \( L^p(S^q) \) norm of \( D_{\phi} f \) can be estimated by the Sobolev norm of \( f \). A careful discretization of the integral using our quadrature formulas in [18] leads to the desired construction of ZF networks yielding the optimal order of decay for the degree of approximation. We note that similar representations have been used several times in the literature on neural networks by many authors, for example, Barron [1], Kurkova and Sanguineti [14], and Murata [21]. In [15], we have studied the question of tractability of the degree of approximation of functions admitting a similar representation. One novelty of the present work is that we give explicit constructions rather than asserting the existence of the approximation.

To obtain the generalization of Theorem 1.1 for weighted integrals with the weight \( w \), we will construct a ZF network approximation \( \sum_{\zeta \in C_n} w_{\zeta} \phi(\cdot \cdot \zeta) \) to the function \( \int_{S^q} \phi(\cdot \cdot x) \, w(x) \, d\mu_q^*(x) \). The error bounds will be obtained by observing that

\[
\sum_{\zeta \in C_n} w_{\zeta} f(\zeta) = \int_{S^q} \left\{ \sum_{\zeta \in C_n} w_{\zeta} \phi(y \cdot \zeta) \right\} D_{\phi} f(y) \, d\mu_q^*(y),
\]

and

\[
\int_{S^q} f(x) w(x) \, d\mu_q^*(x) = \int_{S^q} \left\{ \int_{S^q} \phi(y \cdot x) w(x) \, d\mu_q^*(x) \right\} D_{\phi} f(y) \, d\mu_q^*(y).
\]

Indeed, using an argument similar to the proof of [15, Theorem 1.1], it is not difficult to see that in the case of functions admitting a representation of the form (1.3), the quadrature bounds for weighted integrals are actually equivalent to the approximation bounds on ZF networks.

In Section 2, we develop most of the notations needed in the statement and proof of our main results, formulated in Section 3. The proofs of the results in Section 3 are given in Section 4.
2. Notations and background

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 $\mathbb{R}^{q+1}$. A spherical cap with center $x_0$ and radius $\varepsilon$ is defined by

$$\mathbb{S}_\varepsilon^q(x_0) := \{ x \in \mathbb{S}^q : \| x - x_0 \| \leq \varepsilon \},$$

where $\| \cdot \|$ is the Euclidean norm on $\mathbb{R}^{q+1}$.

The volume (surface area) measure on $\mathbb{S}^q$ will be denoted by $\mu^*_q$. We note that the volume element is invariant under arbitrary coordinate changes. The volume of $\mathbb{S}^q$ is

$$\omega_q := \int_{\mathbb{S}^q} d\mu^*_q = \frac{2\pi^{(q+1)/2}}{\Gamma((q + 1)/2)}.$$

If $v$ is any (possibly signed) measure on $\mathbb{S}^q$, its total variation measure will be denoted by $|v|$. In the sequel, we will write $\int f \, dv$ in place of $\int_{\mathbb{S}^q} f \, dv$.

If $1 \leq p \leq \infty$, $B \subseteq \mathbb{S}^q$ is $v$-measurable, and $f : B \rightarrow \mathbb{R}$ is $v$-measurable, we define the $L^p(v; B)$ norms of $f$ by

$$\|f\|_{v; B, p} := \begin{cases} \int_B |f(\xi)|^p |v|(\xi)^{1/p} & \text{if } 1 \leq p < \infty, \\ \|v\|_{\text{ess sup}_{\xi \in B} |f(\xi)|} & \text{if } p = \infty. \end{cases} \quad (2.2)$$

The space of all $v$-measurable functions on $B$ such that $\|f\|_{v; B, p} < \infty$ will be denoted by $L^p(v; B)$, with the usual convention that two functions are considered equal as elements of this space if they are equal $|v|$-almost everywhere. The space of all uniformly continuous, bounded functions on $B$ will be denoted by $C(B)$, and the symbol $X^p(v; B)$ will denote $L^p(v; B)$ if $1 \leq p < \infty$ and $C(B)$ if $p = \infty$ (equipped with the norm of $L^\infty(v; B)$). Strictly speaking, the class $L^p(v; B)$ consists of equivalence classes. If $f \in L^p(v; B)$ is $v$-almost everywhere equal to a function in $C(B)$, we will assume that this continuous function is chosen as the reprenter of its class. We will omit the mention of the set if it is $\mathbb{S}^q$. Thus, $X^p(\mu^*_q) := X^p(\mu^*_q; \mathbb{S}^q)$, etc. We will also omit the mention of the measure if it is the measure $\mu^*_q$. Thus, $X^p := X^p(\mu^*_q)$, etc. We will denote the conjugate exponent of $p$ by $p'$; i.e., $p'$ is defined by $p/(p-1)$ if $1 < p < \infty$, 1 if $p = \infty$, and $\infty$ if $p = 1$.

We will write $w_q(x) := (1 - x^2)^{(q-2)/2}$, $x \in (-1, 1)$, and the space of all Lebesgue measurable functions $f$ on $[-1, 1]$, for which

$$\|f\|_{w_q; p} := \begin{cases} \left\{ \int_{-1}^1 |f(t)|^p w_q(t) \, dt \right\}^{1/p} & \text{if } 1 \leq p < \infty, \\ \text{ess sup}_{t \in [-1, 1]} |f(t)| & \text{if } p = \infty, \end{cases}$$

is finite, by $L^p(w_q)$, where, as usual, functions that are equal almost everywhere are identified. The class of all continuous functions on $[-1, 1]$ will be denoted by $C([-1, 1])$, and the class $X^p(w_q)$ will denote $L^p(w_q)$ if $1 \leq p < \infty$ and $C([-1, 1])$ if $p = \infty$.

In the remainder of this paper, we adopt the following convention regarding constants. The letters $c, c_1, \ldots$ will denote generic positive constants depending on the fixed parameters of the problem under consideration (for example, $q$, the norms, the smoothness parameter $\beta$ to be introduced below, etc.), and other explicitly indicated quantities. Their values may be different at different occurrences, even within a single formula.

For $x \in \mathbb{R}$, we write $x_+ := x$ if $x > 0$, and $x_+ := 0$ if $x \leq 0$. 
2.1. Spherical polynomials and smoothness classes

For a fixed integer $\ell \geq 0$, the restriction to $S^q$ of a homogeneous harmonic polynomial of degree $\ell$ is called a spherical harmonic of degree $\ell$. Most of the following information is based on [20,29, Section IV.2], although we use a different notation. The class of all spherical harmonics of degree $\ell$ will be denoted by $H^q_\ell$, and for any $t \geq 0$, the class of all spherical polynomials of degree $\ell \leq t$ will be denoted by $\Pi^q_{\ell,t}$. We note that $\Pi^q_{\ell,t} = \bigoplus_{\ell=0}^{t} H^q_\ell$ for any $t \geq 0$. Using the somewhat unusual notation $\Pi^q_{\ell,t}$ for all real $t \geq 0$ will help us to avoid the more cumbersome notation using integer parts, when we wish to make a statement, for example, regarding polynomials of degree at most $n/2$ for an integer $n$, regardless of whether $n$ is odd or even. The spaces $H^q_\ell$'s are mutually orthogonal relative to the inner product of $L^2$. Of course, for any integer $n \geq 0$, $\Pi^q_n = \bigoplus_{\ell=0}^{n} H^q_\ell$, and it comprises the restriction to $S^q$ of all algebraic polynomials in $q+1$ variables of total degree not exceeding $n$. The dimension of $H^q_\ell$ is given by

$$d^q_\ell := \dim H^q_\ell = \begin{cases} 
\frac{2\ell + q - 1}{\ell + q - 1} \left( \frac{\ell + q - 1}{\ell} \right) & \text{if } \ell \geq 1, \\
1 & \text{if } \ell = 0, 
\end{cases}$$

(2.3)

and that of $\Pi^q_n$ is $\sum_{\ell=0}^{n} d^q_\ell$. Furthermore, $L^2 = L^2$-closure\{ $\bigoplus_\ell H^q_\ell$ \}. Hence, if we choose an orthonormal basis $\{Y_{\ell,k} : k = 1, \ldots, d^q_\ell\}$ for each $H^q_\ell$, then the set $\{Y_{\ell,k} : \ell = 0, 1, \ldots \text{ and } k = 1, \ldots, d^q_\ell\}$ is an orthonormal basis for $L^2$.

One has the well-known addition formula [20] :

$$\sum_{k=1}^{d^q_\ell} Y_{\ell,k}(x) Y_{\ell,k}(y) = \frac{d^q_\ell}{\omega_q} P_\ell(q+1; x \cdot y), \quad \ell = 0, 1, \ldots ,$$

(2.4)

where $P_\ell(q+1; x)$ is the degree-$\ell$ Legendre polynomial in $q+1$-dimensions.

The Legendre polynomials are normalized so that $P_\ell(q+1; 1) = 1$, and satisfy the orthogonality relations [20, Lemma 10]

$$\int_{-1}^{1} P_\ell(q+1; x) P_k(q+1; x) w_q(x) \, dx = \omega_q \frac{\omega_{q-1}}{\omega_q} d^q_\ell \delta_{\ell,k}.$$  

(2.5)

In the sequel, $p_\ell(q+1; x)$ will denote the orthonormalized polynomial $P_\ell(q+1; x)$, so that for $x, y \in S^q$,

$$\omega_{q-1}^{-1} p_\ell(q+1; 1) p_\ell(q+1; x \cdot y) = \frac{d^q_\ell}{\omega_q} P_\ell(q+1; x \cdot y) = \sum_{k=1}^{d^q_\ell} Y_{\ell,k}(x) Y_{\ell,k}(y).$$

(2.6)

Let $S$ be the space of all infinitely often differentiable functions on $S^q$, endowed with the locally convex topology induced by the supremum norms of all the derivatives of such functions, and $S^*$ be the dual of this space. For $x^* \in S^*$, we define

$$\hat{x}^*(\ell, k) := x^*(Y_{\ell,k}), \quad k = 1, \ldots, d^q_\ell, \quad \ell = 0, 1, \ldots.$$

(2.7)
A most common example is when \( x^* \) is defined by \( g \mapsto \int g(\xi) f(\xi) \, d\mu^*_q(\xi) \) for some integrable function \( f \) on \( S^q \). In this case, we identify \( x^* \) with \( f \) and use the notation \( \hat{f}(\ell, k) \) to denote the corresponding \( x^*(\ell, k) \). If \( f \in S^* \) and \( \beta \in \mathbb{R} \), then \( \hat{\beta} f \in S^* \) is defined by

\[
\hat{\beta} f(\ell, k) := (\ell + 1)^\beta \hat{f}(\ell, k), \quad k = 1, \ldots, d^q_{\ell}, \quad \ell = 0, 1, \ldots.
\]

For \( \beta > 0 \) and \( 1 \leq p \leq \infty \), the class \( H_{\beta,p} \) consists of \( f \in X^p \) for which \( \hat{\beta} f \in X^p \). If \( f \in L^1(w_q) \), we define the Legendre coefficients of \( \phi \) by

\[
\hat{\phi}(\ell) = \omega_{q-1} \int_{-1}^{1} \phi(t) P_{\ell}(q + 1; t) w_q(t) \, dt.
\]  

This normalization ensures that if \( f \in L^1 \) and \( \phi \in L^1(w_q) \), then the spherical harmonic coefficients of \( \int \phi(\cdot \cdot \cdot y) f(y) \, d\mu^*_q(y) \) are given by \( \hat{\phi}(\ell) \hat{f}(\ell, k) \). We note that the formal expansion of \( \hat{\phi} \) is \( \sum \hat{\phi}(\ell) (d^q_{\ell} / \omega_q) P_{\ell} \).

Let \( a = \{a_n\}_{n=0}^\infty \) be a sequence of real numbers. The forward difference operators are defined by

\[
\Delta a_n := \Delta^1 a_n := a_{n+1} - a_n, \quad \Delta^r a_n := \Delta(\Delta^{r-1} a_n), \quad r = 2, 3, \ldots,
\]

and \( \Delta^0 a_n := a_n \). If \( K \geq 1 \) is an integer and \( z \in \mathbb{R} \), we define

\[
|||a|||_{z,K} := \sup_{0 \leq r \leq K, \nu \geq 0} (v + 1)^{z+r} |\Delta^r a_v|,
\]

\[
|||a|||_{z,K}^* := \sup_{r=1}^{\infty} \sum_{v=0}^{\infty} (v + 1)^{z+r-1} |\Delta^r a_v|.
\]  

and denote by \( B_{z,K} \) (respectively, \( B_{z,K}^* \)) the set of all sequences \( a \) for which \( |||a|||_{z,K} < \infty \) (respectively, \( |||a|||_{z,K} + |||a|||_{z,K}^* < \infty \)). If \( z = 0 \), we will omit it from the notation here. We note that \( |||a|||_{z,K} \) involves a bound on the sequence itself, while \( |||a|||_{z,K}^* \) involves only the forward differences of the sequence.

We summarize certain facts in this connection in the following proposition. The estimate (2.10) is the Bernstein inequality (cf. [9, Proposition 4.3(c)]), the estimate (2.11) is the Nikolskii inequality (cf. [19, Proposition 2.1]). The parts (b) and (c) are proved in [9] (Theorem 2.1 and Lemma 4.3, respectively).

**Proposition 2.1.** Let \( 1 \leq p, r \leq \infty \), \( \beta > 0 \), \( \phi \in L^1(w_q) \), \( K > (q + 1)/2 \) be an integer, and \( \{((\ell + 1)^{-\beta} \hat{\phi}(\ell))^{-1}\} \in B_{K}^* \).

(a) We have

\[
\|\hat{\beta} P\|_p \leq cn^\beta \|P\|_p, \quad P \in \Pi_n^q, \quad n = 1, 2, \ldots,
\]

and

\[
\|P\|_p \leq cn^{(q/r-q/p)+} \|P\|_r, \quad P \in \Pi_n^q, \quad n = 1, 2, \ldots.
\]  

(b) For any \( f \in H_{\beta,p} \), there is a unique \( D \phi f \in X^p \) such that

\[
\hat{f}(\ell, k) = \hat{\phi}(\ell) \hat{D} \phi f(\ell, k), \quad k = 1, \ldots, d^q_{\ell}, \quad \ell = 0, 1, \ldots.
\]
In particular, for almost all \( x \in \mathbb{S}^q \),

\[
f(x) = \int \phi(x \cdot y) D_\phi f(y) \, d\mu_q^*(y). 
\]  

(2.12)

Moreover, \( \|D_\phi f\|_p \leq c \|\hat{\beta} f\|_p \).

(c) If \( h \) is a sequence of nonnegative numbers such that \( h(\ell) = 0 \) for all sufficiently large \( \ell \), \( a \) is a sequence such that \( a^\beta := \{ (\ell + 1)^\beta a(\ell) \} \in \mathcal{B}_K \), then for any \( \alpha < \beta, \alpha \in \mathbb{R} \),

\[
|||a(v)h(v)||^\infty_{r=0}|||^*_a K \leq c |||a|||_K \left\{ h(0) + |||h|||^*_a \right\}. 
\]  

(2.13)

2.2. Polynomial summability operators

Let \( h : [0, \infty) \to \mathbb{R} \), and for some constant \( c > 0 \) (depending on \( h \)), \( h(x) = 0 \) if \( x > c \). We extend \( h \) by setting \( h(x) = 0 \) if \( x < 0 \). We define the kernel

\[
\Phi_y(h; t) := \begin{cases} 
\sum_{\ell = 0}^{\infty} h(\ell/y) \frac{d^\alpha q}{\alpha q} \mathcal{P}_\ell(q+1; t) & t \in \mathbb{R}, \ y \geq 1, \\
0 & t \in \mathbb{R}, \ y < 1. 
\end{cases} 
\]  

(2.14)

If \( \mu \) is a possibly signed measure, the summability operator is defined for \( f \in L^1(\mu; \mathbb{S}^q) \) by

\[
\sigma_y(\mu; h, f, x) := \int \Phi_y(h, x \cdot \xi) f(\xi) \, d\mu(\xi).
\]  

(2.15)

We will denote \( \sigma_y(\mu_q^*; h, f) \) by \( \sigma_y^*(h, f) \).

Let \( K \geq 1 \) be an integer. We will write \( h \in A_K \) if each of the following conditions is satisfied:

(i) \( h : [0, \infty) \to [0, \infty) \), (ii) \( h \) is a \( K - 1 \) times iterated integral of a function of bounded variation,

(iii) \( h(x) = 0 \) if \( x \leq c_1 \), and (iv) \( h(x) = 0 \) if \( x > c \). Here, \( c, c_1 \) are positive constants depending upon \( h \). Further, we will write \( h \in A_K^+ \) if \( h \in A_K \), \( h(x) = 1 \) for \( x \in [0, 1/2] \) and \( h(x) = 0 \) for \( x > 1 \). If \( y \geq 1 \), the sequence \( h_y \) is defined by \( h_y(\ell) = h(\ell/y) \). If \( h \in A_K \), a repeated use of the mean value theorem shows that

\[
|||h_y|||^*_a \leq c(h)y^2, \quad y \geq 1, \quad x \in \mathbb{R}. 
\]  

(2.16)

In the sequel, we will assume \( h \) to be a fixed function, and the dependence of various constants on \( h \) will not be indicated.

If \( \{ \mu_n \} \) is a sequence of measures, we will write \( \{ \mu_n \} \preceq_p \mu_q^* \) if for each integer \( n \geq 0 \), every \( \mu_q^* \)-measurable function is also \( \mu_n \)-measurable, and \( \| f \|_{\mu_q^*; \mathbb{S}^q}, P \leq c(p, \{ \mu_n \})\| f \|_p \) for every \( f \in X^\beta \).

Let \( x \geq 0 \). A (possibly signed) measure \( \nu \) will be called an \( M-Z \) measure of order \( x \) if

\[
\|P\|_{\nu, \mathbb{S}^q, p} \leq M_{v,q,p}(x)\|P\|_{\mu_q^*, \mathbb{S}^q, p}, \quad P \in \Pi_q^x, \quad 1 \leq p \leq \infty 
\]  

(2.17)

for a constant \( M_{v,q,p}(x) \geq 1 \) independent of the polynomial \( P \). The constant \( M_{v,q,p}(x) \) will be called the \( M-Z \) constant for \( \nu \) (\( x \), and \( p \)). The measure \( \nu \) will be called a quadrature measure of order \( x \) if

\[
\int P(\xi) \, d\mu_q^*(\xi) = \int P(\xi) \, d\nu(\xi), \quad P \in \Pi_q^x, 
\]  

(2.18)
and an M–Z quadrature measure of order \(x\) it is both an M–Z measure of order \(x\) and a quadrature measure of order \(x\). We note that \(\mu^*_q\) itself is an M–Z quadrature measure of order \(x\) for each \(x \geq 0\), with \(M_{\mu^*_q} \cdot q, p(x) = 1\). In [18], we have established the existence of discrete M–Z quadrature measures with uniformly bounded M–Z constants, supported on sufficiently dense scattered data sets.

We recall certain properties of the operators \(\sigma_y\). Part (a) of the following proposition is proved in [17, Proposition 4.1], a proof of part (b) can be found in [9, Proposition 4.3].

**Proposition 2.2.** Let \(K \geq q + 1\) be an integer, \(h \in \mathcal{A}_K\), \(h(x) = 0\) if \(x \geq D\), \(\{\mu_n\}\) be a sequence of measures such that each \(\mu_n\) is an M–Z measure of order \(2Dn\), with the M–Z constants bounded independently of \(n\). Let \(1 \leq p \leq \infty\), and \(\beta > 0\).

(a) For \(f \in L^p(\mu_n; \mathbb{S}^d)\),

\[
\|\sigma_n(\mu_n; h, f)\|_{\mu_n; \mathbb{S}^d, p} \leq c\|\sigma_n(\mu_n; h, f)\|_p \leq c\|f\|_{\mu_n; \mathbb{S}^d, p}.
\]

(b) If \(f \in H_{\beta, p}, \{\mu_n\} \preceq_p \mu_q^*\), each \(\mu_n\) is an M–Z quadrature measure of order \(2n\), and \(h \in \mathcal{A}_K^*\), then for \(0 \leq \gamma \leq \beta\),

\[
\|\partial_\gamma f - \partial_\gamma \sigma_n(\mu_n; h, f)\|_p \leq cn^{-\beta}\|\partial_\gamma f\|_p.
\]

3. Main results

Our first theorem concerns approximation by ZF networks. The statement involves two sets of measures \(\mu_n\) and \(v_n\). Out of these, we may always choose \(v_n\) to be a discrete measure supported on \(c2^{nq}\) points. With this choice, the integral expression in (3.1) below defines a ZF network with \(c2^{nq}\) neurons. In the case \(r = p\), the estimate is then optimal in the sense of \(n\)-widths (cf. [19] for further details on \(n\)-widths). In the case when \(r = p = \infty\), we may also choose \(\mu_n\)’s to be discrete measures. These may be different from \(v_n\). Thus, one may use the quadrature measures \(\mu_n\) based on scattered sites, as constructed in [18] to get the coefficients in the ZF network in (3.1) based on the values of the target function at these sites, and use other measures in place of \(v_n\) to achieve such other goals as fast computation. Alternatively, one may use \(\mu_n = v_n\) as in [19].

**Theorem 3.1.** For \(n \geq 1\), let \(\mu_n, v_n\) be M–Z quadrature measures of order \(6(2^n)\), with uniformly bounded M–Z constants. Let \(\phi \in L^1(w_q), K > q\) be an integer, \(1 \leq p \leq \infty, \beta > q/p'\). Let \(h \in \mathcal{A}_K^*\) be a nonnegative and nonincreasing function, \(\{(\ell + 1)^{\beta}\phi(\ell)\}_{\ell=0}^{\infty} \in \mathcal{B}_K\), and \(\{(\ell + 1)^{-\beta}\phi(\ell)^{-1}\}_{\ell=0}^{\infty} \in \mathcal{B}^*_K\). Let \(1 \leq r \leq \infty, (q/r - q/p)_+ < \gamma \leq \beta, f \in X^p \cap H_{r, r}, \text{ and } \{\mu_n\} \preceq_r \mu_q^*, \{\mu_n\} \preceq_p \mu_q^*\). Then

\[
\left\| f - \int D_{\phi} \sigma_2^n(\mu_n; h, f, \xi) \phi(\phi(\xi)) d v_n(\xi) \right\|_p \leq c2^{-n(\gamma - (q/r - q/p)_+)} \|\partial_\gamma f\|_r.
\]

We have the stability estimate

\[
\left\| \int D_{\phi} \sigma_2^n(\mu_n; h, f, \xi) \phi(\phi(\xi)) d v_n(\xi) \right\|_p \leq c\|f\|_{\mu_n; p}.
\]
An immediate example of the activation function $\phi$ satisfying all the conditions in Theorem 3.1 is the Green’s function for the pseudo-differential operator $(-\Delta^s + (q - 1)^2/4)^{\beta/2}$, where $\Delta^s$ is the Laplace–Beltrami operator [7]. In this case, $\phi(\ell) = c(\ell + (q - 1)/2)^{-\beta}$, $\ell = 0, 1, \ldots$. Another simple example is the function $\phi(t) = (1 - t)^s$ with $s = (\beta - q)/2$, $s$ not equal to a positive integer. For this function, we have (cf. [31, Section 9.3(4)], [25, Chapter 4, Section 5.1])

$$\hat{\phi}(\ell) = c \frac{\Gamma(\ell - s)}{\Gamma(\ell + s + q)} = (\ell + 1)^{-\beta} \sum_{k=0}^{\infty} \frac{c_k}{(\ell + 1)^k}.$$

Another example is given by Narcowich and Ward [23]. Let $s = (\beta + 1)/2$, and assume that $s - q/2 - 1$ is a positive integer. Let

$$\Psi_{s}(x) := \frac{1}{(2\pi)^{q+1}} \int_{\mathbb{R}^{q+1}} \frac{\exp(i x \cdot y)}{(1 + \|y\|^2)^s} \, dy,$$

where $\| \cdot \|$ is the Euclidean norm and $dy$ is the Lebesgue measure on $\mathbb{R}^{q+1}$. It is known (cf. [29, Chapter IV, Corollary 1.2]) that for a suitable function $\Psi_s : [0, \infty) \rightarrow \mathbb{R}, \Psi_s(x) = \Psi_s(\|x\|^2)$ for all $x \in \mathbb{R}^{q+1}$. Narcowich and Ward have shown that for the function $\phi$ defined by $\phi(t) = \Psi_s(2 - 2t), t \in [-1, 1],$

$$\hat{\phi}(\ell) = (\ell + 1)^{-\beta} \sum_{k=0}^{\infty} \frac{c_k}{(\ell + 1)^k}.$$

Hence, $\phi$ satisfies the conditions of Theorem 3.1.

We illustrate the use of Theorem 3.1 in a special case. Starting with a data of the form $(\zeta, f(\zeta))_{\zeta \in C}$, where $C$ is a finite set of points on $S^q$, we wish to construct a ZF network to approximate $f$. Let $p$, $\beta$, $\phi$, and $h$ be as in Theorem 3.1. For each integer $m \geq 0$, let $C_m$ be a set of judiciously chosen points on $S^q$ such that there exists positive quadrature formulas of order $6(2^m)$ based on $C_m$; i.e., there exist weights $W_{\zeta,m} > 0$ such that

$$\sum_{\zeta \in C_m} W_{\zeta,m} P(\zeta) = \int P \, d\mu_q^*, \quad P \in \Pi_{6(2^m)}^q.$$

We use as $v_m$ in Theorem 3.1, the measure that associates the mass $W_{\zeta,m}$ with $\zeta, \zeta \in C_m$, and carry out the following steps to construct a ZF network.

1. Find an integer $n$, and weights $w_\zeta \geq 0, \zeta \in C$, such that

$$\sum_{\zeta \in C} w_\zeta P(\zeta) = \int_{S^q} P \, d\mu_q^*, \quad P \in \Pi_{6(2^n)}^q.$$

The measure $\mu_n$ in Theorem 3.1 associates the mass $w_\zeta$ with $\zeta, \zeta \in C$.

2. Let

$$\tilde{\Phi}(t) := \sum_{\ell=0}^{2^n} h\left(\frac{\ell}{2^n}\right) \frac{d^q}{\omega_q \phi(\ell)} \mathcal{P}_\ell(q + 1; t), \quad t \in \mathbb{R}. $$

The function $\tilde{\Phi}$ is the kernel of the operator $D_\phi \sigma_{2^n}^*(h)$ in Theorem 3.1.
3. Let

\[ Z_C(f, x) := \sum_{\zeta \in C_n} W_{\zeta, n} \left\{ \sum_{\zeta \in C} w_{\zeta} \tilde{\Phi}(\zeta \cdot \xi) f(\xi) \right\} \phi(x \cdot \zeta) \]

\[ = \sum_{\zeta \in C} w_{\zeta} \left\{ \sum_{\zeta \in C_n} W_{\zeta, n} \tilde{\Phi}(\zeta \cdot \xi) \phi(x \cdot \zeta) \right\} f(\xi). \]

We observe that the operator \( Z_C \) is a linear operator, defined using only the values \( f(\xi), \xi \in C \). The function \( Z_C(f) \) is a ZF network, whose construction is universal; i.e., does not require any a priori assumptions on \( f \), nor the solution of any optimization problem. At this time, the construction of quadrature formulas to satisfy (3.4) seems to be a difficult problem. We are currently working on the development of good algorithms to achieve this. However, the existence of such formulas is guaranteed if the set \( C \) is sufficiently dense, as described more precisely in [18]. Moreover, the results in [18] guarantee that the integer \( n \) in Step 3 can be found large enough so that \( 2^n \sim \delta_C^{-1} \), where \( \delta_C \) is the mesh norm (fill distance) for \( C \), defined by

\[ \delta_C := \sup_{x \in \mathbb{S}^q} \min_{\xi \in C} \text{dist}(x, \xi). \]

In the case when \( 0 < \gamma \leq \beta \), and \( \partial_\gamma f \in X^p \), Theorem 3.1 thus guarantees that \( \|f - Z_C(f)\|_p = O(\delta_C^p) \). Moreover, (3.2) ensures that if the actual data happens to be \( (\zeta, g(\zeta)), \zeta \in C \), then

\[ \|f - Z_C(g)\|_p \leq c \delta_C^p \|\partial_\gamma f\|_p + c_1 \|f - g\|_{\mu_n}. \]

Moreover, by choosing the set \( C_n \) judiciously, \( Z_C(f) \) can be constructed as a network with a substantially smaller number of neurons than the number of elements in the given set of sites, \( C \). Finally, we note that we do not require \( \phi \) to be a spherical basis function.

Our next theorem is a consequence of Theorem 3.1, and describes error bounds for quadrature formulas that are exact for high degree polynomials. We note again that if the measure \( v_n \) is a discrete measure, the integral expression on the right-hand side of (3.5) is a linear combination of values of \( P \) at the points in the support of \( v_n \).

**Theorem 3.2.** For \( n \geq 1 \), let \( v_n \) be an M–Z quadrature measure of order \( 6(2^n) \). Let \( 1 \leq p \), \( r \leq \infty \), \( \beta > q/p \), \( w \in X^r \), and \( f \in H_{\beta, p} \). \( K > q \) be an integer, and \( h \in A^\beta_K \) be a nonnegative and nonincreasing function. Then

\[ \int P(y)w(y) d\mu_q^*(y) = \int P(y)\sigma_{2^n}^*(h, w, y) d\nu_n(y), \quad P \in \Pi^q_{2n-1}, \quad (3.5) \]

\[ \|\sigma_{2^n}^*(h, w)\|_{\nu_n; \mathbb{S}^q, r} \leq c\|w\|_r, \quad (3.6) \]

and

\[ \left| \int f w d\mu_q^* - \int f(y)\sigma_{2^n}^*(h, w, y) d\nu_n(y) \right| \leq c 2^{-n(\beta - (q/r - q/p^2))} \|w\|_r \|\partial_\beta f\|_p. \quad (3.7) \]

We remark that \( \beta > q/p \) implies that \( \beta - (q/r - q/p^2) = \beta - q/p + q - q/r > 0 \). The condition that \( \|\partial_\beta f\|_p \leq \infty \) with \( \beta > q/p \) implies that \( f \) is a continuous function on \( \mathbb{S}^q \). Hence,
both the integrals on the left-hand side of (3.7) are well defined, even if \( w \in L^1 \), and \( v_n \) is a discrete measure. The estimate in this case is the same as that given by a straightforward application of the Sobolev embedding, namely, \( O(2^{-n(\beta-q/p)}) \). With the stronger assumption that \( w \in X^r \), the estimate improves as \( r \) increases until \( r \geq 1/p' \). For \( r \geq 1/p' \), the estimate is the optimal \( O(2^{-n\beta}) \) as in Theorem 1.1. In particular, in the case when \( w \equiv 1 \), \( \sigma_n^2(h, w) \equiv 1 \) as well, and we may choose \( r = \infty \) to obtain

\[
\left| \int f \, d\mu_q^* - \int f \, dv_n \right| \leq c 2^{-n\beta} \| \partial f \|_p,
\]

generalizing Theorem 1.1 for all \( X^p \). Finally, we remark that in the case when \( p = 2 \), \( w \equiv 1 \), Hesse [10] has shown that for every \( \beta > q/2 \), and any discrete measure \( v_n \) supported on \( 2^{nq} \) points, there exists \( f \in H\beta,2 \) such that

\[
\left| \int f \, d\mu_q^* - \int f \, dv_n \right| \geq c 2^{-n\beta} \| \partial f \|_2.
\]

Thus, the estimate (3.7) cannot be improved in the case of discrete measures supported on a number of points proportional to the dimension of the polynomial space for which an exact quadrature formula is desired.

Next, we give certain equivalent characterizations of M–Z measures, partly to demonstrate that our conditions on the measures \( v_n \) are the same as the continuity conditions required by Hesse and Sloan in [11]. After the paper was submitted, we learnt that the equivalence of parts (a) and (b) of Theorem 3.3 is proved by Dai [5, Theorem 2.1] in the case when the measures are positive, but for all \( p \in (0, \infty) \). If \( r > 0 \), \( A \geq 1 \), and \( v \) is a (possibly signed) measure, we will say that \( v \) is \((A, r)\)-continuous if for every spherical cap \( C \), we have

\[
|v|(C) \leq A(\mu_q^*(C) + r^q).
\]

(3.8)

We observe that \( \mu_q^* \) is \((1, r)\)-continuous for every \( r > 0 \), and so is any measure \( \mu \) of the form \( \mu(B) = \int_B f \, d\mu \) for \( f \in L^\infty \), \( \| f \|_\infty \leq 1 \). It is clear that there is some redundancy in the notation: a signed measure \( v \) is \((A, r)\)-continuous if and only if the positive measure \( |v|/A \) is \((1, r)\)-continuous. Indeed, throughout most of the paper, the constant \( A \) will be unimportant, and will be accordingly replaced by the generic constants \( c, c_1 \), etc. Nevertheless, our formulation above helps us to define the notion for one measure and yet, to avoid a cumbersome formulation of the conditions on different measures.

**Theorem 3.3.** Let \( \{v_n\} \) be a sequence of possibly signed measures on \( S^q \). The following statements (a)–(c) are equivalent.

(a) There exist constants \( c_1, c_2 \), such that each \( v_n \) is \((c_1, c_2/n)\)-continuous.
(b) Each \( v_n \) is an M–Z measure of order \( n \), with M–Z constant bounded independently of \( n \).
(c) For every \( \alpha > 0 \), each \( v_n \) is an M–Z measure of order \( \alpha n \) with M–Z constant bounded independently of \( n \).

If \( v_n \) is a positive quadrature measure of order \( n \), then \( v_n \) is an M–Z quadrature measure of order \( n \), with M–Z constant bounded independently of \( n \).
4. Proofs

Our proofs involve a very detailed analysis of the kernels introduced in Section 2.2. We find it convenient to introduce the kernels in an even more general manner. If \( h \) is a sequence with \( h(\ell) = 0 \) for all sufficiently large \( \ell \), we write

\[
\Psi(h, t) := \omega_q^{-1} \sum_{\ell=0}^{\infty} h(\ell) p_\ell(q + 1; t) p_\ell(q + 1; t) = \sum_{\ell=0}^{\infty} h(\ell) d_{\ell,q}^q \mathcal{P}_\ell(q + 1; t). \tag{4.1}
\]

If \( a \) is another sequence, we denote the sequence \( \{a(\ell)h(\ell)\}_{\ell=0}^{\infty} \) by \( ah \).

Our analysis of the kernels is summarized in the following proposition, motivated by the arguments in the paper [11] of Hesse and Sloan. In particular, the conditions on the measures are formulated in terms of the continuity conditions, rather than M–Z conditions.

**Proposition 4.1.** Let \( D \geq 1, h \) be a sequence of nonnegative numbers such that \( h(\ell) = 0 \) for all \( \ell \geq D \), and \( \Delta h(\ell) = 0 \) if \( 0 \leq \ell \leq C_1 D \) for some constant \( C_1, 0 < C_1 < 1 \). Let \( \beta > 0, K > q \) be an integer, and \( a \) be a sequence such that \( a^{(\ell)} := \{((\ell + 1)^{\beta} a(\ell))_{\ell=0}^{\infty} \in \mathcal{B}_K \). Suppose \( A \geq 1, C_2 > 0, C_2/D \leq \rho \leq 1, \) and \( \mu \) is a (possibly signed) \((A, \rho)\)-continuous measure on \( \mathbb{S}^q \). Let \( 1 \leq p \leq \infty \). Then there exists a constant \( c \) depending only on \( C_1, C_2, q, K \), such that for every \( x \in \mathbb{S}^q \),

\[
\int |\Psi(h, x \cdot y)| d|\mu|(y) \leq c A |||h||| K(D\rho)^q, \tag{4.2}
\]

\[
\int |\Psi(ah, x \cdot y)| d|\mu|(y) \leq c A |||a^{(\ell)}||| K \left\{ h(0)((D\rho)^q + \rho^{q/2+1/2-K}) + |||h||| K(D\rho)^q D^{-\beta} \right\} =: M_1, \tag{4.3}
\]

\[
\int |\Psi(ah, x \cdot y)| d\mu^*_q(y) \leq c |||a^{(\ell)}||| K (h(0) + D^{-\beta} |||h||| K) =: M_2. \tag{4.4}
\]

Hence,

\[
|||\Psi(ah, x \cdot \circ)|||_\mu; \mathbb{S}^q, p \leq M_1^{1/p} (D^q M_2)^{1/p'}, \tag{4.5}
\]

and

\[
\left\| \int |\Psi(ah, \circ \cdot y)| d|\mu|(y) \right\|_p \leq c (|\mu|(\mathbb{S}^q) M_2)^{1/p} M_1^{1/p'}. \tag{4.6}
\]

The proof of this proposition utilizes certain estimates on the kernels, summarized in the following lemma. We observe that different analogues of the estimate (4.7) below are also obtained independently by Brown and Dai [3] and Narcowich et al. [22].

**Lemma 4.1.** Let \( D \geq 1, h \) be a sequence of nonnegative numbers such that \( h(\ell) = 0 \) for all \( \ell \geq D \), \( \beta > 0, K > q \) be an integer, and \( a \) be a sequence such that \( a^{(\ell)} := \{((\ell + 1)^{\beta} a(\ell))_{\ell=0}^{\infty} \in \mathcal{B}_K \). There exists a constant \( c \), depending only on \( q \) and \( K \), such that

\[
|\Psi(h, \cos \theta)| \leq c \begin{cases} D^q |||h||| K & \text{if } 0 \leq \theta \leq 1/D, \\ |||h||| q/2+1/2-K, K & \theta^{q/2+1/2-K} & \text{if } 1/D \leq \theta \leq \pi/2, \\ |||h||| q-K, K & \text{if } \pi/2 < \theta \leq \pi, \end{cases} \tag{4.7}
\]
and
\[ |\Psi(ah, \cos \theta)| \leq c \begin{cases} D^q |||a^\beta|||_K \{h(0) + |||h|||^*_q/2+1/2-K-K, K\} \theta^{-q/2+1/2-K} & \text{if } 0 \leq \theta \leq 1/D, \\ |||a^\beta|||_K \{h(0) + |||h|||^*_q/2+1/2-K-K, K\} & \text{if } 1/D \leq \theta \leq \pi/2, \\ |||a^\beta|||_K \{h(0) + |||h|||^*_q/2+1/2-K-K, K\} & \text{if } \pi/2 < \theta \leq \pi. \end{cases} \]

**Proof.** The estimate (4.7) follows from [16, Lemma 4.10]. In order to prove (4.8), we use (4.7) with ah in place of h, and use (2.13). \(\square\)

**Proof of Proposition 4.1.** To simplify our notations, we make some normalizations, valid only in this proof. We write \(g_0 = h(0)/||h||||h||^*_K\). We may replace a by \(a|||a^\beta|||_K, h by \hbar/||h||||h||^*_K\), and \(\mu by |\mu|/A\), and assume thereby that \(|||a^\beta|||_K = \{h|||^*_K = A = 1, and that \(\mu\) is a \((1, \rho)\)-continuous positive measure. Let \(x \in S^q\). In this proof only, we denote for \(0 \leq \phi \leq \theta \leq \pi\) the band \(\{y \in S^q : \cos \theta \leq x \cdot y \leq \cos \phi\}\) by \(B(\phi, \theta)\), and write \(F(\theta) := \mu(B(0, \theta))\). Let \(r\) be chosen so that \(r = 2 \sin(r/2)\). Then \(2r/\pi \leq \rho \leq r\), and \(C_2/D \leq r \leq \pi/3\). Our assumption that \(\mu\) is \((1, \rho)\)-continuous implies that
\[ F(\theta) \leq c \max(r^q, \theta^q) \leq c \max(\rho^q, \theta^q), \quad \theta \in [0, \pi]. \]

Finally, we note that the conditions that \(\Delta h(\ell) = 0 \text{ for } \ell \leq C_1D\) and \(h(\ell) = 0 \text{ for } \ell \geq D\) and the definition (2.9) imply that
\[ |||h|||^*_s, K \leq c D^s, \quad s \in \mathbb{R}. \]

The proofs of (4.2) and (4.3) are very similar. We prove (4.3), and indicate the differences for the proof of (4.2).

In view of the first estimate in (4.8), (4.10), and (4.9), we have
\[ \int_{B(0, r)} |\Psi(ah, x \cdot y)| d\mu(y) \leq c D^q (g_0 + D^{-\beta}) F(r) \leq c (\rho D)^q (g_0 + D^{-\beta}). \]

Similarly, in view of the last estimate in (4.8), (4.10), and (4.9), we obtain that
\[ \int_{B(\pi/2, \pi)} |\Psi(ah, x \cdot y)| d\mu(y) \leq c (g_0 + D^{q-K-\beta}). \]

The second estimate in (4.8) and (4.10) show that
\[ \int_{B(r, \pi/2)} |\Psi(ah, x \cdot y)| d\mu(y) \leq c (g_0 + D^{q/2+1/2-K-\beta}) \int_r^{\pi/2} \theta^{-q/2+1/2-K} dF(\theta). \]

Using integration by parts and (4.9), and keeping in mind that \(\rho \leq r \leq \pi/3, K > (q+1)/2\), we obtain that
\[ \int_r^{\pi/2} \theta^{-q/2+1/2-K} dF(\theta) \leq c + c_1 r^{q/2+1/2-K} + c_2 \int_r^{\pi/2} \theta^{-q/2-1/2-K} d\theta \leq c_3 + c_4 r^{q/2+1/2-K} \leq c_5 \rho^{q/2+1/2-K}. \]
Hence, recalling that \( D \rho \geq C_2 \),
\[
\int_{B(r,\pi/2)} |\Psi(ah, x \cdot y)| \, d\mu(y) \leq c(g_0 + D^{q/2+1/2-K-\beta})\rho^{q/2+1/2-K} \\
\leq cg_0\rho^{q/2+1/2-K} + c_1 D^{-\beta}. \tag{4.15}
\]

The estimates (4.11), (4.12), and (4.15) lead to (4.3).

The proof of (4.2) is almost the same as that of (4.3), except that we use (4.7) in place of (4.8). The main difference is that we no longer need to keep the term \( g_0 \) in the estimates above. Thus, (4.7), (4.9), and (4.10) show that in place of (4.11), we have
\[
\int_{B(0,r)} |\Psi(h, x \cdot y)| \, d\mu(y) \leq cD^{q} F(r) \leq c(D\rho)^q, \tag{4.16}
\]
in place of (4.12), we have
\[
\int_{B(\pi/2,\pi)} |\Psi(h, x \cdot y)| \, d\mu(y) \leq cD^{q-K}, \tag{4.17}
\]
and in place of (4.13), we have
\[
\int_{B(r,\pi/2)} |\Psi(h, x \cdot y)| \, d\mu(y) \leq cD^{q/2+1/2-K} \int_{r}^{\pi/2} \theta^{q/2+1/2-K} \, dF(\theta). \tag{4.18}
\]
In view of (4.14),
\[
\int_{B(r,\pi/2)} |\Psi(h, x \cdot y)| \, d\mu(y) \leq c(D\rho)^{q/2+1/2-K} \leq c. \tag{4.19}
\]
The estimates (4.16), (4.17), and (4.18) now lead to (4.2).

To obtain (4.4), we use [16, Lemma 4.6] to get
\[
\int |\Psi(ah, x \cdot y)| \, d\mu_q(y) = c \int_{-1}^{1} |\Psi(ah, t)| w_q(t) \, dt \leq c \|\|a\|\|_K. \tag{4.19}
\]
In view of (2.13), and our normalizations, we see that (cf. (4.10))
\[
\|\|a\|\|_K^* \leq c(g_0 + D^{-\beta}).
\]
Together with (4.19), this implies (4.4).

The estimate (4.3) is the same as (4.5) with \( p = 1 \). The estimate (4.8) implies (4.5) with \( p = \infty \). To obtain the estimate (4.5) for \( 1 < p < \infty \), we use the convexity estimate
\[
\|g\|_{\mu;S^q,p} \leq \|g\|_{\mu;S^q,1}^{1/p} \|g\|_{\mu;S^q,\infty}^{1/p'},
\]
which holds for any \( \mu \)-measurable function \( g \) on \( S^q \).

The estimate (4.3) is the same as (4.6) with \( p = \infty \) and (4.4) implies (4.6) with \( p = 1 \). The estimate (4.6) for \( 1 < p < \infty \) is obtained using the above convexity estimate. \( \square \)

Since the conditions on the measures in Proposition 4.1 are formulated in terms of the continuity conditions, we now prove Theorem 3.3 first, so that we may use Proposition 4.1 when the measures involved are M–Z measures. The proof of this theorem uses the Markov–Stieltjes inequalities (cf. [8]); the formulation appropriate in the context of this paper is given in the following lemma.
Lemma 4.2. Let $C$ be any spherical cap, centered at $x_0 \in \mathbb{S}^q$, and $\chi_C$ be its characteristic function. For any integer $n \geq 1$, there exist univariate polynomials $R_1$, $R_2$ of degree at most $n$, the polynomials depending on the radius of $C$ and $q$ alone, such that for any $x_0, x \in \mathbb{S}^q$,

$$R_1(x \cdot x_0) \leq \chi_C(x) \leq R_2(x \cdot x_0), \quad (4.20)$$

and

$$\int (R_2(x \cdot x_0) - R_1(x \cdot x_0)) \, d\mu_q^*(x) \leq c(n^*(C) + 1/n^q). \quad (4.21)$$

Proof. Let $C = \mathbb{S}_{t_0}^q(x_0)$, where $t_0 = 2 \sin(\theta_0/2)$, $0 \leq \theta_0 \leq \pi$. Let $r_0 = \cos \theta_0$. We note that

$$c_1 \theta_0^q \leq \mu_q^*(C) \leq c_2 \theta_0^q.$$

We observe that $\chi_C(x) = \chi(x \cdot x_0), x \in \mathbb{S}^q$, where in this proof only,

$$\chi(t) := \begin{cases} 1 & \text{if } t \in [r_0, 1], \\ 0 & \text{if } t \in [-1, r_0). \end{cases}$$

Let $n'$ be the integer part of $(n + 2)/2$, and

$$n^* := \begin{cases} n' & \text{if } p_{n'-1}(q + 1; r_0) \neq 0, \\ n' - 1 & \text{if } p_{n'-1}(q + 1; r_0) = 0. \end{cases}$$

In view of [8, Theorem I.5.2], there exist univariate polynomials $R_1$ and $R_2$ of degree at most $2n^* - 2 \leq n$ such that

$$R_1(t) \leq \chi(t) \leq R_2(t), \quad t \in [-1, 1], \quad (4.22)$$

and (cf. [8, p. 29], the remark before display (5.4) there)

$$\int_{-1}^1 (R_2(t) - R_1(t))w_q(t) \, dt \leq \left( \sum_{k=0}^{n^*-1} p_k(q + 1; r_0)^2 \right)^{-1}. \quad (4.23)$$

In view of [24, Lemma 5, p. 108], we obtain

$$\left( \sum_{k=0}^{n^*-1} p_k(q + 1; r_0)^2 \right)^{-1} \leq \frac{c}{n} (\theta_0 + 1/n)^{q-1} \leq c(n^*(C) + n^{-q}). \quad (4.24)$$

The estimate (4.22) implies (4.20), and the estimates (4.23) and (4.24) imply (4.21). □

Proof of Theorem 3.3. We prove first that (a) $\Rightarrow$ (c). Let $K > q$ be an integer, and $h \in \mathcal{A}_K^*$. We apply (4.2) with $\nu_n$ (respectively, $\mu_q^*$) in place of $\mu$, $2\kappa n$ in place of $D$, $c_2/n$ in place of $\rho$. Then (2.16) implies that

$$\max \left( \sup_{n \geq 0, x \in \mathbb{S}^q} \int |\Phi_{2\kappa n}(h, x \cdot y)| \, d\mu_q^*(y), \sup_{n \geq 0, x \in \mathbb{S}^q} \int |\Phi_{2\kappa n}(h, x \cdot y)| \, d\nu_n(y) \right) \leq c. \quad (4.25)$$

Therefore, a simple application of the Riesz–Thorin interpolation theorem (cf. [16, Lemma 4.1]) leads to

$$\|\sigma_{2\kappa n}^+(h, f)\|_{\nu_n; \mathbb{S}^q, p} \leq c \|f\|_p, \quad f \in L^p. \quad (4.26)$$

Since $\sigma_{2\kappa n}^+(h, P) = P$ for $P \in \Pi_{2\kappa n}$, this implies part (c).
Clearly, part (c) implies part (b).

Next, let (b) hold, $C$ be a spherical cap, centered at $x_0 \in S^q$. We obtain polynomials $R_1, R_2$ as in Lemma 4.2. Then $R_2(x \cdot x_0) \geq 0$ for all $x \in S^q$. Using (2.17) with $p = 1$, we get

$$|v_n|(C) \leq \int R_2(x \cdot x_0) d|v_n|(x) \leq c \int R_2(x \cdot x_0) d\mu_q^*(x)$$

$$\leq c \int R_1(x \cdot x_0) d\mu_q^*(x) + c \int (R_2(x \cdot x_0) - R_1(x \cdot x_0)) d\mu_q^*(x)$$

$$\leq c\mu_q^*(C) + c_3(\mu_q^*(C) + 1/n^q).$$

This proves part (a).

Finally, let $v_n$ be a positive quadrature measure of order $n$, $C$ be a spherical cap, centered at $x_0 \in S^q$. We obtain polynomials $R_1, R_2$ as in Lemma 4.2. Since $v_n$ is a positive measure, (4.20) implies that

$$\int R_1 d\mu_q^* = \int R_1 dv_n \leq v_n(C) \leq \int R_2 dv_n = \int R_2 d\mu_q^*,$$

and

$$\int R_1 d\mu_q^* \leq \mu_q^*(C) \leq \int R_2 d\mu_q^*.$$  (4.28)

Hence, using (4.21), we deduce that

$$|v_n(C) - \mu_q^*(C)| \leq (R_2 - R_1) d\mu_q^* \leq c(\mu_q^*(C) + 1/n^q).$$

This implies that $v_n$ is $(c_1, 1/n)$-continuous, and hence, an M–Z measure with M–Z constant bounded independently of $n$.  


In order to prove Theorem 3.1, we need some additional preparation. The following lemma gives a sufficient condition for the existence of a function in a suitable $L^p(w_q)$ with a given sequence of coefficients.

**Lemma 4.3.** Let $1 \leq p \leq \infty$, $\beta > q/p'$, $K > q$ be an integer, and $a$ be a sequence of real numbers such that $a^\beta := \{(\ell + 1)^\beta a(\ell)\} \in B_K$. Then there exists $\phi \in X^p(w_q)$ such that $\hat\phi(\ell) = a(\ell)$, $\ell = 0, 1, \ldots$.

**Proof.** In this proof only, let $h$ be a nonincreasing, nonnegative function in $A_{K}^+$, $g(t) := h(t) - h(2t)$, and for $y \geq 1$, $g_y$ be the sequence defined by $g_y(\ell) = g(\ell/y)$, and $h_y(\ell) = h(\ell/y)$. Then the mean value theorem implies as in (2.16) that

$$\|g_y\|_{s,K} \leq cy^s, \quad y \geq 1, \quad s \in \mathbb{R}.  \quad (4.29)$$

Let $m \geq 0$. We now use Proposition 4.1 with $\mu_q^*$ in place of $\mu$, $g_{2m+1}$ in place of $h$, $D = 2^{m+1}$, and $\rho = 1/2^m$. Since $g_{2m+1}(0) = 0$, the quantities $M_1, M_2$ defined there are both $\leq c2^{-m\beta}$. The estimate (4.5) implies that for any $x \in S^q$,

$$\|\Psi(a g_{2m+1}, 0)\|_{w_q, p} = c\|\Psi(a g_{2m+1}, x \cdot 0)\|_p \leq c2^{-m(\beta - q/p')}.$$  (4.30)
Since $\beta > q / p'$, $\sum_{m=0}^{\infty} \| \Psi (ag_{2m+1}, o) \|_w p < \infty$. Thus, $\Psi (ah_1, o) + \sum_{m=0}^{\infty} \Psi (ag_{2m+1}, o)$ converges in the sense of $X^p (w_q)$ to some function $\phi \in X^p (w_q)$. It is easy to check that $\phi (\ell) = a (\ell)$, $\ell = 0, 1, \ldots$. □

The following Proposition 4.2 may be thought of as an analogue of the bounds on quadrature formulas in [11] in the case of approximation bounds.

**Proposition 4.2.** Let $1 \leq p \leq \infty$, $\beta > q / p'$, $\phi \in L^1 (w_q)$, and for some integer $K > q$, $\{(\ell + 1)^{\beta} \phi (\ell)\} \in B_K$, and $h$ be a nonincreasing, nonnegative function in $A^*_K$. Then $\phi \in L^p (w_q)$. Moreover, let for each $n \geq 0$, $\mu_n$ be an $M-Z$ measure of order $2^n$, and the $M-Z$ constants be bounded independently of $n$. With $\psi_x (y) := \phi (x \cdot y)$, $x, y \in \mathbb{S}^q$, we have for integer $n \geq 0$,

$$
\left\| \int \phi (\cdot \cdot y) - \sigma_{2n}^m (h, \psi_y o, y) \right\|_{d |\mu_n | (y)} \leq c 2^{-n \beta}.
$$

(4.31)

**Proof.** In this proof only, let $g (t) := h (t) - h (2t)$, and for $y \in \mathbb{S}^q$, $g_y (\ell) = g (\ell / y)$. We note that (4.29) holds. In this proof only, let $a$ be defined by $a (\ell) = \phi (\ell)$. Let $m \geq n$. We observe that for any integers $\ell, k$, $\hat{\psi}_x (\ell, k) = \hat{\phi} (\ell) Y_{\ell, k} (x)$, and $h (\ell / 2^m) = 0$ if $\ell \geq 2^m$. Therefore, for any $y \in \mathbb{S}^q$, we have

$$
\sigma_{2m+1}^n (h, \psi_x, y) - \sigma_{2m}^n (h, \psi_x, y)
= \sum_{\ell=0}^{2m+1} h (\ell / 2^{m+1}) \hat{\phi} (\ell) \sum_{k=0}^{d^1 \ell} Y_{\ell, k} (x) Y_{\ell, k} (y) - \sum_{\ell=0}^{2m+1} h (\ell / 2^m) \hat{\phi} (\ell) \sum_{k=0}^{d^1 \ell} Y_{\ell, k} (x) Y_{\ell, k} (y)
= \Psi (ag_{2m+1}, x \cdot y).
$$

Since $\beta > q / p'$, the same argument as in the proof of Lemma 4.3 shows that for every $y \in \mathbb{S}^q$,

$$
\psi_x (y) = \sigma_1^n (h, \psi_x, y) + \sum_{m=0}^{\infty} (\sigma_{2m+1}^n (h, \psi_x, y) - \sigma_{2m}^n (h, \psi_x, y))
= \sigma_1^n (h, \psi_x, y) + \sum_{m=0}^{\infty} \Psi (ag_{2m+1}, x \cdot y)
$$

(4.32)

with convergence in the sense of $X^p$ with respect to the variable $x$. This implies that

$$
\left\| \int | \psi_x (y) - \sigma_{2n}^m (h, \psi_y o, y) | d |\mu_n | (y) \right\|_p \\
\leq \left\| \int \sum_{m=n}^{\infty} | \Psi (ag_{2m+1} o, y) | d |\mu_n | (y) \right\|_p \\
\leq \sum_{m=n}^{\infty} \left\| \int | \Psi (ag_{2m+1} o, y) | d |\mu_n | (y) \right\|_p.
$$

(4.33)

We again use Proposition 4.1 with $\mu_n$ in place of $\mu$, $2^{-n}$ in place of $\rho$, $g_{2m+1}$ in place of $h$, $2^{m+1}$ in place of $D$. In view of (4.29), and the fact that $g_{2m+1} (0) = 0$ for each $m$, we see that the quantities $M_1, |\mu_n | (\mathbb{S}^q)$ $M_2$ can be estimated by $c 2^{-m \beta} (2^{m-n}) q$ and $c 2^{-m \beta}$ respectively. Therefore, (4.6)
shows that for $m \geq n$,
\[
\left\| \int |\Psi(\mathbf{a}g_{2n+1}, \circ \cdot \mathbf{y})|d|\mu_n|\mathbf{y} \right\| \leq c2^{-m\beta}(2^{m-n}q/p') .
\]
The estimate (4.33) now yields
\[
\left\| \int |\phi(\circ \cdot \mathbf{y}) - \sigma_{2n}^*(h, \psi_\circ, \mathbf{y})|d|\mu_n|\mathbf{y} \right\| \leq \frac{c}{2^{mq/p'}} \sum_{m=n}^{\infty} 2^{-m(b/q/p')} \leq c2^{-n\beta},
\]
as required. □

We are finally in a position to prove Theorems 3.1 and 3.2.

\textbf{Proof of Theorem 3.1.} We consider the operator
\[
T_n(g, \mathbf{x}) := T_n(\phi, h, g, \mathbf{x}) := \int \phi(x \cdot y)\sigma_{2n+1}^*(h, g, y) dv_n(y) - \int \phi(x \cdot y)\sigma_{2n+1}^*(h, g, y) d\mu_q^*(y).
\]
We note that $P := \int \phi(\circ \cdot y)\sigma_{2n+1}^*(h, g, y) d\mu_q^*(y) \in \Pi_{2n+1}$, and for $x, y \in S^q$,
\[
\sigma_{2n+1}^*(h, \phi(\circ \cdot y), x) = \int \phi(y \cdot \xi)\Phi_{2n+1}(h, \xi \cdot x) d\mu_q^*(\xi)
\]
\[
= \sum_{\ell=0}^{2n+1} \hat{\phi}(\ell)h(\ell/2n+1) \sum_{k=1}^{d_\ell^q} Y_{\ell,k}(x)Y_{\ell,k}(y)
\]
\[
= \int \phi(x \cdot \xi)\Phi_{2n+1}(h, \xi \cdot y) d\mu_q^*(\xi) = \sigma_{2n+1}^*(h, \phi(\circ \cdot x), y).
\]
Therefore, since $v_n$ is a quadrature measure of order $6(2^n)$,
\[
P(x) = \sigma_{2n+2}^*(h, P, x) = \int \sigma_{2n+2}^*(h, \phi(\circ \cdot o), y)\sigma_{2n+1}^*(h, g, y) d\mu_q^*(y)
\]
\[
= \int \sigma_{2n+2}^*(h, \phi(\circ \cdot o), y)\sigma_{2n+1}^*(h, g, y) dv_n(y).
\]
Thus,
\[
T_n(g, \mathbf{x}) = \int (\phi(x \cdot y) - \sigma_{2n+2}^*(h, \phi(\circ \cdot o), y)) \sigma_{2n+1}^*(h, g, y) dv_n(y).
\]
For $x \in S^q$ and $g \in L^\infty$, we have from (4.31) (used with $v_n$ in place of $\mu_n$ and $p = 1$):
\[
|T_n(g, \mathbf{x})| \leq c2^{-n\beta}\|\sigma_{2n+1}^*(h, g)\|_\infty \leq c2^{-n\beta}\|g\|_\infty;
\]
i.e.,
\[
\|T_n(g)\|_\infty \leq c2^{-n\beta}\|g\|_\infty.
\]
Using Fubini’s theorem and (4.31) with $\mu_q^*$ in place of $\mu_n$, we obtain
\[
\|T_n(g)\|_1 \leq c2^{-n\beta}\|\sigma_{2n+1}^*(h, g)\|_{v_n, S^q}.
\]
Since $v_n$ is an M–Z quadrature measure of order $6(2^n)$, (2.19) (with $\mu_q^*$ in place of $\mu_n$) implies that
\[
\| T_n(g) \|_1 \leq c2^{-n\beta} \| \sigma_{2n+1}^* (h, g) \|_{v_n; 1} \leq c2^{-n\beta} \| \sigma_{2n+1}^* (h, g) \|_1 \leq c2^{-n\beta} \| g \|_1.
\]
Together with (4.36), the Riesz–Thorin interpolation theorem now implies that for all $p, 1 \leq p \leq \infty$,
\[
\| T_n(g) \|_p \leq c2^{-n\beta} \| g \|_p.
\] (4.37)

Since $D \phi \sigma_{2n}^* (\mu_n; h, f) \in \Pi_{2n}^q$, we have $\sigma_{2n}^* (h, D \phi \sigma_{2n}^* (\mu_n; h, f) ) = D \phi \sigma_{2n}^* (\mu_n; h, f)$, and hence, conclude using Proposition 2.1(b) that
\[
\left\| \int \phi (\circ \cdot y) D \phi \sigma_{2n}^* (\mu_n; h, f, y) d\nu_n (y) - \sigma_{2n}^* (\mu_n; h, f) \right\|_p
= \left\| \int \phi (\circ \cdot y) D \phi \sigma_{2n}^* (\mu_n; h, f, y) d\nu_n (y) - \int \phi (\circ \cdot y) D \phi \sigma_{2n}^* (\mu_n; h, f, y) d\mu_q^* (y) \right\|_p
= \| T_n (D \phi \sigma_{2n}^* (\mu_n; h, f)) \|_p \leq c2^{-n\beta} \| D \phi \sigma_{2n}^* (\mu_n; h, f) \|_p
\leq c2^{-n\beta} \| \partial_\gamma \sigma_{2n}^* (\mu_n; h, f) \|_p.
\] (4.38)

In view of the Bernstein inequality (2.10), the Nikolskii inequality (2.11), the simultaneous approximation property (2.20) (with $\beta = \gamma$), and the fact that $\{\mu_n\} \preceq_r \mu_q^*$, $\{\mu_n\} \preceq_p \mu_q^*$, we have
\[
\| \partial_\gamma \sigma_{2n}^* (\mu_n; h, f) \|_p \leq c2^{n(\beta - \gamma)} \| \partial_\gamma \sigma_{2n}^* (\mu_n; h, f) \|_p
\leq c2^{n(\beta - \gamma + (q/r - q/p)_+)} \| \partial_\gamma \sigma_{2n}^* (\mu_n; h, f) \|_r
\leq c2^{n(\beta - \gamma + (q/r - q/p)_+)} \| \partial_\gamma f \|_r.
\]

Thus, (4.38) implies that
\[
\left\| \sigma_{2n}^* (\mu_n; h, f) - \int \phi (\circ \cdot y) D \phi \sigma_{2n}^* (\mu_n; h, f, y) d\nu_n (y) \right\|_p
\leq c2^{-n(\gamma - (q/r - q/p)_+)} \| \partial_\gamma f \|_r.
\] (4.39)

Using the Nikolskii inequalities (2.11) and the approximation bound (2.20) again (with 0 in place of $\gamma$ and $\gamma$ in place of $\beta$), we obtain
\[
\| \sigma_{2m+1} (\mu_{m+1}; h, f) - \sigma_{2m} (\mu_m; h, f) \|_p
\leq c2^{m(q/r - q/p)_+} \| \sigma_{2m+1} (\mu_{m+1}; h, f) - \sigma_{2m} (\mu_m; h, f) \|_r
\leq c2^{m(q/r - q/p)_+} \left\{ \| \sigma_{2m+1} (\mu_{m+1}; h, f) - f \|_r + \| f - \sigma_{2m} (\mu_m; h, f) \|_r \right\}
\leq c2^{m(q/r - q/p)_+} \| \partial_\gamma f \|_r.
\]

Since $\gamma > (q/r - q/p)_+$, this implies that
\[
\| f - \sigma_{2n}^* (\mu_n; h, f) \|_p \leq \sum_{m=n}^{\infty} \| \sigma_{2m+1} (\mu_{m+1}; h, f) - \sigma_{2m} (\mu_m; h, f) \|_p
\leq c2^{-n(\gamma - (q/r - q/p)_+)} \| \partial_\gamma f \|_r.
\]
Together with the estimate (4.39), this implies (3.1).
We note that this proves (3.2). The estimate (3.6) is proved in the same way as (4.26).

Proof of Theorem 3.2. Let $P \in \Pi^q_{2^n-1}$. Then $\hat{P}(\ell, k) = 0$ if $\ell > 2^n-1$. Since $h(\ell/2^n) = 1$ if $\ell \leq 2^n-1$, and $v_n$ is a quadrature measure of order $6(2^n)$, we have

$$
\int P(y)\sigma^*_2n(h, w, y) \, dv_n(y) = \int P(y)\sigma^*_2n(h, w, y) \, d\mu^*_q(y)
$$

$$
= \sum_{\ell=0}^{2^n-1} \sum_{k=1}^{d^q} \hat{P}(\ell, k) \hat{w}(\ell, k) = \int P(y)w(y) \, d\mu^*_q(y).
$$

This proves (3.5). The estimate (3.6) is proved in the same way as (4.26).

In view of Lemma 4.3, there exists $K_\beta \in X^{p'}(w_q)$ such that $\hat{K}_\beta(\ell) = (\ell + 1)^{-\beta}$, $\ell = 0, 1, \ldots$. We note that

$$
\partial_{-\beta} w(x) = \int K_\beta(x \cdot y) w(y) \, d\mu^*_q(y)
$$

for almost all $x \in \mathbb{S}^q$, where the fact that $K_\beta \in L^1(w_q)$ ensures the existence of the integral. We apply Theorem 3.1 with the kernel $K_\beta$ in place of $\phi$, and the function $\partial_{-\beta} w$ in place of $f$. Since our choice of $\phi$ implies that

$$
\mathcal{D}_\beta \sigma^*_2n(h, \partial_{-\beta} w) = \sigma^*_2n(h, w),
$$

the estimate (3.1) with $p'$ in place of $p$ and $\mu^*_q$ in place of $\mu_n$ implies that

$$
\left\| \partial_{-\beta} w - \int \sigma^*_2n(h, w, y) K_\beta(\cdot \cdot y) \, dv_n(y) \right\|_{p'} \leq c2^{-n(\beta-(q/r-q/p')_+)} \|w\|_r.
$$

(4.40)

Hence,

$$
\left| \int f w \, d\mu^*_q - \int \sigma^*_2n(h, w, y) f(y) \, dv_n(y) \right|
$$

$$
= \int w(y) \int K_\beta(x \cdot y) \partial_{-\beta} f(x) \, d\mu^*_q(x) \, d\mu^*_q(y)
$$

$$
- \int \sigma^*_2n(h, w, y) \int K_\beta(x \cdot y) \partial_{-\beta} f(x) \, d\mu^*_q(x) \, dv_n(y)
$$

$$
= \left\| \int \left\{ \int w(y) K_\beta(x \cdot y) \, d\mu^*_q(y) - \int K_\beta(x \cdot y) \sigma^*_2n(h, w, y) \, dv_n(y) \right\} \partial_{-\beta} f(x) \, d\mu^*_q(x) \right\|_{p'}
$$

$$
\leq \left\| \partial_{-\beta} w - \int \sigma^*_2n(h, w, y) K_\beta(\cdot \cdot y) \, dv_n(y) \right\|_{p'} \|\partial_{-\beta} f\|_p
$$

$$
\leq c2^{-n(\beta-(q/r-q/p')_+)} \|w\|_r \|\partial_{-\beta} f\|_p.
$$

\[\square\]
Acknowledgements

The research of the author was supported, in part, by Grant DMS-0204704 from the National Science Foundation and Grant W911NF-04-1-0339 from the US Army Research Office. We are grateful to Kerstin Hesse and Ian Sloan for their manuscripts [10–12] as well as several useful discussions. We would also like to thank Feng Dai for [5], and the referees for their suggestions for improvement in the first submitted version of the paper.

References