APPROXIMATION BY RIDGE FUNCTIONS AND NEURAL NETWORKS[*]

PENCHO P. PETRUSHEV[†]

Abstract. We investigate the efficiency of approximation by linear combinations of ridge functions in the metric of $L_2(\mathbf{B}^d)$ with $\mathbf{B}^d$ the unit ball in $\mathbf{R}^d$. If X_n is an n-dimensional linear space of univariate functions in $L_2(I)$, $I = [-1,1]$, and Ω is a subset of the unit sphere $\mathbf{S}^{d-1}$ in $\mathbf{R}^d$ of cardinality m, then the space $Y_n := \mathrm{span}\{r(\mathbf{x} \cdot \xi) : r \in X_n, \omega \in \Omega\}$ is a linear space of ridge functions of dimension $\leq mn$. We show that if X_n provides order of approximation $O(n^{-r})$ for univariate functions with r derivatives in $L_2(I)$, and Ω are properly chosen sets of cardinality $O(n^{d-1})$, then Y_n will provide approximation of order $O(n^{-r-d/2+1/2})$ for every function $f \in L_2(\mathbf{B}^d)$ with smoothness of order $r + d/2 - 1/2$ in $L_2(\mathbf{B}^d)$. Thus, the theorems we obtain show that this form of ridge approximation has the same efficiency of approximation as other more traditional methods of multivariate approximation such as polynomials, splines, or wavelets. The theorems we obtain can be applied to show that a feed-forward neural network with one hidden layer of computational nodes given by certain sigmoidal function σ will also have this approximation efficiency. Minimal requirements are made of the sigmoidal functions and in particular our results hold for the unit-impulse function $\sigma = \chi_{[0,\infty)}$.

Key words. approximation error, ridge functions, neural networks

AMS subject classifications. 41A15, 41A25, 41A29

PII. S0036141097322959

1. Introduction. A ridge function is a multivariate function of the form $r(\mathbf{x} \cdot \omega)$, where r is a univariate function, ω is a fixed vector in $\mathbf{R}^d$, the variable $\mathbf{x} \in \mathbf{R}^d$, and $\mathbf{x} \cdot \omega$ is the inner product of $\mathbf{x}$ and ω. These functions appear naturally in harmonic analysis, special function theory, and in several applications such as tomography and neural networks. In most applications, we are interested in representing or approximating a general function f on a domain $\mathbf{\Omega} \subset \mathbf{R}^d$ by linear combinations of ridge functions. It is surprising therefore that the most fundamental questions concerning the efficiency of approximation by ridge functions are unanswered.

In this paper, we shall consider approximating functions in $L_2(\mathbf{B}^d)$, $\mathbf{B}^d$ the unit ball in $\mathbf{R}^d$, $d \geq 2$, by linear combinations of ridge functions. Using extension theorems, the set $\mathbf{B}^d$ can be replaced by more general sets $\mathbf{\Omega} \subset \mathbf{R}^d$.

Let X_n be a linear space of univariate functions in $L_2(I)$, $I := [-1,1]$ and let $\Omega_n \subset \mathbf{S}^{d-1}$ be a finite subset of the unit sphere $\mathbf{S}^{d-1}$ in $\mathbf{R}^d$. Then

$$(1.1) \qquad Y_n := \mathrm{span}\{r(\mathbf{x} \cdot \omega) : r \in X_n, \ \omega \in \Omega_n\}$$

is a space of multivariate ridge functions of dimension $\leq n \#\Omega_n$, where $\#\Omega_n$ is the cardinality of Ω_n. We shall relate the approximation efficiency of Y_n to that of X_n and the distribution of the vectors of Ω_n in $\mathbf{S}^{d-1}$.

Let $W^s(L_2(I))$ denote the univariate Sobolev spaces. We say that a sequence of spaces $X_n, n = 1, 2, \ldots, \dim(X_n) = n$, provides *approximation of order s* if

$$(1.2) \qquad E(g, X_n)_{L_2(I)} \leq c(s) n^{-s} \|g\|_{W^s(L_2(I))}, \quad g \in W^s(L_2(I)),$$

[†] Department of Mathematics, University of South Carolina, Columbia, SC 29208 (pencho@math.sc.edu).

where

$$E(g, X_n)_{L_2(I)} := \inf_{r \in X_n} \|g - r\|_{L_2(I)}$$

is the error in approximating the univariate function g in the $L_2(I)$ norm by the elements of X_n. We denote similarly the multivariate Sobolev space $W^s(L_2(\mathbf{B}^d))$ on $\mathbf{B}^d$ and the approximation error

$$E(f, Y_n)_{L_2(\mathbf{B}^d)} := \inf_{R \in Y_n} \|f - R\|_{L_2(\mathbf{B}^d)}$$

for any $f \in L_2(\mathbf{B}^d)$. Our main result, given in section 8, shows that for any sequence of spaces $X_n, n = 1, 2, \ldots$, which provide approximation of order s, and for appropriately chosen sets Ω_n with $\#\Omega_n = O(n^{d-1})$, the sequence of spaces $Y_n, n = 1, 2, \ldots$, given in (1.1), provide the following approximation: for $\lambda := s + (d-1)/2$,

$$(1.3) \qquad E(f, Y_n)_{L_2(\mathbf{B}^d)} \le c(\lambda) n^{-\lambda} \|f\|_{W^\lambda(L_2(\mathbf{B}^d))}, \quad f \in W^\lambda(L_2(\mathbf{B}^d)).$$

Note that there is in a certain sense an unexpected gain in the multivariate approximation order $s + (d-1)/2$ over the univariate order s. This gain will be explained later (see section 9).

One can generate the space Y_n appearing in (1.3) by using very general univariate spaces X_n such as splines or wavelets. In particular, our results apply to feed-forward neural networks using a very general activation function σ. A complete discussion of the application to neural networks is given in section 9. In this introduction, we wish to illustrate the typical result by considering the following simple example. Let $\sigma = \chi_{[0,\infty)}$ and define X_n as the univariate space spanned by $\sigma(x - k/n), 0 \le k < n$. Then, defining Y_n for this X_n as described above, we obtain a space of dimension $O(n^d)$ of certain piecewise constant functions. The space Y_n can be realized computationally by a feed-forward neural network with $O(n^{d-1})$ computational nodes. In this case (see section 9 for details), (1.3) provides the approximation order $1 + \frac{d-1}{2}$. One might expect the estimate (1.3) to be 1 since we are using piecewise constants in the approximation. As noted in (1.3), the gain of $\frac{d-1}{2}$ in the approximation rate persists in general (see also Theorem 8.2).

There is a standard method in approximation theory (see [DL, Chapter 7]) which derives from (1.3) the estimate

$$(1.4) \qquad E(f, Y_n)_{L_2(\mathbf{B}^d)} \le c \left(\omega_r(f, n^{-1})_{L_2(\mathbf{B}^d)} + \|f\|_{L_2(\mathbf{B}^d)} n^{-r} \right), \quad f \in L_2(\mathbf{B}^d)$$

with ω_r the rth order modulus of smoothness of f. In the case that Y_n contains all polynomials of total degree $< r$ (in d variables), the last term on the right can be eliminated.

Since Y_n is a linear space of dimension $O(n^d)$ then it follows from the general theory of n-widths that for all $m > 0$,

$$(1.5) \qquad \sup_{\|f\|_{W^m(L_2(\mathbf{B}^d))} \le 1} E(f, Y_n) \ge c_0 n^{-m}$$

with $c_0 > 0$ a constant depending only on m and d. In this sense, the estimates (1.3) cannot be improved.

We also note that (1.3) shows that, in general, linear spaces of ridge functions are at least as efficient as other methods of multivariate approximation such as polynomials, wavelets, and splines.

This paper is an extension of the results from [DOP], where we considered the case $d = 2$. Throughout the paper we assume that $d > 2$, although most of the statements hold when $d = 2$.

The results of this paper differ from other work in this field in the following respects. We are able to begin with a very general class of univariate spaces X_n. Other authors (most notably Mhaskar and Micchelli [MM], [MM1], and Mhaskar [M]) have also considered approximation problems of the type treated here. The work of Mhaskar and Micchelli does not give the best order of approximation. Mhaskar [M] has given best possible results but only in the case that X_n is generated using a rather restrictive class of sigmoidal functions.

Our results are, for the present, limited to approximation in L_2, and it remains an important open question in ridge approximation to understand to what extent results such as those presented in this paper are valid in L_p, $p \neq 2$.

It is also an interesting question to understand which sets $\Omega_n \subset \mathbf{S}^{d-1}$, when used in defining the spaces Y_n, will provide the approximation order of (1.3). In the case $d = 2$, as was shown in [DOP], n equally spaced points on $\mathbf{S}^1$ are the most natural choice. There is no direct analogy of equally spaced points in $\mathbf{S}^{d-1}$, $d > 2$. It will become clear from section 4 that any set Ω_n which permits a cubature formula that is exact for spherical polynomials of degree $\leq n$ and with good localization properties will provide spaces Y_n which satisfy (1.3). Since we could not find in the literature examples of such sets Ω_n, we construct some in section 4. There should be more elegant and more natural constructions than ours. In some sense, one might expect that a natural quadrature formula might provide the analogue of equally spaced points in $\mathbf{S}^{d-1}$, $d > 2$.

We prove (1.3) by first understanding well the structure of ridge polynomials. Our main vehicle (given in section 3) is a fundamental orthogonal decomposition of a general function $f \in L_2(\mathbf{B}^d)$ into ridge polynomials. This decomposition uses the univariate Gegenbauer polynomials.

An outline of this paper is the following. The properties we need about Gegenbauer polynomials are given in section 2. In section 3, we give the fundamental orthogonal decomposition of functions in $L_2(\mathbf{B}^d)$ in terms of ridge polynomials. In section 4, we give our construction of cubature (quadrature) formulas. In sections 5–6, we introduce smoothness spaces (the Sobolev spaces) and recall their characterization by polynomial approximation. In section 7, we prove the main theorem about approximation by ridge functions. In section 8, we discuss how to improve the theorem of section 7 to be more amenable to applications. In section 9, we give some applications of our results, in particular to feed-forward neural networks.

Throughout the paper, the constants are denoted by $c, c_1, \ldots$ and they may vary at every occurrence. The constants usually depend on some parameters (like the dimension d) that will be sometimes indicated explicitly.

2. The Gegenbauer (ultraspherical) polynomials. Special functions appear naturally when we represent a general function in terms of ridge polynomials as will be done in the next section. In particular, the Gegenbauer polynomials will play an important role in this paper. In this section, we shall present the essential properties of Gegenbauer polynomials and bring out their role in the Radon transform. We refer the reader to [E] and [Sz] as general references for this section.

The Gegenbauer polynomials are usually defined by the following generating function

$$(1 - 2tz + z^2)^{-\lambda} = \sum_{m=0}^{\infty} C_m^{\lambda}(t) z^m,$$

where $|z| < 1$, $|t| \leq 1$, and $\lambda > 0$. The coefficients $C_m^{\lambda}(t)$ are algebraic polynomials of degree m which are called the Gegenbauer polynomials associated with λ. The family of polynomials $\{C_m^{\lambda}\}_{m=0}^{\infty}$ is a complete orthogonal system for the weighted space $L_2(I, w)$, $I := [-1, 1]$, $w(t) := w_{\lambda}(t) := (1 - t^2)^{\lambda - \frac{1}{2}}$ and we have

$$(2.1) \quad \int_I C_m^{\lambda}(t) C_n^{\lambda}(t) w(t) \, dt = \left\{ \begin{array}{ll} 0, & m \neq n \\ h_{n,\lambda}, & m = n \end{array} \right. \quad \text{with } h_{n,\lambda} := \frac{\pi^{1/2} (2\lambda)_n \Gamma(\lambda + \frac{1}{2})}{(n + \lambda) n! \Gamma(\lambda)},$$

where we use here and later the standard notation

$$(a)_0 := 0, \quad (a)_n := a(a+1) \ldots (a + n - 1) = \Gamma(a + n)/\Gamma(a).$$

Also, we have

$$(2.2) \qquad C_n^{\lambda}(-t) = (-1)^n C_n^{\lambda}(t), \quad C_n^{\lambda}(1) = \frac{(2\lambda)_n}{n!}, \qquad \text{and} \quad C_0^{\lambda}(t) = 1.$$

The Gegenbauer polynomials can also be defined by the following identity (called Rodrigues' formula):

$$(2.3) \ C_n^{\lambda}(t) = (-1)^n \alpha_{n,\lambda} (1 - t^2)^{-\lambda + \frac{1}{2}} \left(\frac{d}{dt} \right)^n (1 - t^2)^{n + \lambda - \frac{1}{2}}, \ \alpha_{n,\lambda} := \frac{(2\lambda)_n}{n! 2^n (\lambda + \frac{1}{2})_n}.$$

There is an identity that relates Gegenbauer polynomials with different weights:

$$(2.4) \qquad \left(\frac{d}{dt} \right)^m C_n^{\lambda}(t) = 2^m (\lambda)_m C_{n-m}^{\lambda + m}(t), \quad m = 1, 2, \ldots, n.$$

Special cases of the Gegenbauer polynomials are the Legendre polynomials P_n and the Chebyshev polynomials of second kind U_n which correspond to $\lambda = 1/2$ and $\lambda = 1$, respectively. Namely,

$$P_n(t) := \frac{(-1)^n}{2^n n!} \left(\frac{d}{dt} \right)^n (1 - t^2)^n = C_n^{1/2}(t),$$

$$U_n(t) := \frac{\sin(n+1) \arccos t}{\sqrt{1 - t^2}} = C_n^1(t).$$

The Chebyshev polynomials of the first kind $T_n(t) := \cos n \arccos t$ can be considered as the Gegenbauer polynomials C_n^0 associated with the weight $w_0(t) = (1 - t^2)^{-\frac{1}{2}}$.

We shall also need the Gegenbauer polynomials C_n^{λ} when $\lambda < 0$ and, in particular, when $\lambda = -1, -2, \ldots$. Note that $\alpha_{n,\lambda} = 0$ when $\lambda = -1, -2, \ldots$ and $n > 2\nu$. Therefore, we cannot use (2.3) to define $C_n^{-\nu}$ when $\nu = 1, 2, \ldots$. However, we can define (see [Sz, Chapter IV])

$$(2.5) \qquad C_n^{\lambda}(t) := \alpha (1 - t^2)^{-\lambda + \frac{1}{2}} \left(\frac{d}{dt} \right)^n (1 - t^2)^{n + \lambda - \frac{1}{2}}, \quad \lambda < 0,$$

where α is any constant independent of t. To our goals the normalization of C_n^{λ} ($\lambda < 0$) is not essential. Identity (2.4) remains valid except for a constant factor (see [Sz, Chapter IV]): for any λ, we have

$$(2.6) \qquad \left(\frac{d}{dt} \right)^m C_n^{\lambda}(t) = c C_{n-m}^{\lambda + m}(t), \quad m = 1, 2, \ldots, n,$$

where c is independent of t.

The Gegenbauer polynomials play a fundamental role in inverting the Radon transform. We shall show in Lemma 2.1 that follows that the Gegenbauer polynomials C_n^λ for $\lambda = k$ and $\lambda = k + 1/2$ (k an integer) are eigenfunctions for certain differential operators that occur in the Radon transform inversion formula. These operators will play an important role in defining an equivalent norm for the weighted Sobolev spaces $W^s(L_2(I, w))$ (see section 8).

We begin with a brief discussion of the Hilbert transform H on $\mathbf{R}$ and its analogue $\mathbf{H}$ for the interval $I := [-1, 1]$. For any $g \in L_1(I)$ we define

$$(2.7) \qquad \mathbf{H}g := Hg^\diamond \qquad \text{with} \qquad g^\diamond(t) := \begin{cases} g(t), & t \in I, \\ 0, & t \in (-\infty, \infty) \setminus I, \end{cases}$$

where $Hg^\diamond$ is the Hilbert transform of $g^\diamond$. It follows that

$$\mathbf{H}g(t) = \frac{1}{\pi}\text{p.v.} \int_{\mathbf{R}^1} \frac{g^\diamond(s)}{t - s}\, ds = \frac{1}{\pi}\text{p.v.} \int_I \frac{g(s)}{t - s}\, ds.$$

The analogue of the Hilbert transform on the circle $\mathbf{T}$ is the conjugate operator (see [Z, Chapter II]). If $g \in L_1(\mathbf{T})$, we denote its conjugate function by

$$\tilde{g}(\tau) := \frac{1}{2\pi}\text{p.v.} \int_{\mathbf{T}} g(\theta) \cot \frac{\tau - \theta}{2}\, d\theta.$$

For any (nonnegative) weight function w, let $L_2^0(I, w)$ be the space of all $g \in L_2(I, w)$ with weighted mean value zero: $\int_I g(t)\, w(t) dt = 0$. The following proposition gives some properties of $\mathbf{H}$ which we shall use.

PROPOSITION 2.1. *If $g \in L_1(I)$, we define $Tg(\theta) := \text{sgn}\,\theta g(\cos\theta)\sin\theta$ for $\theta \in [-\pi, \pi)$. The Hilbert transform $\mathbf{H}$ satisfies the following properties:*
(a) *If $g \in L_1(I)$ then*

$$(2.8) \qquad \mathbf{H}g(\cos\tau) = -\frac{1}{\sin\tau}\widetilde{Tg}(\tau) \quad \text{a.e. on } (0, \pi).$$

(b) *We have, on $(-1, 1)$, $\mathbf{H}w_1^{-1} = 0$,*

$$(2.9) \qquad \mathbf{H}[w_1^{-1}T_{n+1}] = -U_n \quad \text{and} \quad \mathbf{H}[w_1 U_n] = T_{n+1} \quad \text{for } n = 0, 1, \ldots,$$

and hence

$$(2.10) \qquad \mathbf{H}\frac{d}{dt}[w_1 U_n] = (n + 1)U_n.$$

(c) *The functions $V_n := w_1^{-1}T_n$, $n = 0, 1, \ldots$ (in analogy to $\{U_n\}_{n=0}^\infty$) form a complete orthogonal system for $L_2(I, w_1)$.*
(d) *$\mathbf{H}$ is a one-to-one mapping of $L_2^0(I, w_1)$ onto $L_2(I, w_1)$ with*

$$\mathbf{H}^{-1}h = -\frac{1}{w_1}\mathbf{H}(w_1 h) \quad \text{for } h \in L_2(I, w_1)$$

and

$$(2.11) \qquad \|\mathbf{H}g\|_{L_2(I, w_1)} = \|g\|_{L_2(I, w_1)} \quad \text{for } g \in L_2^0(I, w_1).$$

(e) *The operators* $\mathbf{H}$ *and* $\frac{d}{dt}$ *commute: for any polynomial P, we have*

$$\mathbf{H}\left(\frac{d}{dt}(w_1 P)\right) = \frac{d}{dt}\left(\mathbf{H}(w_1 P)\right).$$

Proof. (a) We apply the substitution $s = \cos\theta$ to the integral that defines $\mathbf{H}g$ and replace t by $\cos\tau$, $0 < \tau < \pi$ and obtain

$$\mathbf{H}g(\cos\tau) = \frac{1}{\pi}\text{p.v.}\int_0^\pi \frac{Tg(\theta)}{\cos\tau - \cos\theta}\,d\theta$$
$$= \frac{1}{2\pi}\text{p.v.}\int_{-\pi}^\pi \frac{Tg(\theta)}{\cos\tau - \cos\theta}\,d\theta,$$

since the integrand is even. Note that $\text{p.v.}\int_I \ldots ds = \text{p.v.}\int_0^\pi \ldots d\theta$ above since the substituting function and its inverse are smooth enough. Now, we use the identity

$$\frac{1}{\cos\tau - \cos\theta} = -\frac{1}{2\sin\tau}\left(\cot\frac{\tau-\theta}{2} + \cot\frac{\tau+\theta}{2}\right)$$

to obtain

$$\mathbf{H}g(\cos\tau) = -\frac{1}{2\sin\tau}\left[\frac{1}{2\pi}\text{p.v.}\int_{-\pi}^\pi Tg(\theta)\cot\frac{\tau-\theta}{2}\,d\theta + \frac{1}{2\pi}\text{p.v.}\int_{-\pi}^\pi Tg(\theta)\cot\frac{\tau+\theta}{2}\,d\theta\right].$$

After substituting $\theta = -\theta'$ in the second integral above and using that Tg is even, we see that the two integrals are equal and, therefore, we obtain (a).

(b) For any function $g \in L_1(I, w_1^{-1})$, we have $T[w_1^{-1}g](\theta) = g(\cos\theta)$. Since the conjugate function of $\cos n\theta$ is $\sin n\theta$, $n = 0, 1, \ldots$, the first two statements in (b) follow from (a). Similar calculations give the last two statements.

(c) This is trivial.

(d) This follows from (b) by using the two bases for $L_2(I, w_1)$ given in (c).

(e) This follows from (2.10). $\square$

We shall next show that the Gegenbauer polynomials are eigenfunctions of certain differential operators that arise in inverting the Radon transform. For functions g defined on $\mathbf{B}^d$, we introduce the following differential operators:

$$(2.12) \qquad \Lambda g := \left(\frac{d}{dt}\right)^{d-1}\left[w_{d/2}g\right],$$

and

$$(2.13) \qquad \mathcal{D} := \Lambda, \ d \text{ odd}, \quad \mathcal{D} := \mathbf{H}\Lambda, \ d \text{ even}.$$

LEMMA 2.1. *Let $d \geq 2$ and define $\mathcal{U}_n := C_n^{d/2}$ for $n = 0, 1, \ldots$. Then we have*

$$(2.14) \qquad \mathcal{D}\mathcal{U}_n = (-1)^{\left[\frac{d-1}{2}\right]}\mu_n \mathcal{U}_n, \quad n = 0, 1, \ldots,$$

and

$$(2.15) \qquad \Lambda^2 \mathcal{U}_n = (-1)^{d-1}\mu_n^2 \mathcal{U}_n, \quad n = 0, 1, \ldots,$$

where

$$(2.16) \qquad \mu_n = (n+1)_{d-1} \asymp n^{d-1}, \quad n = 0, 1, \ldots.$$

Proof. We first consider (2.14) in the case when d is odd, $d = 2k + 1$. From (2.3) and (2.4), we find

$$\mathcal{D}\mathcal{U}_n = \mathcal{D}C_n^{k+1/2} = \left(\frac{d}{dt}\right)^{2k}\left[w_1^{2k}C_n^{k+1/2}\right] = c\left(\frac{d}{dt}\right)^{n+2k}w_1^{2n+2k}$$

$$= c_1\left(\frac{d}{dt}\right)^k C_{n+k}^{1/2} = c_2 C_n^{k+1/2} = c_2\mathcal{U}_n.$$

By examining the coefficients of t^n we obtain that $c_2 = (-1)^k\mu_n$. Thus (2.14) is proved in this case.

Assume now that d is even, $d =: 2k$. Then, again using (2.3), (2.4), and (2.10) (recall that $C_n^1 = U_n$) and the commutativity of $\frac{d}{dt}$ and $\mathbf{H}$, we obtain

$$\mathcal{D}\mathcal{U}_n = \mathcal{D}C_n^k = \left(\frac{d}{dt}\right)^{2k-1}\mathbf{H}\left[w_1^{2k-1}C_n^k\right] = c\left(\frac{d}{dt}\right)^{2k-1}\mathbf{H}\left(\frac{d}{dt}\right)^n w_1^{2n+2k-1}$$

$$= c\left(\frac{d}{dt}\right)^{k-1}\mathbf{H}\left(\frac{d}{dt}\right)^{n+k}w_1^{2n+2k-1} = c_1\left(\frac{d}{dt}\right)^{k-1}\mathbf{H}\frac{d}{dt}\left[w_1 C_{n+k-1}^1\right]$$

$$= c_2\left(\frac{d}{dt}\right)^{k-1}C_{n+k-1}^1 = c_3 C_n^k = c_3 C_n^{d/2} = c_3\mathcal{U}_n.$$

We can calculate c_3 as follows. Let $C_n^k(t) =: c_n t^n + \cdots$ and $U_r(t) =: a_r t^r + \cdots$ with $r := n + 2k - 2$. We find

$$\left(\frac{d}{dt}\right)^{2k-1}\mathbf{H}\left[w_1^{2k-1}C_n^k(t)\right] = c_n\left(\frac{d}{dt}\right)^{2k-1}\mathbf{H}\left[w_1(t)\left((-1)^{k-1}t^{n+2k-2} + \cdots\right)\right]$$

$$= (-1)^{k-1}\frac{c_n}{a_r}\left(\frac{d}{dt}\right)^{2k-2}\left[\mathbf{H}\frac{d}{dt}\left[w_1(t)U_{n+k-2}(t) + \cdots\right]\right]$$

$$= (-1)^{k-1}\frac{c_n}{a_r}(n+2k-1)\left(\frac{d}{dt}\right)^{2k-2}\left[U_{n+2k-2}(t) + \cdots\right]$$

$$= (-1)^{k-1}c_n(n+2k-1)\left(\frac{d}{dt}\right)^{2k-2}\left(t^{n+2k-2} + \cdots\right)$$

$$= (-1)^{k-1}(n+1)_{2k-1}c_n\left(t^n + \cdots\right) = (-1)^{k-1}(n+1)_{2k-1}C_n^k(t),$$

where we used identities (2.3), (2.4), and (2.10). Thus (2.14) is proved in this case as well.

Finally, we consider (2.15). From (2.3) and (2.5), respectively, we have

$$\Lambda\mathcal{U}_n = \Lambda C_n^{d/2} = c\left(\frac{d}{dt}\right)^{n+d-1}\left[(1-t^2)^{n+d/2-1/2}\right]$$

$$= c\left(\frac{d}{dt}\right)^{n+d-1}\left[(1-t^2)^{n+d-1}(1-t^2)^{-d/2+1/2}\right]$$

$$= c_1(1-t^2)^{-d/2+1/2}C_{n+d-1}^{-d/2+1}.$$

Hence, applying Λ once again and using (2.6) gives

$$\Lambda^2\mathcal{U}_n = \Lambda^2 C_n^{d/2} = c_1\left(\frac{d}{dt}\right)^{d-1}C_{n+d-1}^{-d/2+1} = c_2 C_n^{d/2} = c_2\mathcal{U}_n.$$

By calculating the coefficients of t^n, we find that $c_2 = (-1)^{d-1}\mu_n^2$ and we arrive at (2.15). $\quad\square$

3. An orthogonal decomposition of $L_2(\mathbf{B}^d)$ in terms of ridge polynomials. Since we are interested in approximating functions $f \in L_2(\mathbf{B}^d)$ by elements from spaces of ridge functions, it is natural to find a decomposition of f in terms of fundamental building blocks of ridge functions. We shall show in this section that we can take as the building blocks certain ridge polynomials. We begin by describing this decomposition.

If $f, g \in L_2(\mathbf{B}^d)$, we define the inner product

$$(3.1) \qquad\qquad \langle f, g \rangle := \int_{\mathbf{B}^d} f(\mathbf{x})\overline{g(\mathbf{x})} \, d\mathbf{x}.$$

This inner product induces the norm

$$\|f\|_{L_2(\mathbf{B}^d)} := \left(\int_{\mathbf{B}^d} |f(\mathbf{x})|^2 \, d\mathbf{x} \right)^{1/2}.$$

We also define, for $f, g \in L_2(\mathbf{S}^{d-1})$, the inner product

$$(3.2) \qquad\qquad (f, g) := \int_{\mathbf{S}^{d-1}} f(\xi)\overline{g(\xi)} \, d\xi$$

and the norm

$$\|f\|_{L_2(\mathbf{S}^{d-1})} := \left(\int_{\mathbf{S}^{d-1}} |f(\xi)|^2 \, d\xi \right)^{1/2},$$

where $d\xi$ stands for the area (volume) element on $\mathbf{S}^{d-1}$ the unit sphere in $\mathbf{R}^d$.

The Gegenbauer polynomials $C_n^{d/2}$ are the building blocks for our decomposition. Let

$$(3.3) \qquad\qquad \mathcal{U}_n := \left(h_{n,d/2} \right)^{-1/2} C_n^{d/2}, \quad n = 0, 1, \ldots,$$

where $h_{n,d/2}$ is from (2.1). Then $\|\mathcal{U}_n\|_{L_2(I,w)} = 1$ and hence $\{\mathcal{U}_n\}_{n=0}^{\infty}$ is a complete orthonormal system for the weighted space $L_2(I, w)$, $w(t) := w_{d/2}(t) = (1-t^2)^{\frac{d-1}{2}}$. Of course, $\mathcal{U}_n$ depends on the space dimension d, but we are suppressing this dependence in our notation. The reader should think of the space dimension d as arbitrary but fixed throughout.

Let $\mathcal{P}_n$ denote the set of all algebraic polynomials of total degree n in d real variables. That is, each $P \in \mathcal{P}_n$ is a linear combination of monomials $\mathbf{x}^{\mathbf{m}} := x_1^{m_1} \ldots x_d^{m_d}$ with $\mathbf{x} := (x_1, \ldots, x_d)$, $\mathbf{m}$ is a d-tuple $(m_1, \ldots, m_d)$ of nonnegative integers, and $|\mathbf{m}| := m_1 + \cdots + m_d \leq n$.

The polynomials $\mathcal{U}_n(\xi \cdot \mathbf{x})$, $\xi \in \mathbf{S}^{d-1}$, are in $\mathcal{P}_n$ and $\mathcal{U}_n(\xi \cdot \mathbf{x})$ are orthogonal to $\mathcal{P}_{n-1}$ in $L_2(\mathbf{B}^d)$ (proved in the appendix):

$$(3.4) \qquad \int_{\mathbf{B}^d} \mathcal{U}_n(\xi \cdot \mathbf{x}) P(\mathbf{x}) \, d\mathbf{x} = 0 \quad \text{for } \xi \in \mathbf{S}^{d-1} \text{ and } P \in \mathcal{P}_{n-1}.$$

This is why the ridge polynomials $\mathcal{U}_n(\xi \cdot \mathbf{x})$ occur in our decomposition of $L_2(\mathbf{B}^d)$.

THEOREM 3.1. *Each function $f \in L_2(\mathbf{B}^d)$ can be represented uniquely as*

$$(3.5) \qquad\qquad f \stackrel{L_2}{=} \sum_{n=0}^{\infty} Q_n(f),$$

where

$$(3.6) \qquad Q_n(f, \mathbf{x}) := \nu_n \int_{\mathbf{S}^{d-1}} A_n(f, \xi) \mathcal{U}_n(\xi \cdot \mathbf{x}) \, d\xi$$

with

$$(3.7) \qquad A_n(f, \xi) := \int_{\mathbf{B}^d} f(\mathbf{y}) \mathcal{U}_n(\xi \cdot \mathbf{y}) \, d\mathbf{y},$$

and

$$(3.8) \qquad \nu_n := \frac{(n+1)_{d-1}}{2(2\pi)^{d-1}} = \frac{(n+1)(n+2) \cdots (n+d-1)}{2(2\pi)^{d-1}}.$$

Moreover, the operators Q_n, $n = 0, 1, \ldots$, are the orthogonal projectors from $L_2(\mathbf{B}^d)$ onto $\mathcal{P}_n \ominus \mathcal{P}_{n-1}$ and the following Parseval identity holds

$$(3.9) \qquad \|f\|_{L_2(\mathbf{B}^d)}^2 = \sum_{n=0}^{\infty} \|Q_n(f)\|_{L_2(\mathbf{B}^d)}^2 = \sum_{n=0}^{\infty} \nu_n \|A_n\|_{L_2(\mathbf{S}^{d-1})}^2.$$

Next we make a few remarks which will help explain the nature of this decomposition.

(i) For each $n = 0, 1, \ldots$, the function $Q_n(f)$ is an algebraic polynomial (in d variables) of degree n. Indeed, each of the $\mathcal{U}_n(\xi \cdot \mathbf{x})$ is a ridge polynomial of degree n and $Q_n(f)$ is a linear combination of these.

(ii) For each $n = 0, 1, \ldots$, the function $A_n(f, \xi)$, $\xi \in \mathbf{S}^{d-1}$, is a spherical polynomial of degree n. This follows from the fact that each of the $\mathcal{U}_n(\xi \cdot \mathbf{x})$, $x \in \mathbf{B}^d$, is of this type.

(iii) The constants ν_n, $n = 0, 1, \ldots$, are eigenvalues which occur in the Radon inversion formula (see (3.26)).

(iv) Among other reasons, the polynomials $\mathcal{U}_n$ occur in this formula because for each $\xi \in \mathbf{S}^{d-1}$ the weight $w_{d/2}(t) = (1 - t^2)^{(d-1)/2}$ is a constant multiple of the $d-1$-dimensional volume of the intersection of $\mathbf{B}^d$ with the hyperplane $\mathbf{x} \cdot \xi = t$.

(v) The orthogonality of the functions $Q_n(f)$ occurs because for each $\xi \in \mathbf{S}^{d-1}$, the polynomial $\mathcal{U}_n(\mathbf{x} \cdot \xi)$ is orthogonal to all algebraic polynomials of degree $< n$ on $\mathbf{B}^d$ (see (3.4)).

(vi) One can imagine that the integral representation of $Q_n(f)$ can be rewritten as a discrete sum by using some sort of quadrature formula on $\mathbf{S}^{d-1}$ and thereby obtain a discrete decomposition of f in terms of ridge polynomials. In the case $d = 2$, one can simply take the canonical quadrature formula for integrating spherical polynomials (i.e., trigonometric polynomials) which uses equally spaced points on the unit circle. This then gives the orthonormal system $\{\mathcal{U}_n(\omega \cdot \mathbf{x})\}$, $\omega \in \Omega_n$, $n = 0, 1, \ldots$, where $\Omega_n := \{(\cos k\pi/n, \sin k\pi/n)\}_{k=1}^n$. This was used in [DOP] as the vehicle for proving approximation results for ridge functions in two variables. In the case $d \geq 3$, we know no analogous quadrature formula. This necessitates a substantial effort (executed in the following section) to derive (less elegant) quadrature formulas which can be used to discretize the integral representation of $Q_n(f)$.

(vii) The decomposition of Theorem 3.1 is in essence known (see, e.g., [RK] and [LS] for the case $d = 2$). However, we could find no reference which gives it in the above form.

There are several ways in which the decomposition of Theorem 3.1 can be derived. One approach is to derive it from the theory of spherical harmonics. A second approach is Radon transforms and in particular (3.5) is a rewriting of the Radon inversion formula (see [RK]). We shall briefly explain this at the end of this section.

We shall give a simple and direct proof of this decomposition using fundamental identities for the ridge polynomials $\mathcal{U}_n(\xi \cdot \mathbf{x})$, $\xi \in \mathbf{S}^{d-1}$. To keep our exposition more fluid, we shall state these identities without proof and relegate the proofs to the appendix.

We start with the following two fundamental identities (proved in the appendix): for each $\xi, \eta \in \mathbf{S}^{d-1}$, we have

$$(3.10) \qquad \int_{\mathbf{B}^d} \mathcal{U}_n(\xi \cdot \mathbf{x}) \mathcal{U}_n(\eta \cdot \mathbf{x}) \, d\mathbf{x} = \frac{\mathcal{U}_n(\xi \cdot \eta)}{\mathcal{U}_n(1)},$$

and, for each $\eta \in \mathbf{S}^{d-1}$, we have

$$(3.11) \qquad \int_{\mathbf{S}^{d-1}} \mathcal{U}_n(\xi \cdot \mathbf{x}) \mathcal{U}_n(\xi \cdot \eta) \, d\xi = \frac{\mathcal{U}_n(1)}{\nu_n} \mathcal{U}_n(\eta \cdot \mathbf{x}).$$

Proof of Theorem 3.1. Let $f \in L_2(\mathbf{B}^d)$, $d > 2$. From remark (i) and (3.4), it follows that $Q_n(f)$ is in $\mathcal{P}_n \ominus \mathcal{P}_{n-1}$. From identities (3.10) and (3.11), we have $Q_n(g) = g$ whenever $g(\mathbf{x}) = \mathcal{U}_n(\eta \cdot \mathbf{x})$, $\eta \in \mathbf{S}^{d-1}$. Therefore, $Q_n^2 = Q_n$ and hence Q_n is a projector onto a subspace Y_n of $\mathcal{P}_n \ominus \mathcal{P}_{n-1}$. Thus, to prove (3.5), it remains only to show that

$$(3.12) \qquad \dim(Y_n) = \dim(\mathcal{P}_n \ominus \mathcal{P}_{n-1}) = \dim(\mathcal{P}_n^h),$$

where $\mathcal{P}_n^h$ denotes the space of all homogeneous polynomials of degree n.

To prove (3.12), we recall a few well-known facts about spherical harmonics which can be found in Stein and Weiss [SW, Chapter 4]; see also [Se]. Let $\mathcal{H}_n$ denote the space of spherical harmonics of degree n; i.e., $\mathcal{H}_n$ is the set of those functions on $\mathbf{S}^{d-1}$ which are the restriction to $\mathbf{S}^{d-1}$ of a function from $\mathcal{P}_n^h$ which is harmonic in $\mathbf{B}^d$. The spherical harmonics of degree n are orthogonal to those of dimension $m \neq n$ with respect to the inner product (3.2). We have

$$(3.13) \qquad \dim(\mathcal{H}_n) = N(d, n) := \binom{n + d - 1}{n} - \binom{n + d - 3}{n - 2}$$

and

$$(3.14) \qquad \dim(\mathcal{P}_n^h) = \dim(\mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon),$$

where $\epsilon = 0$ if n is even and $\epsilon = 1$ if n is odd.

Write

$$K_n(t) := \frac{N(d, n)}{|\mathbf{S}^{d-1}| C_n^{(d-2)/2}(1)} C_n^{(d-2)/2}(t),$$

where $|\mathbf{S}^{d-1}| := \int_{\mathbf{S}^{d-1}} 1 \, d\xi = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the surface area of $\mathbf{S}^{d-1}$. The function $K_n(\xi \cdot \eta)$ is the reproducing kernel for $\mathcal{H}_n$; i.e.,

$$(3.15) \qquad \int_{\mathbf{S}^{d-1}} S(\xi) K_n(\xi \cdot \eta) d\xi = S(\eta), \quad S \in \mathcal{H}_n.$$

Moreover, a simple identity for Gegenbauer polynomials (see Appendix A3) gives that

$$(3.16) \qquad K_n + K_{n-2} + \cdots + K_\epsilon = \frac{C_n^{d/2}}{|\mathbf{S}^{d-1}|} = \frac{\nu_n \mathcal{U}_n}{\mathcal{U}_n(1)}.$$

Hence, the right side of (3.16) is the reproducing kernel for $\mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$; i.e.,

$$(3.17) \qquad \int_{\mathbf{S}^{d-1}} S(\xi) \frac{\nu_n}{\mathcal{U}_n(1)} \mathcal{U}_n(\xi \cdot \eta) d\xi = S(\eta), \quad S \in \mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon.$$

Note that $A_n(f, \xi)$ is a spherical polynomial of degree n since $\mathcal{U}_n(\xi \cdot \mathbf{y})$ is a spherical polynomial of degree n in ξ. We have $\mathcal{U}_n(-t) = (-1)^n \mathcal{U}_n(t)$ (see (2.2)) and hence $A_n(f, -\xi) = (-1)^n A_n(f, \xi)$. Therefore, $A_n(f) \in \mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$. Thus, Q_n can be considered as a linear operator mapping $\mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$ into Y_n. On the other hand, after multiplying both sides of (3.6) by $\mathcal{U}_n(\eta \cdot \mathbf{x})$ and integrating over $\mathbf{B}^d$ we obtain

$$\int_{\mathbf{B}^d} Q_n(f, \mathbf{x}) \mathcal{U}_n(\eta \cdot \mathbf{x}) \, d\mathbf{x} = \int_{\mathbf{S}^{d-1}} A_n(f, \xi) \left(\nu_n \int_{\mathbf{B}^d} \mathcal{U}_n(\eta \cdot \mathbf{x}) \mathcal{U}_n(\xi \cdot \mathbf{x}) \, d\mathbf{x} \right) d\xi$$
$$= \int_{\mathbf{S}^{d-1}} A_n(f, \xi) \frac{\nu_n}{\mathcal{U}_n(1)} \mathcal{U}_n(\eta \cdot \xi) \, d\xi = A_n(f, \xi),$$

where we used (3.10) and (3.17). Hence, A_n is an operator mapping Y_n onto $\mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$ and it is the inverse operator of Q_n. Therefore, $\dim(Y_n) = \dim(\mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon)$ which together with (3.14) implies (3.12).

Since $Q_n(f)$ is in $\mathcal{P}_n \ominus \mathcal{P}_{n-1}$, it is orthogonal to $Q_j(f)$, $j \neq n$, and therefore we have the first equality in (3.9). For the proof of the second equality in (3.9), we use (3.10) to write

$$\int_{\mathbf{B}^d} Q_n(f, \mathbf{x})^2 \, d\mathbf{x} = \nu_n^2 \int_{\mathbf{S}^{d-1}} \int_{\mathbf{S}^{d-1}} \int_{\mathbf{B}^d} A_n(f, \xi) A_n(f, \eta) \mathcal{U}_n(\xi \cdot \mathbf{x}) U_n(\eta \cdot \mathbf{x}) \, d\mathbf{x} \, d\xi \, d\eta$$
$$= \nu_n^2 \int_{\mathbf{S}^{d-1}} \int_{\mathbf{S}^{d-1}} A_n(f, \xi) A_n(f, \eta) \frac{\mathcal{U}_n(\xi \cdot \eta)}{\mathcal{U}_n(1)} \, d\xi \, d\eta.$$

Since $A_n(f) \in \mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$, then we can use (3.17) to complete the integral with respect to η above. We get

$$\int_{\mathbf{B}^d} Q_n(f, \mathbf{x})^2 \, d\mathbf{x} = \nu_n \int_{\mathbf{S}^{d-1}} A_n(f, \xi)^2 \, d\xi.$$

This completes the proof of (3.9) and the theorem. $\qquad \square$

In the same way that we have proved (3.9) of Theorem 3.1 we obtain the following formulas for calculating inner products:

$$(3.18) \qquad \langle f, g \rangle = \sum_{n=0}^{\infty} \nu_n \int_{\mathbf{S}^{d-1}} A_n(f, \xi) A_n(g, \xi) \, d\xi.$$

We next consider the decomposition (3.5) for ridge functions. Let r be a univariate function in $L_2(I, w)$, $w := w_{d/2}$. Then

$$(3.19) \qquad r(t) = \sum_{n=0}^{\infty} \hat{r}(n) \mathcal{U}_n(t), \quad \hat{r}(n) := \int_I r(t) \mathcal{U}_n(t) w(t) \, dt.$$

It follows that for any $\rho \in \mathbf{S}^{d-1}$, the ridge function $R(\mathbf{x}) := r(\rho \cdot \mathbf{x})$ has the representation

$$(3.20) \qquad R(\mathbf{x}) = \sum_{n=0}^{\infty} \hat{r}(n) \mathcal{U}_n(\rho \cdot \mathbf{x}).$$

Using (3.10) and (3.4), we see that

$$(3.21) \qquad A_n(R, \xi) = \hat{r}(n) \frac{\mathcal{U}_n(\rho \cdot \xi)}{\mathcal{U}_n(1)}.$$

Moreover, if R_1 and R_2 are two such ridge functions corresponding to r_1, ρ_1 and r_2, ρ_2, respectively, then from (3.4), (3.11), and (3.18) we have

$$(3.22) \qquad \langle R_1, R_2 \rangle = \sum_{n=0}^{\infty} \hat{r}_1(n) \hat{r}_2(n) \frac{\mathcal{U}_n(\rho_1 \cdot \rho_2)}{\mathcal{U}_n(1)}.$$

There is another approach to deducing the decomposition of Theorem 3.1 which we want to mention since it brings out the connections between this paper and Radon transforms. For each $f \in L_1(\mathbf{B}^d)$ the Radon transform is defined by

$$(3.23) \qquad \mathcal{R}(f; \xi, t) := \int_{\xi^{\perp} \bigcap \mathbf{B}^d} f(t\xi + \mathbf{y}) \, d\mathbf{y},$$

where $\xi \in \mathbf{S}^{d-1}$, $t \in [-1, 1]$, and $\xi^{\perp} := \{\mathbf{y} \in \mathbf{R}^d : \mathbf{y} \cdot \xi = 0\}$. So, the integration is over the intersection of the hyperplane $\mathbf{y} \cdot \xi = t$ and $\mathbf{B}^d$.

We can recover f from its Radon transform by using the Radon transform inversion formula. The Radon transform inversion formula uses the operator (see, e.g., [L])

$$(3.24) \qquad Kg(t) := K_t g(t) := \begin{cases} \dfrac{(-1)^{\frac{d-1}{2}}}{2(2\pi)^{d-1}} \left(\dfrac{d}{dt}\right)^{d-1} g(t) & \text{for } d \text{ odd}, \\[2ex] \dfrac{(-1)^{\frac{d}{2}-1}}{2(2\pi)^{d-1}} \mathbf{H} \left(\dfrac{d}{dt}\right)^{d-1} g(t) & \text{for } d \text{ even}, \end{cases}$$

where $\mathbf{H}$ is the Hilbert transform (see (2.7)). The following relation is the Radon inversion formula for functions defined on $\mathbf{B}^d$: for every sufficiently smooth function f supported on $\mathbf{B}^d$

$$(3.25) \qquad f(\mathbf{x}) = \int_{\mathbf{S}^{d-1}} h(\xi, \mathbf{x} \cdot \xi) \, d\xi \quad \text{with} \quad h(\xi, t) := K_t \mathcal{R}(f; \xi, t).$$

Lemma 2.1 gives that the functions $\mathcal{U}_n$ are eigenfunctions for the operator $K(w\bullet)$:

$$(3.26) \qquad K(w\mathcal{U}_n) = \nu_n \mathcal{U}_n.$$

We now show the idea of using the Radon inversion formula to derive a representation of f in terms of the ridge polynomials $\{\mathcal{U}_n(\mathbf{x} \cdot \xi)\}$. Since $\{\mathcal{U}_n\}_{n=0}^{\infty}$ is a complete orthonormal system for $L_2(I, w)$, we can expand $\mathcal{R}(f; \xi, \bullet)/w$ in terms of the $\{\mathcal{U}_n\}_{n=0}^{\infty}$ to obtain

$$(3.27) \qquad \frac{\mathcal{R}(f; \xi, \bullet)}{w} = \sum_{n=0}^{\infty} A_n(\xi) \mathcal{U}_n,$$

with

$$(3.28) \qquad A_n(\xi) := \int_I \mathcal{R}(f;\xi,t)\,\mathcal{U}_n(t)\,dt = \int_{\mathbf{B}^d} f(\mathbf{y})\,\mathcal{U}_n(\mathbf{y}\cdot\xi)\,d\mathbf{y}.$$

After multiplying both sides of (3.27) by the weight w and applying the operator $K := K_t$ we get

$$K\mathcal{R}(f;\xi,\bullet) = \sum_{n=0}^{\infty} A_n(\xi)K\,[w\,\mathcal{U}_n] = \sum_{n=0}^{\infty} \nu_n A_n(\xi)\mathcal{U}_n,$$

where we used (3.26). Finally, we insert the above in (3.25) and find

$$f(\mathbf{x}) = \int_{\mathbf{S}^{d-1}} K\mathcal{R}(f;\xi,\mathbf{x}\cdot\xi)\,d\xi = \sum_{n=0}^{\infty} \nu_n \int_{\mathbf{S}^{d-1}} A_n(\xi)\mathcal{U}_n(\mathbf{x}\cdot\xi)\,d\xi$$

which is the decomposition from Theorem 3.1. We leave the details of verifying this approach to the reader.

4. Discrete representation of functions and norms. In this section we shall deduce from Theorem 3.1 a discrete representation of functions by ridge polynomials. To this end we shall use a cubature formula for integration on $\mathbf{S}^{d-1}$, $d > 2$. We need a cubature formula that is exact for all spherical polynomials of degree n. In the case $d = 2$ we used in [DOP] a quadrature formula with equally spaced nodes on the unit circle. Unfortunately, we do not know any "equally spaced points" on $\mathbf{S}^{d-1}$, $d > 2$. Also, we do not know effectively any cubature formula with near equally spaced nodes on $\mathbf{S}^{d-1}$. For this reason we shall use a cubature formula, determined by using spherical coordinates on $\mathbf{S}^{d-1}$. The results of this section are somewhat technical and the reader may just wish to read them briefly at first and proceed to section 5.

The spherical coordinates $(\theta,\phi) := (\theta_1,\theta_2,\ldots,\theta_{d-2},\phi)$ on $\mathbf{S}^{d-1}$ are defined as usual by

$\xi_1 = \cos\theta_1,\ \ \xi_2 = \sin\theta_1\cos\theta_2,\ \ \ldots,\ \ \xi_{d-2} = \sin\theta_1\sin\theta_2\ldots\sin\theta_{d-3}\cos\theta_{d-2},$
$\xi_{d-1} = \sin\theta_1\sin\theta_2\ldots\sin\theta_{d-3}\sin\theta_{d-2}\cos\phi,\ \ \xi_d = \sin\theta_1\sin\theta_2\ldots\sin\theta_{d-3}\sin\theta_{d-2}\sin\phi,$

$0 \leq \theta_j \leq \pi,\ j = 1,2,\ldots,d-2;\ 0 \leq \phi < 2\pi$. We shall denote these identities in vector form briefly by $\xi := \xi(\theta,\phi)$. In these coordinates, the surface element $d\xi$ of $\mathbf{S}^{d-1}$ becomes

$$(4.1)\quad d\xi = (\sin\theta_1)^{d-2}(\sin\theta_2)^{d-3}\ldots\sin\theta_{d-2}\,d\theta_1\,d\theta_2\ldots d\theta_{d-2}\,d\phi =: J(\theta)\,d\theta\,d\phi.$$

We have the following identity for integration in spherical coordinates

$$(4.2) \qquad \int_{\mathbf{S}^{d-1}} f(\xi)\,d\xi = \int_0^{\pi}\ldots\int_0^{\pi}\int_0^{2\pi} f(\xi(\theta,\phi))J(\theta)\,d\theta_1\ldots d\theta_{d-2}\,d\phi,$$

where $J(\theta)$ is the Jacobian given by (4.1). We shall use this to define our cubature.

We wish to construct a cubature that is exact for all spherical polynomials of degree $2n$. Every spherical polynomial of degree $2n$ can obviously be represented in spherical coordinates as a linear combination of terms

$$(4.3) \qquad\qquad (\cos\phi)^{k_{d-1}}(\sin\phi)^{\ell_{d-1}}\prod_{j=1}^{d-2}(\cos\theta_j)^{k_j}(\sin\theta_j)^{\ell_j},$$

where $k_j, \ell_j \geq 0$, and $\max\{k_j + \ell_j : j = 1, 2, \ldots, d-1\} \leq 2(d-1)n$. Also, the Jacobian J is represented in the same terms (see (4.1)). So, we need quadrature formulas for integration over $[0, 2\pi]$ and $[0, \pi]$ that are exact for trigonometric polynomial of degree $2(d-1)n + d - 2$.

We shall use the following quadrature formula for integration on $[0, 2\pi]$ with respect to ϕ

$$(4.4) \qquad Q_{\phi,k}(g) := \sum_{j=0}^{2k} \varrho_j g(\gamma_j) \quad \sim \quad \int_0^{2\pi} g(\phi)\, d\phi,$$

where $\gamma_j := \frac{\pi}{2k+1} + \frac{2\pi j}{2k+1}$ and $\varrho_j := \frac{2\pi}{2k+1}$. The quadrature (4.4) is exact for all trigonometric polynomials of degree k (see [Z, Chapter X]).

Since $0 \leq \theta_j \leq \pi$, we need a quadrature for integration over $[0, \pi]$ that is exact for all trigonometric polynomials of degree k. In addition to this, the quadrature should have good localization properties. We also need to control (asymptotically) the nodes and the coefficients of the quadrature. Since we do not know any quadrature like this, we shall construct one in the following lemma.

LEMMA 4.1. *For any $k = 1, 2, \ldots$, there exists a quadrature*

$$(4.5) \qquad Q_{\theta,k}(g) = \sum_{j=0}^{2k} \lambda_j g(\beta_j) \quad \sim \quad \int_0^{\pi} g(\theta)\, d\theta$$

with the following properties:
  (a) $Q_{\theta,k}(g)$ *is exact for all trigonometric polynomials of degree k;*
  (b) $0 < \beta_0 < \beta_1 < \cdots < \beta_{2k} < \pi$,

$$(4.6) \qquad \beta_j - \beta_{j-1} \leq \pi k^{-1}, \quad j = 0, 1, \ldots, 2k+1;$$

  (c)

$$(4.7) \qquad 0 < \lambda_j \leq c\left(\beta_{j+1} - \beta_{j-1}\right), \quad j = 0, 1, \ldots, 2k,$$

where $\beta_{-1} := 0$, $\beta_{2k+1} := \pi$, and c is an absolute constant.

The exact values of the nodes β_j and the coefficients λ_j of the quadrature (4.5) *are given in Remark* 4.1 *below.*

Proof. For symmetry reasons we shall prove the lemma with the interval of integration $[0, \pi]$ replaced by $[-\pi/2, \pi/2]$. We shall build a quadrature

$$(4.8) \qquad Q_k(g) = \sum_{j=-k}^{k} \lambda_j g(\theta_j) \quad \sim \quad \int_{-\pi/2}^{\pi/2} g(\theta)\, d\theta$$

with symmetric nodes θ_j and coefficients λ_j ($\theta_{-j} = \theta_j$, $\theta_0 := 0$, and $\lambda_{-j} = \lambda_j$). Then $Q_k(g)$ will be automatically exact for odd polynomials. Therefore, it is enough to construct $Q_k(g)$ exact only for all even trigonometric polynomials of degree k. To this end it is sufficient to have

$$(4.9) \quad Q_k(P(\cos\theta)) := \sum_{j=-k}^{k} \lambda_j P(\cos\theta_j) = \int_{-\pi/2}^{\pi/2} P(\cos\theta)\, d\theta = 2\int_0^{\pi/2} P(\cos\theta)\, d\theta$$

for each algebraic polynomial P of degree k. We shall apply the substitution $\theta :=$ $\theta(\alpha) := \arccos\left(\cos^2 \frac{\alpha}{2}\right)$ to the last integral in (4.9). Simple calculations show that

$$(4.10) \qquad \Delta(\alpha) := \frac{d}{d\alpha}\theta(\alpha) = \frac{\cos\frac{\alpha}{2}}{\sqrt{1 + \cos^2\frac{\alpha}{2}}}$$

and hence $\theta(\alpha)$ is increasing on $[0,\pi]$ and maps $[0,\pi]$ on $[0,\pi/2]$. We obtain

$$(4.11) \quad \int_0^{\pi/2} P(\cos\theta)\,d\theta = \int_0^\pi P\left(\cos^2\frac{\alpha}{2}\right)\Delta(\alpha)\,d\alpha = \frac{1}{2}\int_{-\pi}^\pi P\left(\cos^2\frac{\alpha}{2}\right)\Delta(\alpha)\,d\alpha,$$

where we used that the integrand is even. We now extend $\Delta(\alpha)$ 2π-periodically by $\Delta(\alpha) := |\cos\frac{\alpha}{2}|/\sqrt{1+\cos^2\frac{\alpha}{2}}$.

We shall use the Dirichlet kernel $D_k(u) := \frac{\sin(k+1/2)u}{2\sin u/2}$ to interpolate the trigonometric polynomial of degree m: $P\left(\cos^2\frac{\alpha}{2}\right) = P\left(\frac{1+\cos\alpha}{2}\right)$ at the points $\alpha_j := \frac{2\pi j}{2k+1}$, $j = 0, \pm 1, \ldots, \pm k$. We have (see [Z, Chapter X])

$$P\left(\cos^2\frac{\alpha}{2}\right) = \frac{2}{2k+1}\sum_{j=-k}^k P\left(\cos^2\frac{\alpha_j}{2}\right)D_k(\alpha - \alpha_j).$$

This and (4.11) imply

$$\int_0^{\pi/2} P(\cos\alpha)\,d\alpha = \frac{1}{2k+1}\sum_{j=-k}^k P\left(\cos^2\frac{\alpha_j}{2}\right)\int_{-\pi}^\pi \Delta(\alpha)D_k(\alpha - \alpha_j)\,d\alpha$$

$$(4.12) \qquad\qquad = \sum_{j=-k}^k \eta_j P\left(\cos^2\frac{\alpha_j}{2}\right)$$

$$= \eta_0 P\left(\cos^2\frac{\alpha_0}{2}\right) + \sum_{j=1}^k 2\eta_j P\left(\cos^2\frac{\alpha_j}{2}\right),$$

where

$$(4.13) \qquad \eta_j := \frac{1}{2k+1}\int_{-\pi}^\pi \Delta(\alpha)D_k(\alpha - \alpha_j)\,d\alpha$$

and we used that $\eta_{-j} = \eta_j$ since Δ is even and $\alpha_{-j} = -\alpha_j$.

We now define the nodes and the coefficients of our quadrature. Set

$$\theta_j := \arccos\left(\cos^2\frac{\alpha_j}{2}\right) \text{ for } j = 0, 1, \ldots, k \text{ and } \theta_j := -\theta_{-j} \text{ for } j = -1, -2, \ldots, -k.$$

Also, set

$$\lambda_j := 2\eta_j \text{ for } j = 0, 1, \ldots, k \text{ and } \lambda_j := \lambda_{-j} \text{ for } j = -1, -2, \ldots, -k.$$

We obtain by (4.9), (4.12) and the symmetry that quadrature (4.8) with the above defined nodes and coefficients is exact for all trigonometric polynomials of degree m. It remains to prove that the nodes and the coefficients of the quadrature satisfy the required properties.

We have $\quad \theta_j - \theta_{j-1} = \theta'(\zeta_j)(\alpha_j - \alpha_{j-1}) = \Delta(\zeta_j)(\alpha_j - \alpha_{j-1}) \quad$ for some $\quad \zeta_j \in (\alpha_{j-1}, \alpha_j)$ and hence, by (4.10),

$$(4.14) \quad \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha_j}{2}\right)\frac{2\pi}{2k+1} \le \theta_j - \theta_{j-1} \le \left(\cos\frac{\alpha_j}{2}\right)\frac{2\pi}{2k+1} < \frac{\pi}{k}, \quad j = 1, 2, \ldots, k.$$

Therefore, the proof will be completed if we show that

$$(4.15) \qquad\qquad 0 < \eta_j \le ck^{-1}\cos\frac{\alpha_j}{2}, \quad j = 1, 2, \ldots, k.$$

By (4.13) it follows that

$$\eta_j = \pi(2k+1)^{-1}S_k(\Delta)(\alpha_j), \quad \text{where} \quad S_k(\Delta)(\alpha) := \tfrac{1}{\pi}\int_{-\pi}^{\pi}\Delta(\beta)D_k(\beta - \alpha)\,d\beta$$

is the kth partial Fourier sum of Δ. In order to simplify our further calculations we shift Δ by π and obtain

$$\varphi(\alpha) := \Delta(\alpha + \pi) = \left|\sin\frac{\alpha}{2}\right|\bigg/\sqrt{1 + \sin^2\frac{\alpha}{2}}\ .$$

The function φ is even and, therefore, its Fourier coefficients associated with $\sin\nu\alpha$ are all equal to zero. Let

$$a_0 := \frac{1}{2\pi}\int_0^{2\pi}\varphi(\alpha)\,d\alpha \quad \text{and} \quad a_\nu := \frac{1}{\pi}\int_0^{2\pi}\varphi(\alpha)\cos\nu\alpha\,d\alpha, \quad \nu = 1, 2, \ldots$$

be the Fourier coefficients of φ associated with $\cos\nu\alpha$. Obviously, $a_0 > 0$. Let $\nu = 1, 2, \ldots$. Then using integration by parts (twice) we get

$$\begin{aligned}
a_\nu &= -\frac{1}{\pi\nu}\int_0^{2\pi}\varphi'(\alpha)\sin\nu\alpha\,d\alpha \\
&= \frac{1}{\pi\nu^2}\left[\varphi'(2\pi^-) - \varphi'(0^+)\right] - \frac{1}{\pi\nu^2}\int_0^{2\pi}\varphi''(\alpha)\cos\nu\alpha\,d\alpha \\
&= \frac{1}{\pi\nu^2}\int_0^{2\pi}\varphi''(\alpha)(1 - \cos\nu\alpha)\,d\alpha.
\end{aligned}$$

Therefore,

$$(4.16) \qquad\qquad a_\nu = \frac{2}{\pi\nu^2}\int_0^{2\pi}\varphi''(\alpha)\sin^2\frac{\nu\alpha}{2}\,d\alpha.$$

Simple calculations show that

$$\varphi'(\alpha) = \frac{1}{2}\cos\frac{\alpha}{2}\bigg/\left(1 + \sin\frac{\alpha}{2}\right)^{3/2} \text{ and } \varphi''(\alpha) < 0 \text{ for } 0 < \alpha < 2\pi.$$

This and (4.16) imply that $a_\nu < 0$ and

$$|a_\nu| \le \frac{2}{\pi\nu^2}\int_0^{2\pi}|\varphi''(\alpha)|\,d\alpha = -\frac{2}{\pi\nu^2}\int_0^{2\pi}\varphi''(\alpha)\,d\alpha = -\frac{2}{\pi\nu^2}\left[\varphi'(2\pi^-) - \varphi'(0^+)\right] = \frac{2}{\pi\nu^2}.$$

Thus, we have

$$(4.17) \qquad\qquad -\frac{2}{\pi\nu^2} \le a_\nu < 0, \quad \nu = 1, 2, \ldots.$$

Therefore $\varphi(\alpha) = a_0 + \sum_{\nu=1}^{\infty} a_\nu \cos \nu\alpha$, where $a_0 > 0$ and $a_\nu < 0$, $\nu = 1, 2, \ldots$, and hence

$$S_k(\varphi)(\alpha) = a_0 + \sum_{\nu=1}^{k} a_\nu \cos \nu\alpha \geq a_0 + \sum_{\nu=1}^{k} a_\nu = S_k(\varphi)(0)$$

$$= S_k(\varphi)(0) - \varphi(0) = -\sum_{\nu=k+1}^{\infty} a_\nu > 0.$$

Thus $S_k(\varphi)(\alpha) > 0$ for $\alpha \in [-\pi, \pi)$ and hence $S_k(\Delta)(\alpha) > 0$ for $\alpha \in [-\pi, \pi)$ which implies the lower bound in (4.15).

The inequalities (4.17) imply

$$\|\Delta - S_k(\Delta)\|_C = \|\varphi - S_k(\varphi)\|_C \leq ck^{-1}.$$

Using this, we obtain

$$|\eta_j| \leq ck^{-1}|S_k(\Delta)(\alpha_j)| \leq ck^{-1} \left(|\Delta(\alpha_j)| + \|\Delta - S_k(\Delta)\|_C\right)$$
$$\leq ck^{-1}\left(\cos\frac{\alpha_j}{2} + k^{-1}\right) \leq ck^{-1}\left(\cos\frac{\alpha_j}{2} + \cos\frac{\alpha_k}{2}\right) \leq ck^{-1}\cos\frac{\alpha_j}{2},$$

where we used that $\cos(\alpha_k/2) = \cos(\pi k/(2k+1)) > ck^{-1}$. Thus the upper estimate in (4.15) is proved. Lemma 4.1 is proved. $\square$

REMARK 4.1. *The exact values of the nodes β_j and the coefficients λ_j of the quadrature* (4.5) *from Lemma* 4.1 *are the following:*

$$\beta_j = \frac{\pi}{2} - \arccos\left(\cos^2\frac{(k-j)\pi}{2k+1}\right), \quad j = 0, 1, \ldots, k,$$

and $\beta_j = \pi - \beta_{2k-j}$, $j = k+1, k+2, \ldots, 2k$;

$$\lambda_j = \frac{2}{2k+1}\int_{-\pi}^{\pi}\cos\frac{\alpha}{2}\left(1 + \cos^2\frac{\alpha}{2}\right)^{-1/2}D_k\left(\alpha - \frac{2\pi(k-j)}{2k+1}\right)d\alpha, \quad j = 0, 1, \ldots, k,$$

and $\lambda_j = \lambda_{2k-j}$, $j = k+1, k+2, \ldots, 2k$, where D_k is the Dirichlet kernel of degree k.

We are now in a position to construct our cubature formula for integration over $\mathbf{S}^{d-1}$ $(d > 2)$. We shall use (4.2), (4.3), and the quadratures from (4.4) and (4.5).

Definition of cubature $\mathbf{Q}_n$. Given $n = 1, 2, \ldots$ we select $k := 2(d-1)n + d - 2$. Let $\mathcal{J}_n$ be the set of all indices $\mathbf{j} := (j_1, \ldots, j_{d-1})$ such that $0 \leq j_\nu \leq 2k$; i.e., $\mathcal{J}_n := \{0, 1, \ldots, 2k\}^{d-1}$. Note that the cardinality of $\mathcal{J}_n$ is $\#\mathcal{J}_n = (2k+1)^{d-1} \asymp n^{d-1}$. Set $\beta_{\mathbf{j}} := (\beta_{j_1}, \ldots, \beta_{j_{d-2}})$, $\gamma_{\mathbf{j}} := \gamma_{j_{d-1}}$, $\omega_{\mathbf{j}} := \xi(\beta_{\mathbf{j}}, \gamma_{\mathbf{j}})$, and $\lambda_{\mathbf{j}} := J(\beta_{\mathbf{j}})\varrho_{j_{d-1}}\prod_{\nu=1}^{d-2}\lambda_{j_\nu}$, where γ_j, ϱ_j, β_j, and λ_j are the nodes and the coefficients of quadratures (4.4) and (4.5), respectively, and J is from (4.1). We define

(4.18) $$\mathbf{Q}_n(f) := \sum_{\mathbf{j} \in \mathcal{J}_n} \lambda_{\mathbf{j}} f(\omega_{\mathbf{j}}) \quad \sim \quad \int_{\mathbf{S}^{d-1}} f(\xi)\, d\xi.$$

When it is possible we shall write this cubature with the following simpler indices. Let Ω_n be the set of all nodes $\omega = \omega_{\mathbf{j}}$, and $\lambda_\omega := \lambda_{\omega_{\mathbf{j}}} := \lambda_{\mathbf{j}}$, $\mathbf{j} \in \mathcal{J}_n$. Then cubature (4.18) can be rewritten in the form

(4.19) $$\mathbf{Q}_n(f) := \sum_{\omega \in \Omega_n} \lambda_\omega f(\omega) \quad \sim \quad \int_{\mathbf{S}^{d-1}} f(\xi)\, d\xi.$$

Observe that $\#\Omega_n \asymp n^{d-1}$.

As we mentioned in the beginning of this section, every spherical polynomial of degree $2n$ can be represented in spherical coordinates as a linear combination of terms like those in (4.3) and the Jacobian J is represented in a similar way (see (4.1)). On the other hand, quadratures (4.4) and (4.5) are exact for trigonometric polynomials of degree $k := 2(d-1)n + d - 2$. Therefore (see (4.2)), cubature (4.18) (or (4.19)) is exact for all spherical polynomials of degree $2n$; i.e., for every spherical polynomial S of degree $\leq 2n$ we have

$$(4.20) \qquad \mathbf{Q}_n(S) := \sum_{\omega \in \Omega_n} \lambda_\omega S(\omega) = \int_{\mathbf{S}^{d-1}} S(\xi)\, d\xi.$$

Note that $\lambda_\omega > 0$ and, since (4.20) holds for $S = 1$, then

$$(4.21) \qquad \sum_{\omega \in \Omega_n} \lambda_\omega = \int_{\mathbf{S}^{d-1}} 1\, d\xi =: |\mathbf{S}^{d-1}|.$$

Identity (4.20) implies discrete representations of the projection $Q_m(f)$ of any $f \in L_2(\mathbf{B}^d)$ onto $\mathcal{P}_m \ominus \mathcal{P}_{m-1}$ and $\|A_m(f)\|_{L_2(\mathbf{S}^{d-1})}$ (see (3.6) and (3.7) from Theorem 3.1). Namely, since $A_m(f,\xi)\mathcal{U}_m(\mathbf{x} \cdot \xi)$ and $A_m^2(f,\xi)$, for $m \leq n$, are spherical polynomials of degree $\leq 2m \leq 2n$, then

$$(4.22) \quad Q_m(f,\mathbf{x}) := \nu_m \int_{\mathbf{S}^{d-1}} A_m(f,\xi)\mathcal{U}_m(\mathbf{x} \cdot \xi)\, d\xi = \nu_m \sum_{\omega \in \Omega_n} \lambda_\omega A_m(f,\omega)\mathcal{U}_m(\mathbf{x} \cdot \omega)$$

and

$$(4.23) \qquad \|A_m\|_{L_2(\mathbf{S}^{d-1})}^2 := \int_{\mathbf{S}^{d-1}} |A_m(f,\xi)|^2\, d\xi = \sum_{\omega \in \Omega_n} \lambda_\omega |A_m(f,\omega)|^2.$$

Since quadratures (4.4) and (4.5) have good localization properties, then cubature (4.18) (or (4.19)) has such properties. We shall use them to prove the following lemma.

LEMMA 4.2. *Let $n = 1, 2, \ldots$, $m = 1, 2, \ldots$, and let, for $\gamma \in (0, \pi]$, $\mathcal{K}_m(\cos \gamma) := c_0 \min\{m^{d-1}, m^{d-1}/(m\gamma)^d\}$ with $c_0 > 0$ a constant. Then we have*

$$(4.24) \qquad \mathbf{Q}_n(\mathcal{K}_m(\bullet \cdot \eta)) \leq c[1 + (m/n)^{d-1}] \quad \text{for } \eta \in \mathbf{S}^{d-1},$$

where $\mathbf{Q}_n$ is the cubature from (4.19) *and c depends only on d and c_0.*

Proof. In what follows, we shall assume that $n > n_0$, where n_0 is sufficiently large and depends only on the dimension d. Estimate (4.24) obviously holds for $n \leq n_0$ by (4.21). We first construct a tiling of $\mathbf{S}^{d-1}$ which is determined by the nodes of cubature (4.18). We associate with each node $\omega_{\mathbf{j}}$ the spherical box (tile) $T_{\mathbf{j}}$ consisting of all points $\xi \in \mathbf{S}^{d-1}$ for which $\xi = \xi(\theta, \phi)$ with

$$(\theta, \phi) \in [a_{j_1}, a_{j_1+1}) \times \cdots \times [a_{j_{d-2}}, a_{j_{d-2}+1}) \times [b_{j_{d-1}}, b_{j_{d-1}+1}),$$

where $a_j := \frac{1}{2}(\beta_j + \beta_{j-1})$ and $b_j := \frac{1}{2}(\gamma_j + \gamma_{j-1})$ with β_j from (4.5) and γ_j from (4.4). Observe that $\omega_{\mathbf{j}} \in T_{\mathbf{j}}$ is the (spherical) center of $T_{\mathbf{j}}$. Obviously $T_{\mathbf{j}} \bigcap T_{\mathbf{i}} = \emptyset$, $\mathbf{j} \neq \mathbf{i}$, and the tiles $T_{\mathbf{j}}$ cover $\mathbf{S}^{d-1}$ excluding small regions around the poles. The most important property of our cubature is that

$$(4.25) \qquad 0 < \lambda_{\mathbf{j}} \leq c \int_{T_{\mathbf{j}}} 1\, d\xi =: c|T_{\mathbf{j}}| \quad \text{for } \mathbf{j} \in \mathcal{J}_n.$$

This property follows readily by (4.7), the definition of γ_j from (4.4), and the definition of our cubature (see (4.18)).

The second important property of our tiling is that the diameter of each tile $T_{\mathbf{j}}$ is $\leq cn^{-1}$. We let $\rho(\xi, \eta) := \arccos \xi \cdot \eta$, $\xi, \eta \in \mathbf{S}^{d-1}$ denote the angular distance on $\mathbf{S}^{d-1}$ (the angle between vectors ξ and η). It is easily seen that $\rho(\xi, \eta)$ satisfies the axioms for a distance on $\mathbf{S}^{d-1}$. Since $\beta_j - \beta_{j-1} \leq cn^{-1}$, by (4.6), and $\gamma_j - \gamma_{j-1} \leq cn^{-1}$, by the definition of γ_j, then

$$(4.26) \qquad \sup\{\rho(\xi, \eta) : \xi, \eta \in T_{\mathbf{j}}\} < c_1 n^{-1},$$

where c_1 depends only on d.

Suppose that $\eta \in \mathbf{S}^{d-1}$ is fixed. We select a new coordinate system such that $\eta = \mathbf{e}'_1 := (1, 0, \ldots, 0)$ is its first coordinate vector. This can be done by a suitable rotation of the old coordinate system. For $\xi \in \mathbf{S}^{d-1}$, we shall denote by $\theta' := (\theta'_1, \ldots, \theta'_{d-2})$ and ϕ' the new spherical coordinates of ξ.

We define, for $\nu = 1, 2, \ldots, n$,

$$\mathcal{Z}_\nu := \left\{ \xi \in \mathbf{S}^{d-1} : \frac{\pi(\nu-1)}{n} \leq \rho(\xi, \mathbf{e}'_1) \leq \frac{\pi\nu}{n} \right\}$$

$$= \left\{ \xi \in \mathbf{S}^{d-1} : \frac{\pi(\nu-1)}{n} \leq \theta'_1 \leq \frac{\pi\nu}{n} \right\}$$

and

$$\mathcal{Z}_\nu^* := \left\{ \xi \in \mathbf{S}^{d-1} : \max\left\{ \frac{\pi(\nu-1) - c_1}{n}, -\pi \right\} \leq \theta'_1 \leq \min\left\{ \frac{\pi\nu + c_1}{n}, \pi \right\} \right\},$$

where c_1 is from (4.26). Obviously $\bigcup_{\nu=1}^n \mathcal{Z}_\nu = \mathbf{S}^{d-1}$.

Let $\mathcal{T}_\nu$ be the set of all tiles $T_{\mathbf{j}}$ with centers $\omega_{\mathbf{j}} \in \mathcal{Z}_\nu$. It follows by (4.26) that $\bigcup_{T \in \mathcal{T}_\nu} T \subset \mathcal{Z}_\nu^*$ and hence

$$(4.27) \qquad \sum_{T \in \mathcal{T}_\nu} |T| \leq |\mathcal{Z}_\nu^*| := \int_{\mathcal{Z}_\nu^*} 1 \, d\xi \leq c \int_{\mathcal{Z}_\nu} 1 \, d\xi =: c|\mathcal{Z}_\nu|, \quad \nu = 1, 2, \ldots, n.$$

We are now ready to estimate $\mathbf{Q}_n(\mathcal{K}_m(\bullet \cdot \eta))$. If $\nu = 1$, then we obtain, using (4.25), (4.27), and the assumptions of the lemma,

$$\sum_{\omega_{\mathbf{j}} \in \mathcal{Z}_1} \lambda_{\mathbf{j}} \mathcal{K}_m(\omega_{\mathbf{j}} \cdot \mathbf{e}'_1) \leq c \max\{\mathcal{K}_m(\cos\theta_1) : 0 \leq \theta'_1 \leq \pi/n\} \sum_{T \in \mathcal{T}_1} |T|$$

$$\leq cm^{d-1} |\mathcal{Z}_1| \leq cm^{d-1} \int_0^{\pi/n} \sin^{d-2}\theta'_1 \, d\theta'_1 \leq c(m/n)^{d-1}.$$

If $\nu \geq 2$, then

$$\sum_{\omega_{\mathbf{j}} \in \mathcal{Z}_\nu} \lambda_{\mathbf{j}} \mathcal{K}_m(\omega_{\mathbf{j}} \cdot \mathbf{e}'_1) \leq c \max\{\mathcal{K}_m(\cos\theta'_1) : \pi(\nu-1)/n \leq \theta'_1 \leq \pi\nu/n\} \sum_{T \in \mathcal{T}_\nu} |T|$$

$$\leq c\mathcal{K}_m\left(\cos\frac{\pi(\nu-1)}{n}\right) |\mathcal{Z}_\nu^*| \leq c\mathcal{K}_m\left(\cos\frac{\pi\nu}{n}\right) |\mathcal{Z}_\nu| \leq c \int_{\mathcal{Z}_\nu} \mathcal{K}_m(\xi \cdot \mathbf{e}'_1) \, d\xi,$$

where we used that $\mathcal{K}_m(\cos\frac{\pi(\nu-1)}{n}) \leq c\mathcal{K}_m\left(\cos\frac{\pi\nu}{n}\right)$, $\nu \geq 2$, which follows by the definition of $\mathcal{K}_m(\cos\gamma)$ from the assumptions of the lemma. Therefore,

$$(4.28) \qquad \sum_{\mathbf{j} \in \mathcal{J}_n} \lambda_{\mathbf{j}} \mathcal{K}_m(\omega_{\mathbf{j}} \cdot \mathbf{e}'_1) \leq c(m/n)^{d-1} + c \int_{\mathcal{Z}} \mathcal{K}_m(\xi \cdot \mathbf{e}'_1) \, d\xi,$$

where $\mathcal{Z} := \bigcup_{\nu=2}^{n} \mathcal{Z}_\nu$. We obtain, using again the definition of $\mathcal{K}_m(\cos\gamma)$,

$$
\int_{\mathcal{Z}} \mathcal{K}_m(\xi \cdot \mathbf{e}_1')\, d\xi = |\mathbf{S}^{d-2}| \int_{\pi/n}^{\pi} \mathcal{K}_m(\cos\theta_1') \sin^{d-2}\theta_1'\, d\theta_1'
$$

$$
\leq c \int_0^\infty \min\{m^{d-1}, m^{d-1}/(m\theta_1')^d\}(\theta_1')^{d-2}\, d\theta_1' \leq c < \infty.
$$

The above estimates and (4.28) imply (4.24).   □

We shall deal with discrete sums of spherical polynomial values. For this, we need a rapidly decaying reproducing kernel for the space of spherical polynomials of degree m. The following well-known proposition gives us such a kernel.

PROPOSITION 4.1.   *There exists a constant $m_0 = m_0(d)$ such that for every $m \geq m_0$ there exists an algebraic polynomial W_m of degree dm with the properties:*
(a)

$$
S(\eta) = \int_{\mathbf{S}^{d-1}} W_m(\eta \cdot \xi) S(\xi)\, d\xi, \quad \eta \in \mathbf{S}^{d-1},
$$

for each spherical polynomial S of degree $\leq m$;
(b)

(4.29) $|W_m(\cos\tau)| \leq c_0 \min\{m^{d-1}, m^{d-1}/(m\tau)^d\} \quad for \quad 0 < \tau \leq \pi,$

and hence
(c)

(4.30) $$\int_{\mathbf{S}^{d-1}} |W_m(\eta \cdot \xi)|\, d\xi \leq c < \infty, \quad \eta \in \mathbf{S}^{d-1},$$

where c_0 and c are independent of m and η.

Since we do not have a good reference for Proposition 4.1, we shall show how it can be deduced from the following results of Kogbetliantz and Stein (see also [P]).

PROPOSITION 4.2 (see [K]).   *Let $S_m(t) := \sum_{\nu=0}^{m}(\nu + \lambda)C_\nu^\lambda(t)$, $\lambda > 0$, $m = 0, 1, \ldots$, and let $\sigma_m^{(\delta)}$ be the Cesàro means of order δ of S_m; i.e.,*

(4.31) $$\sigma_m^{(\delta)}(t) := \left(A_m^\delta\right)^{-1} \sum_{\nu=0}^{m} A_{m-\nu}^\delta (\nu + \lambda) C_\nu^\lambda(t) \quad with \quad A_\nu^\delta := \frac{\Gamma(\nu+\delta+1)}{\Gamma(\delta+1)\Gamma(\nu+1)}.$$

Then, for $-1 < \delta \leq 2\lambda + 1$,

(4.32) $$|\sigma_m^{(\delta)}(\cos\gamma)| \leq c \min\left\{ (m+1)^{2\lambda+1}, (m+1)^{2\lambda-\delta} \Big/ \left(\sin\frac{\gamma}{2}\right)^{\delta+1}\right\}, \; 0 < \gamma \leq \pi,$$

with c depending only on λ.

PROPOSITION 4.3 (see [St]).   *For each positive integer r and for $m = 0, 1, \ldots$, there exist $r + 1$ parameters $\alpha_1(m), \ldots, \alpha_{r+1}(m)$ (depending only on m and r) which are uniformly bounded: $|\alpha_\nu(m)| \leq A$, A independent of m, and there exists a fixed integer N, so that the following holds:*

If $\sum_{\nu=0}^{\infty} a_\nu$ is a series of real numbers and if $\sigma_m^{(r)}$, $m = 0, 1, \ldots$, are the Cesàro means of order r of the partial sums S_m, $m = 0, 1, \ldots$, of this series (see (4.31)), then

$$
\tau_m^{(r)} := \alpha_1(m)\sigma_{m-1}^{(r)} + \alpha_2(m)\sigma_{2m-1}^{(r)} + \cdots + \alpha_{r+1}(m)\sigma_{(r+1)m-1}^{(r)}
$$

can be represented in the form

$$\tau_m^{(r)} = \sum_{\nu=0}^{m} a_\nu + \sum_{\nu=m+1}^{(r+1)m} \beta_\nu a_\nu \quad \text{if } m \geq N,$$

where β_ν are constants depending on m and r.

Proof of Proposition 4.1. We have already mentioned in (3.15) that

$$K_m(t) := \frac{N(d,m)}{|\mathbf{S}^{d-1}|C_m^{(d-2)/2}(1)} C_m^{(d-2)/2}(t)$$

gives the reproducing kernel $K_m(\xi \cdot \eta)$ for $\mathcal{H}_m$. Therefore, $\sum_{\nu=0}^{m} K_\nu(\xi \cdot \eta)$ is a reproducing kernel for all spherical polynomials of degree $\leq m$. Simple calculations show that

$$K_m(t) = 2 \left[|\mathbf{S}^{d-1}|(d-2) \right]^{-1} (m+\lambda) C_m^\lambda(t) \quad \text{with } \lambda := (d-2)/2.$$

Therefore, $2 \left[|\mathbf{S}^{d-1}|(d-2) \right]^{-1} \sum_{\nu=0}^{m} (\nu+\lambda) C_\nu^\lambda(t)$ gives a reproducing kernel for the spherical polynomials of degree $\leq m$.

We now apply Proposition 4.2 with $\lambda := (d-2)/2$ and $\delta := 2\lambda+1 = d-1$. Then we apply Proposition 4.3 to the resulting Cesàro means $\{\sigma_\nu^{(r)}\}$ with $r := \delta = d-1$ to conclude that $W_m := 2 \left[|\mathbf{S}^{d-1}|(d-2) \right]^{-1} \tau_m^{(r)}$ satisfies (4.29) (by (4.32) and since $\alpha_\nu(m)$ are uniformly bounded) and $W_m(\xi \cdot \eta)$ is a reproducing kernel for the spherical polynomials of degree $\leq m$ (by Proposition 4.3). $\square$

Lemma 4.2 and Proposition 4.1 allow us to estimate discrete $l_p(\Omega_n)$ norms of spherical polynomials by their $L_p(\mathbf{S}^{d-1})$ norms. In this part we use ideas from [O].

LEMMA 4.3. *Let $n = 1, 2, \ldots$, and let $m \geq m_0$, where m_0 is from Proposition 4.1. Then for every spherical polynomial S of degree m and for $1 \leq p \leq \infty$ we have*

$$(4.33) \qquad \sum_{\omega \in \Omega_n} \lambda_\omega |S(\omega)|^p \leq c[1 + (m/n)^{d-1}] \int_{\mathbf{S}^{d-1}} |S(\xi)|^p \, d\xi,$$

where λ_ω and Ω_n are from (4.19), and c is independent of S, n, and m.

Proof. By Proposition 4.1 we get $S(\omega) = \int_{\mathbf{S}^{d-1}} W_m(\omega \cdot \xi) S(\xi) \, d\xi$, $\omega \in \Omega_n$. We obtain, using Hölder's inequality,

$$|S(\omega)| \leq \int_{\mathbf{S}^{d-1}} |W_m(\omega \cdot \xi) S(\xi)| \, d\xi = \int_{\mathbf{S}^{d-1}} |W_m(\eta \cdot \xi)|^{1-1/p} |W_m(\omega \cdot \xi)|^{1/p} |S(\xi)| \, d\xi$$

$$\leq \left(\int_{\mathbf{S}^{d-1}} |W_m(\omega \cdot \xi)| \, d\xi \right)^{1-1/p} \left(\int_{\mathbf{S}^{d-1}} |W_m(\omega \cdot \xi)| |S(\xi)|^p \, d\xi \right)^{1/p}$$

and hence

$$|S(\omega)|^p \leq A^{p-1} \int_{\mathbf{S}^{d-1}} |W_m(\omega \cdot \xi)| |S(\xi)|^p \, d\xi, \quad \text{where } A := \int_{\mathbf{S}^{d-1}} |W_m(\omega \cdot \xi)| \, d\xi.$$

We now multiply both sides of the above inequality by λ_ω and sum over $\omega \in \Omega_n$ to obtain

$$\sum_{\omega \in \Omega_n} \lambda_\omega |S(\omega)|^p \leq A^{p-1} \int_{\mathbf{S}^{d-1}} \left(\sum_{\omega \in \Omega_n} \lambda_\omega |W_m(\omega \cdot \xi)| \right) |S(\xi)|^p \, d\xi$$

$$\leq A^{p-1} \max_{\xi \in \mathbf{S}^{d-1}} \mathbf{Q}_n(|W_m(\bullet \cdot \xi)|) \int_{\mathbf{S}^{d-1}} |S(\xi)|^p \, d\xi.$$

It follows, by (4.29), that $|W_m(\bullet \cdot \xi)| \leq \mathcal{K}_m(\bullet \cdot \xi)$, where $\mathcal{K}_m$ is defined in Lemma 4.2 with c_0 from Proposition 4.1. Then Proposition 4.1 and Lemma 4.2 imply

$$\max_{\xi \in \mathbf{S}^{d-1}} \mathbf{Q}_n(|W_m(\bullet \cdot \xi)|) \leq \max_{\xi \in \mathbf{S}^{d-1}} \mathbf{Q}_n(\mathcal{K}_m(\bullet \cdot \xi)) \leq c[1 + (m/n)^{d-1}] \quad \text{and} \quad A \leq c$$

which completes the proof of Lemma 4.3. $\square$

The following lemma relates the $L_2(\mathbf{S}^{d-1})$ norms and discrete $l_2(\Omega_n)$ norms of spherical polynomials written in terms of $\nu_m \mathcal{U}_m(\xi \cdot \omega)/\mathcal{U}_m(1)$, the reproducing kernel for the space $\mathcal{H}_m \oplus \mathcal{H}_{m-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$ (see (3.17)).

LEMMA 4.4. *Let* $n = 1, 2, \ldots,$ *and let* $c(\omega)$, $\omega \in \Omega_n$, *be real constants. Let* $m \geq m_0$, *where* m_0 *is from Proposition* 4.1. *Then the spherical polynomial*

$$S(\xi) := \sum_{\omega \in \Omega_n} \lambda_\omega c(\omega) \frac{\nu_m}{\mathcal{U}_m(1)} \mathcal{U}_m(\xi \cdot \omega)$$

satisfies

(4.34) $$\|S\|_{L_2(\mathbf{S}^{d-1})}^2 \leq c[1 + (m/n)^{d-1}] \sum_{\omega \in \Omega_n} \lambda_\omega |c(\omega)|^2.$$

Proof. Using (3.11) we get

$$\|S\|_{L_2(\mathbf{S}^{d-1})}^2 = \int_{\mathbf{S}^{d-1}} |S(\xi)|^2 \, d\xi$$

$$= \sum_{\omega \in \Omega_n} \sum_{\eta \in \Omega_n} \lambda_\omega \lambda_\eta c(\omega) c(\eta) \left(\frac{\nu_m}{\mathcal{U}_m(1)}\right)^2 \int_{\mathbf{S}^{d-1}} \mathcal{U}_m(\xi \cdot \omega) \mathcal{U}_m(\xi \cdot \eta) \, d\xi$$

$$= \sum_{\omega \in \Omega_n} \sum_{\eta \in \Omega_n} \lambda_\omega \lambda_\eta c(\omega) c(\eta) \frac{\nu_m}{\mathcal{U}_m(1)} \mathcal{U}_m(\omega \cdot \eta) = \sum_{\eta \in \Omega_n} \lambda_\eta c(\eta) S(\eta)$$

$$\leq \left(\sum_{\eta \in \Omega_n} \lambda_\eta |c(\eta)|^2\right)^{1/2} \left(\sum_{\eta \in \Omega_n} \lambda_\eta |S(\eta)|^2\right)^{1/2}.$$

By Lemma 4.3, the last quantity above does not exceed $c[1 + (m/n)^{d-1}]^{\frac{1}{2}} \|S\|_{L_2(\mathbf{S}^{d-1})}$. Finally, we divide by $\|S\|_{L_2(\mathbf{S}^{d-1})}$ to complete the proof of the lemma. $\square$

5. Smoothness spaces in $L_2(\mathbf{B}^d)$. In this section, we shall recall results about approximation by algebraic polynomials. As earlier, we let $\mathcal{P}_n$ denote the space of algebraic polynomials in d-variables. For $n \geq 1$, let

$$E_n(f) := E_n(f)_{L_2(\mathbf{B}^d)} := \inf_{P \in \mathcal{P}_n} \|f - P\|_{L_2(\mathbf{B}^d)}$$

be the error in approximating $f \in L_2(\mathbf{B}^d)$ by algebraic polynomials P of degree $\leq n$. By Theorem 3.1 we have the following representation of the polynomial $P_n(f, \mathbf{x})$ of best $L_2(\mathbf{B}^d)$-approximation to f:

(5.1) $$P_n(f, \mathbf{x}) = \sum_{m=0}^{n} \nu_m \int_{\mathbf{S}^{d-1}} A_m(\xi) \mathcal{U}_m(\mathbf{x} \cdot \xi) \, d\xi,$$

where

$$A_n(\xi) := A_n(f, \xi) := \int_{\mathbf{B}^d} f(\mathbf{y}) \mathcal{U}_m(\mathbf{y} \cdot \xi) \, d\mathbf{y}.$$

Since $A_m(\xi)U_m(\mathbf{x} \cdot \xi)$ is a spherical polynomial of degree $\leq 2m \leq 2n$ in ξ, we can use the quadrature formula (4.19) to obtain

$$(5.2) \qquad P_n(f, \mathbf{x}) = \sum_{\omega \in \Omega_n} \lambda_\omega \sum_{m=0}^n \nu_m A_m(\omega)\mathcal{U}_m(\mathbf{x} \cdot \omega).$$

From Theorem 3.1, we have

$$E_n(f)^2 = \|f - P_n(f)\|^2_{L_2(\mathbf{B}^d)} = \sum_{m>n} \nu_m \|A_m(f)\|^2_{L_2(\mathbf{S}^{d-1})}$$

$$(5.3) \qquad \asymp \sum_{m>n} m^{d-1} \|A_m(f)\|^2_{L_2(\mathbf{S}^{d-1})}.$$

For $\alpha > 0$, let $W^\alpha(L_2(\mathbf{B}^d))$ be the Sobolev space for the domain $\mathbf{B}^d$. When $\alpha = k$ is an integer, then a function $f \in L_2(\mathbf{B}^d)$ is in $W^k(L_2(\mathbf{B}^d))$ if and only if its distributional derivatives $D^\nu f$ of order k are in $L_2(\mathbf{B}^d)$, and

$$|f|^2_{W^k(L_2(\mathbf{B}^d))} := \sum_{|\nu|=k} \|D^\nu f\|^2_{L_2(\mathbf{B}^d)}$$

gives the seminorm for $W^k(L_2(\mathbf{B}^d))$. The norm for $W^k(L_2(\mathbf{B}^d))$ is obtained by adding $\|f\|_{L_2(\mathbf{B}^d)}$ to $|f|_{W^k(L_2(\mathbf{B}^d))}$. For other values of α, we obtain W^α as the interpolation space

$$W^\alpha(L_2(\mathbf{B}^d)) = (L_2(\mathbf{B}^d), W^k(L_2(\mathbf{B}^d)))_{\theta,2}, \quad \theta = \alpha/k, \ 0 < \alpha < k,$$

given by the real method of interpolation (see, e.g., Bennett and Sharpley [BS]).

A fundamental result in approximation known as the Jackson theorem states that

$$(5.4) \qquad E_n(f) \leq c(k)n^{-k}\|f\|_{W^k(L_2(\mathbf{B}^d))},$$

where the norm on the right can be replaced by the seminorm if k is an integer. This theorem can be deduced easily from the results on univariate approximation in Chapter 7 of [DL]. By interpolation (see, e.g., [DL, Chapter 7]), one obtains

$$(5.5) \qquad \sum_{n=1}^\infty [n^\alpha E_n(f)]^2 n^{-1} \leq c(\alpha)\|f\|^2_{W^\alpha(L_2(\mathbf{B}^d))}, \quad \alpha > 0,$$

with $c(\alpha)$ depending at most on α. From (5.3) and (5.5), it is easy to deduce that

$$(5.6) \qquad \sum_{n=1}^\infty n^{2\alpha+d-1}\|A_n(f)\|^2_{L_2(\mathbf{S}^{d-1})} \leq c(\alpha)\|f\|^2_{W^\alpha(L_2(\mathbf{B}^d))}, \quad \alpha > 0,$$

with $c(\alpha)$ depending at most on α.

6. Approximation of functions in $L_2(I, w)$. We shall also need certain results about the approximation of univariate functions in $L_2(I, w)$ where $I := [-1, 1]$ and $w := w_{d/2}$. As we know by section 2, the Gegenbauer polynomials $\{\mathcal{U}_m\}_{m=0}^\infty$ form a complete orthonormal system for $L_2(I, w)$ (see (3.3)). For any $g \in L_2(I, w)$ we have

$$(6.1) \qquad g = \sum_{m=0}^n \hat{g}(m)\mathcal{U}_m \quad \text{with} \quad \hat{g}(m) := \int_I g(s)\mathcal{U}_m(s)w(s)\,ds.$$

We shall use approximation of functions in $L_2(I, w)$ as an intermediate tool in establishing our results on ridge approximation. Let $\mathcal{P}_n(I)$ denote the space of univariate algebraic polynomials of degree $\leq n$. For a function $g \in L_2(I, w)$, we let

$$E_n(g)_{L_2(I,w)} := \inf_{p \in \mathcal{P}_n(I)} \|g - p\|_{L_2(I,w)}$$

be the error in approximating g by the elements of $\mathcal{P}_n(I)$. The polynomial

$$(6.2) \qquad p_n := \sum_{m=0}^{n} \hat{g}(m)\mathcal{U}_m$$

is the best $L_2(I, w)$ approximation to g by elements of $\mathcal{P}_n(I)$, and we have

$$(6.3) \qquad E_n(g)_{L_2(I,w)}^2 = \|g - p_n\|_{L_2(I,w)}^2 = \sum_{m>n} |\hat{g}(m)|^2.$$

We introduce the univariate Sobolev spaces $W^\alpha(L_2(I, w))$, $\alpha \in \mathbf{R}$, whose norms are defined by

$$(6.4) \qquad \|g\|_{W^\alpha(L_2(I,w))}^2 := \sum_{m=0}^{\infty} [(m+1)^\alpha |\hat{g}(m)|]^2.$$

It follows that for each $g \in W^\alpha(L_2(I, w))$,

$$(6.5) \qquad E_n(g)_{L_2(I,w)} \leq c(\alpha)n^{-\alpha}\|g\|_{W^\alpha(L_2(I,w))}.$$

Moreover, similar to (5.5), we have

$$(6.6) \qquad \sum_{n=1}^{\infty} [n^\alpha E_n(g)_{L_2(I,w)}]^2 n^{-1} \leq c(\alpha)\|g\|_{W^\alpha(L_2(I,w))}^2, \quad \alpha > 0.$$

There is also a Bernstein-type inequality for polynomials in $\mathcal{P}_n(I)$ with respect to $L_2(I, w)$ which follows trivially from the definition: for every $p \in \mathcal{P}_n(I)$ and $\alpha > 0$,

$$(6.7) \qquad \|p\|_{W^\alpha(L_2(I,w))} \leq (n+1)^\alpha \|p\|_{L_2(I,w)}.$$

It is well known (see [DL, Chapter 7]) that companion inequalities like (6.5) and (6.7) imply a characterization of approximation spaces by interpolation spaces. In our context, the approximation spaces are the Sobolev spaces $W^\alpha(L_2(I, w))$ defined by (6.4) and we therefore obtain for each $0 < \alpha < k$,

$$(6.8) \qquad W^\alpha(L_2(I, w)) = (L_2(I, w), W^k(L_2(I, w)))_{\theta,2}, \quad \theta = \alpha/k.$$

Further properties of the spaces $W^\alpha(L_2(I, w))$ are given in section 7.

7. Approximation by ridge functions. In this section, we assume that X_n is a subspace of $L_2(I, w)$, $w = w_{d/2}$, of dimension n with the following property. There is a real number $s > 0$ such that, for each univariate function $g \in W^s(L_2(I, w))$, there is a function $r \in X_n$ which provides the Jackson estimate

$$(7.1) \qquad \|g - r\|_{L_2(I,w)} \leq c_0 n^{-s}\|g\|_{W^s(L_2(I,w))},$$

with c_0 a constant independent of g and n.

We define Y_n to be the space of functions R in d variables of the form

$$(7.2) \qquad R(\mathbf{x}) = \sum_{\omega \in \Omega_n} r_\omega(\mathbf{x} \cdot \omega), \quad r_\omega \in X_n, \quad \omega \in \Omega_n,$$

where Ω_n is the set of vectors in $\mathbf{S}^{d-1}$ from (4.19). Then Y_n is a linear space of dimension $\leq n\#\Omega_n \leq cn^d$. We prove the following theorem about approximation from Y_n.

THEOREM 7.1. *Let X_n, $n = 1, 2, \ldots$, satisfy inequality (7.1) for some $s > 0$. If f is a function from the space $W^{s+\frac{d-1}{2}}(L_2(\mathbf{B}^d))$, then there is a function R in Y_n such that*

$$(7.3) \qquad \|f - R\|_{L_2(\mathbf{B}^d)} \leq cn^{-s-\frac{d-1}{2}}\|f\|_{W^{s+\frac{d-1}{2}}(L_2(\mathbf{B}^d))}$$

with c a constant depending only on s and d.

REMARK 7.1. *If $s + (d-1)/2 - 1$ is an integer and the space Y_n contains $\mathcal{P}_{s+(d-1)/2-1}$, then we have that $\|f\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))}$ can be replaced by the seminorm $|f|_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))}$.*

An important element of the proof of Theorem 7.1 is the idea to get rid of the "low frequencies" when approximating. To this end we shall use the following geometric construction which was proven for us by Boris Kashin.

LEMMA 7.1. *Let H be a Hilbert space with norm $\|\cdot\|$ and let A, $B \subset H$ be finite-dimensional linear subspaces of H with $\dim A \leq \dim B$. If there exists δ, $0 < \delta < 1/2$, such that*

$$(7.4) \qquad \sup_{\substack{x \in A \\ \|x\| \leq 1}} \inf_{y \in B} \|x - y\| \leq \delta,$$

then there is a constant c depending only on δ and a linear operator $L : A \to B$ such that for every $x \in A$,

$$(7.5) \qquad \|Lx - x\| \leq c \inf_{y \in B} \|x - y\|,$$

and

$$Lx - x \perp A \quad (Lx - x \text{ is orthogonal to } A).$$

Proof. See [DOP, Lemma 6]. □

Proof of Theorem 7.1. Estimate (7.3) trivially holds if $n < m_0$, where $m_0 = m_0(d)$ is the constant from Proposition 4.1.

Suppose that $n \geq m_0$. Let $P = P_n$ be the polynomial in $\mathcal{P}_n$ given by (5.1) (or (5.2)). Since P is the best $L_2(\mathbf{B}^d)$ approximation of f, it satisfies (see (5.4))

$$(7.6) \qquad \|f - P\|_{L_2(\mathbf{B}^d)} \leq cn^{-s-(d-1)/2}\|f\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))}$$

with c and all subsequent constants in this proof depending only on s and d. We shall approximate P by an element R of Y_N, $N = k_0 n$, where k_0 is a sufficiently large constant depending only on s and d.

We have $A_m(P, \xi) = A_m(f, \xi)$, $m \leq n$, and $A_m(P, \xi) = 0$, $m > n$. Since $f \in W^{s+(d-1)/2}(L_2(\mathbf{B}^d))$, we know from (5.6) that

$$(7.7) \qquad \sum_{m=0}^{n}(m+1)^{2s+2(d-1)}\|A_m(f)\|^2_{L_2(\mathbf{S}^{d-1})} \leq c\|f\|^2_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))}.$$

From this, using (4.23), we obtain

$$
(7.8) \qquad \sum_{m=0}^{n} (m+1)^{2s+2(d-1)} \sum_{\omega \in \Omega_n} \lambda_\omega |(A_m(f,\omega))|^2 \leq c \|f\|^2_{W^{r+(d-1)/2}(L_2(\mathbf{B}^d))}.
$$

We introduce the univariate polynomials

$$
(7.9) \qquad p_\omega(t) := \sum_{m=0}^{n} \nu_m A_m(f,\omega) \mathcal{U}_m(t) = \sum_{m=0}^{n} \nu_m A_m(P,\omega) \mathcal{U}_m(t), \quad \omega \in \Omega_n.
$$

We have, by (4.22) and (7.9),

$$
(7.10) \qquad P(\mathbf{x}) = \sum_{m=0}^{n} \nu_m \sum_{\omega \in \Omega_n} \lambda_\omega A_m(P,\omega) \mathcal{U}_m(\mathbf{x} \cdot \omega) = \sum_{\omega \in \Omega_n} \lambda_\omega p_\omega(\mathbf{x} \cdot \omega).
$$

According to (6.4), we have

$$
\begin{aligned}
\|\lambda_\omega^{1/2} p_\omega\|^2_{W^s(L_2(I,w))} &= \sum_{m=0}^{n} (m+1)^{2s} \nu_m^2 \lambda_\omega |A_m(f,\omega)|^2 \\
&\asymp \sum_{m=0}^{n} (m+1)^{2s+2(d-1)} \lambda_\omega |A_m(f,\omega)|^2.
\end{aligned}
$$

Hence, from (7.8),

$$
\begin{aligned}
\sum_{\omega \in \Omega_n} \|\lambda_\omega^{1/2} p_\omega\|^2_{W^s(L_2(I,w))} &\leq c \sum_{m=0}^{n} (m+1)^{2s+2(d-1)} \sum_{\omega \in \Omega_n} \lambda_\omega |(A_m(f,\omega))|^2 \\
(7.11) &\leq c \|f\|^2_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))}.
\end{aligned}
$$

We shall approximate each polynomial p_ω by elements of X_N. We apply Lemma 7.1 in the following setting. We take for H the Hilbert space $L_2(I,w)$ and take $A = \mathcal{P}_n(I)$ and $B = X_N$ with $N \geq k_0 n$ and k_0 a positive integer. We next show that if k_0 is large enough then the assumption (7.4) is satisfied. We mentioned earlier in (6.7) that $\mathcal{P}_n(I)$ satisfies the Bernstein inequality

$$
\|p\|_{W^s(L_2(I,w))} \leq (n+1)^s \|p\|_{L_2(I,w)}, \quad p \in \mathcal{P}_n(I).
$$

If $p \in \mathcal{P}_n(I)$, then from this Bernstein inequality and from (7.1), there is an $r \in X_N$ such that

$$
\|p-r\|_{L_2(I,w)} \leq c_0 N^{-s} \|p\|_{W^s(L_2(I,w))} \leq c_0 N^{-s} 2^s n^s \|p\|_{L_2(I,w)} \leq c_0 2^s k_0^{-s} \|p\|_{L_2(I,w)}.
$$

Thus, if k_0 is large enough, condition (7.4) is satisfied. Therefore, for each $\omega \in \Omega_n$, we can find $r_\omega \in X_N$ such that $r_\omega - p_\omega \perp \mathcal{P}_n(I)$ with respect to the inner product in $L_2(I,w)$ and, by (7.1) and (7.5),

$$
\|p_\omega - r_\omega\|^2_{L_2(I,w)} \leq c n^{-2s} \|p_\omega\|^2_{W^s(L_2(I,w))}.
$$

Therefore,

$$
(7.12) \qquad r_\omega - p_\omega = \sum_{m=n+1}^{\infty} \hat{r}_\omega(m) \mathcal{U}_m
$$

with

$$\hat{r}_\omega(m) := \int_I r_\omega(s) \mathcal{U}_m(s) w(s) \, ds$$

and

(7.13) $$\sum_{m=n+1}^\infty |\hat{r}_\omega(m)|^2 = \|p_w - r_\omega\|^2_{L_2(I,w)} \le cn^{-2s} \|p_\omega\|^2_{W^s(L_2(I,w))}.$$

We define

$$R(\mathbf{x}) := \sum_{\omega \in \Omega_n} \lambda_\omega r_\omega(\mathbf{x} \cdot \omega)$$

which is an element of Y_N. Then we have, by (7.10) and (7.12),

$$R(\mathbf{x}) - P(\mathbf{x}) = \sum_{\omega \in \Omega_n} \sum_{m=n+1}^\infty \lambda_\omega \hat{r}_\omega(m) \mathcal{U}_m(\omega \cdot \mathbf{x}) = \sum_{m=n+1}^\infty \sum_{\omega \in \Omega_n} \lambda_\omega \hat{r}_\omega(m) \mathcal{U}_m(\omega \cdot \mathbf{x}).$$

We write

$$R_m(\mathbf{x}) := \sum_{\omega \in \Omega_n} \lambda_\omega \hat{r}_\omega(m) \mathcal{U}_m(\mathbf{x} \cdot \omega).$$

We have by Theorem 3.1 (see also (3.19)–(3.21))

$$R_m(\mathbf{x}) = \nu_m \int_{\mathbf{S}^{d-1}} A_m(R_m, \xi) \mathcal{U}_m(\xi \cdot \mathbf{x}) \, d\xi,$$

where

$$A_m(R_m, \xi) = \int_{\mathbf{B}^d} R_m(\mathbf{y}) \mathcal{U}_m(\mathbf{y} \cdot \xi) \, d\mathbf{y} = \sum_{\omega \in \Omega_n} \lambda_\omega \hat{r}_\omega(m) \int_{\mathbf{B}^d} \mathcal{U}_m(\omega \cdot \mathbf{y}) \mathcal{U}_m(\xi \cdot \mathbf{y}) \, d\mathbf{y}$$

$$= \sum_{\omega \in \Omega_n} \lambda_\omega \hat{r}_\omega(m) \frac{\mathcal{U}_m(\xi \cdot \omega)}{\mathcal{U}_m(1)}.$$

We now use Theorem 3.1 and Lemma 4.4 to obtain

$$\|R_m\|^2_{L_2(\mathbf{B}^d)} = \nu_m \|A_m(R_m, \omega)\|^2_{L_2(\mathbf{S}^{d-1})} = \nu_m^{-1} \left\| \sum_{\omega \in \Omega_n} \lambda_\omega \hat{r}_\omega(m) \frac{\nu_m}{\mathcal{U}_m(1)} \mathcal{U}_m(\xi \cdot \omega) \right\|^2_{L_2(\mathbf{S}^{d-1})}$$

$$\le c\nu_m^{-1}(m/n)^{d-1} \sum_{\omega \in \Omega_n} \lambda_\omega |\hat{r}_\omega(m)|^2 \le cn^{-d+1} \sum_{\omega \in \Omega_n} \lambda_\omega |\hat{r}_\omega(m)|^2,$$

where we used that $\nu_m \asymp m^{d-1}$ (see (3.8)). From this, (7.11), and (7.13), we find, using the Parseval identity (3.9),

$$\|R - P\|^2_{L_2(\mathbf{B}^d)} = \sum_{m=n+1}^\infty \|R_m\|^2_{L_2(\mathbf{B}^d)} \le cn^{-d+1} \sum_{m=n+1}^\infty \sum_{\omega \in \Omega_n} \lambda_\omega |\hat{r}_\omega(m)|^2$$

$$= cn^{-d+1} \sum_{\omega \in \Omega_n} \lambda_\omega \sum_{m=n+1}^\infty |\hat{r}_\omega(m)|^2$$

$$\le cn^{-2s-d+1} \sum_{\omega \in \Omega_n} \|\lambda_\omega^{1/2} p_\omega\|^2_{W^s(L_2(I,w))}$$

$$\le cn^{-2s-d+1} \|f\|^2_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))}.$$

Thus there is a function $R \in Y_N$, $N = k_0 n$, such that (7.3) holds. Theorem 7.1 is now proved. $\square$

REMARK 7.2. *As in* [DOP], *it is possible to prove Theorem* 7.1 *without using Lemma* 7.1. *In place of this lemma one uses a slightly stronger assumption than estimate* (7.1). *The corresponding proof would be more constructive than the present one. We do not provide the details of this approach but instead refer the reader to* [DOP].

8. Elimination of the weight w. The result of section 7 (Theorem 7.1) gives sufficient conditions on a sequence of univariate spaces X_n, $n = 1, 2, \ldots$, in order that the spaces Y_n defined by (7.2) with Ω_n from (4.19) provide approximation rates for functions in Sobolev spaces $W^\alpha(L_2(\mathbf{B}^d))$ comparable with polynomials and splines. However, the assumption (7.1) imposed on X_n is inconvenient for direct application because of the appearance of the weight $w(t) := w_{d/2}(t) := (1 - t^2)^{(d-1)/2}$. We shall show in this section how the weight factor w can be avoided so that the result of section 7 applies more directly. We shall consider approximation on the ball $\mathbf{B}^d_{1/2} := \{\mathbf{x} \in \mathbf{R}^d : |\mathbf{x}| \leq 1/2\}$ rather than $\mathbf{B}^d$. Approximation on $\mathbf{B}^d$ or other balls follows by a change of variables.

We begin by assuming that we have in hand n-dimensional linear spaces Z_n of univariate functions defined on $J := [-1/2, 1/2]$ which satisfy a Jackson-type estimate similar to (7.1) but with weight $= 1$. Let $W^m(L_2(J))$, $m = 1, 2, \ldots$, be the Sobolev space of functions $g \in L_2(J)$ such that $g^{(m)}$ is in $L_2(J)$. The seminorm and norm for $W^m(L_2(J))$ are defined by

$$|g|_{W^m(L_2(J))} := \|g^{(m)}\|_{L_2(J)} \quad ; \quad \|g\|_{W^m(L_2(J))} := \|g^{(m)}\|_{L_2(J)} + \|g\|_{L_2(J)}.$$

For $0 < s < m$ not an integer, we define $W^s(L_2(J))$ by interpolation:

$$(8.1) \qquad W^s(L_2(J)) := (L_2(J), W^m(L_2(J)))_{\theta,2}, \quad \theta := s/m,$$

with the norm as the interpolation space norm. For a given value of s, different values of $m > s$ give equivalent norms (see [DL]).

Our assumption on Z_n is that for a certain fixed value of s, we have that for each $g \in W^s(L_2(J))$, there is a function $\zeta_n \in Z_n$ such that

$$(8.2) \qquad \|g - \zeta_n\|_{L_2(J)} \leq c(s) n^{-s} \|g\|_{W^s(L_2(J))}$$

with the constant $c(s)$ depending only on s.

Let X_n be the space of univariate functions r such that for some $p \in \mathcal{P}_n(I)$ and some $\zeta \in Z_n$,

$$(8.3) \qquad r(t) = \begin{cases} p(t), & t \in I \setminus J, \\ \zeta(t), & t \in J. \end{cases}$$

We shall show that under the assumption (8.2) on the Z_n, the spaces X_n, $n = 1, 2, \ldots$, satisfy the assumption (7.1). To prove this, we recall the definition (6.4) of the spaces $W^\alpha(L_2(I, w))$ and the operator Λ of (2.12):

$$(8.4) \qquad \Lambda g := \left(\frac{d}{dt}\right)^{d-1} [wg].$$

According to (2.15), we have $\Lambda^2 \mathcal{U}_n = (-1)^{d-1} \mu_n^2 \mathcal{U}_n$. Since $\mu_n \asymp n^{d-1}$ (see (2.16)), it follows that for each $g \in W^{m\lambda}(L_2(I, w))$, $\lambda = 2(d-1)$, $m = 1, 2, \ldots$, we have

$$(8.5) \qquad \|g\|_{W^{m\lambda}(L_2(I,w))} \asymp \|\Lambda^{2m} g\|_{L_2(I,w)}$$

with the constants of equivalency depending only on d.

LEMMA 8.1. *For each $m = k\lambda$, with $\lambda := 2(d-1)$ and k a nonnegative integer, we have*

$$(8.6) \qquad \|g\|_{W^m(L_2(J))} \le c(d,m)\|g\|_{W^m(L_2(I,w))} \quad for \ g \in W^m(L_2(I,w))$$

with the constant $c(d,m)$ depending only on d and m.

Proof. We first observe that the weight w is strictly positive on J and, therefore, w^{-1} is infinitely times differentiable on J. Then the following identity holds:

$$(8.7) \qquad g^{(\ell(d-1))} = \sum_{j=0}^{\ell(d-1)-1} u_j g^{(j)} + u_{\ell(d-1)}\Lambda^\ell g, \quad \ell = 1, 2, \ldots,$$

where u_j are obtained from w^{-1} and its derivatives. Indeed, (8.7) can be proved by induction on ℓ. For $\ell = 1$, (8.7) follows from Leibniz's formula for differentiating the product $g = w^{-1}(wg)$. Suppose that (8.7) holds for some $\ell \ge 1$. Then one writes $\Lambda^\ell g$ as $w^{-1}(w\Lambda^\ell g)$ and differentiates both sides of (8.7) $d-1$ times to prove it for $\ell + 1$.

It follows from (8.7), with $\ell = 2k$ and $m = k\lambda$, that

$$(8.8) \qquad \|g^{(m)}\|_{L_2(J)} \le c \sum_{j=0}^{m-1} \|g^{(j)}\|_{L_2(J)} + c\|\Lambda^k g\|_{L_2(J)}.$$

We shall use next the following well-known inequality (see, e.g., [BS])

$$(8.9) \qquad \|g^{(j)}\|_{L_2(J)} \le c\left(\delta^{-j}\|g\|_{L_2(J)} + \delta^{m-j}\|g^{(m)}\|_{L_2(J)}\right), \quad j = 1, 2, \ldots, m,$$

where $\delta > 0$ is arbitrary and c depends only on m. Combining (8.8) with (8.9) we get, for $0 < \delta < 1$,

$$(8.10) \qquad \|g^{(m)}\|_{L_2(J)} \le c^*\delta^{-m+1}\|g\|_{L_2(J)} + c^*\delta\|g^{(m)}\|_{L_2(J)} + c^*\|\Lambda^k g\|_{L_2(J)},$$

where $c^* > 1$ is independent of δ. We now select δ such that $c^*\delta = 1/2$ and bring the second term on the right in (8.10) to the left-hand side. We obtain

$$\begin{aligned}
\|g^{(m)}\|_{L_2(J)} &\le c(\|g\|_{L_2(J)} + \|\Lambda^k g\|_{L_2(J)}) \\
&\le c(\|wg\|_{L_2(I)} + \|w\Lambda^k g\|_{L_2(I)}) \\
&\le c\|g\|_{W^m(L_2(I,w))}. \qquad \square
\end{aligned}$$

THEOREM 8.1. *If the sequence of spaces Z_n, $n = 1, 2, \ldots$, satisfies (8.2), then the spaces X_n, $n = 1, 2, \ldots$, defined by (8.3) satisfy the Jackson estimates (7.1); i.e., for each univariate function $g \in W^s(L_2(I,w))$, there is a function $r \in X_n$ which provides the Jackson estimate*

$$(8.11) \qquad \|g - r\|_{L_2(I,w)} \le cn^{-s}\|g\|_{W^s(L_2(I,w))},$$

with c a constant independent of g and n.

Proof. Consider the linear operator T that associates with every function $g \in L_2(I,w)$ the restriction of g on J. Since w is strictly positive on J, T is a bounded operator from $L_2(I,w)$ into $L_2(J)$. By Lemma 8.1, T is bounded from $W^m(L_2(I,w))$ into $W^m(L_2(J))$ for each $m = 2k(d-1)$, $k = 1, 2, \ldots$. This implies that, for each

$0 < s \le 2k(d-1)$ and $\theta := s/(2k(d-1))$, we have by interpolation (see (6.8) and (8.1)) that for each $g \in W^s(L_2(I, w))$,

$$
\begin{aligned}
\|g\|_{W^s(L_2(J))} &\asymp \|g\|_{\left(L_2(J), W^{2k(d-1)}(L_2(J))\right)_{\theta,2}} \\
&\le c\|g\|_{\left(L_2(I,w), W^{2k(d-1)}(L_2(I,w))\right)_{\theta,2}} \asymp \|g\|_{W^s(L_2(I,w))}.
\end{aligned}
$$

Now, given $g \in W^s(L_2(I, w))$, we let $\zeta \in Z_n$ satisfy (8.2). Then, from (8.6),

$$
\|g - \zeta\|_{L_2(J)} \le cn^{-s}\|g\|_{W^s(L_2(J))} \le cn^{-s}\|g\|_{W^s(L_2(I,w))}.
$$

Similarly, let p be the best approximation in $L_2(I, w)$ to g from $\mathcal{P}_n(I)$. Then, from (6.5),

$$
\|g - p\|_{L_2(I,w)} \le n^{-s}\|g\|_{W^s(L_2(I,w))}.
$$

It now follows that the function $r \in X_n$ defined by (8.3) for these ζ and p satisfies (8.11). □

THEOREM 8.2. *If the sequence of spaces $Z_n, n = 1, 2, \ldots$, satisfy* (8.2), *then for any function $f \in W^{s+(d-1)/2}(L_2(\mathbf{B}_{1/2}^d))$, there are functions $r_\omega \in Z_n$ such that*

$$
R(\mathbf{x}) = \sum_{\omega \in \Omega_n} r_\omega(\omega \cdot \mathbf{x}) \tag{8.12}
$$

satisfies

$$
\|f - R\|_{L_2(\mathbf{B}_{1/2}^d)} \le cn^{-s-(d-1)/2}\|f\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}_{1/2}^d))} \tag{8.13}
$$

with c independent of f and n.

Proof. We first recall (see, e.g., [A, Chapter IV]) that f can be extended to a function f_0 defined on all of $\mathbf{R}^d$ such that f_0 vanishes outside of $\mathbf{B}_{3/4}^d$ and

$$
\|f_0\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d))} \le c\|f\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}_{1/2}^d))}
$$

with a constant c depending only on s and d.

We define X_n as in (8.3). From Theorem 8.1, we obtain that condition (7.1) is satisfied. Therefore, from Theorem 7.1 there are functions $r_\omega \in X_n$, $\omega \in \Omega_n$, such that the function

$$
R(\mathbf{x}) = \sum_{\omega \in \Omega_n} r_\omega(\omega \cdot \mathbf{x})
$$

satisfies

$$
\begin{aligned}
\|f_0 - R\|_{L_2(\mathbf{B}^d)} &\le cn^{-r-(d-1)/2}\|f_0\|_{W^{r+(d-1)/2}(L_2(\mathbf{B}^d))} \\
&\le cn^{-r-(d-1)/2}\|f\|_{W^{r+(d-1)/2}(L_2(\mathbf{B}_{1/2}^d))}.
\end{aligned} \tag{8.14}
$$

On the ball $\mathbf{B}_{1/2}^d$, $f_0 = f$ and r_ω is in Z_n for each $\omega \in \Omega_n$. Therefore, (8.13) follows from (8.14). □

9. Examples and further remarks. In this section, we shall give some applications of the results of section 8. Theorem 8.2 implies that for any sequence of spaces Z_n, $n = 1, 2, \ldots$, contained in $L_2(J)$, $J = [-1/2, 1/2]$, that satisfy (8.2) we have the estimate (8.13) for $f \in W^{s+(d-1)/2}(L_2(J))$. The condition (8.2) is satisfied by all the standard spaces of approximation such as algebraic polynomials and spline functions (discussed in more detail later in this section). We wish to single out, for further elaboration, one particular example which appears frequently in wavelet theory, as well as in computer aided design.

Let ϕ be a univariate function with compact support on **R**. Let ℓ be the smallest integer such that ϕ or one of its shifts $\phi(x - k)$, $k \in \mathbf{Z}$, is supported on $[0, \ell]$. If necessary, we can redefine ϕ to be one of its integer shifts and thereby require that ϕ is supported on $[0, \ell]$. We denote by $\mathcal{S} := \mathcal{S}(\phi)$ the shift-invariant space which is the $L_2(\mathbf{R})$-closure of finite linear combinations of the shifts $\phi(\cdot - j)$, $j \in \mathbf{Z}$, of ϕ. By dilation, we obtain the univariate spaces

$$\mathcal{S}^k := \{S(2^k \cdot) : S \in \mathcal{S}\}, \quad k \in \mathbf{Z}.$$

The approximation properties of the family of spaces $\mathcal{S}^k$ is well understood. In [BDR], there is a complete characterization (in terms of the Fourier transform of ϕ) of when the spaces $\mathcal{S}^k$ provide the Jackson estimates

$$(9.1) \qquad \operatorname{dist}(g, \mathcal{S}^k)_{L_2(\mathbf{R})} \le C 2^{-ks} \|g\|_{W^s(L_2(\mathbf{R}))}.$$

For an integer s, we say that ϕ satisfies the Strang–Fix conditions of order s if

$$(9.2) \qquad \hat{\phi}(0) \ne 0, \text{ and } D^j \hat{\phi}(2k\pi) = 0, \ k \in \mathbf{Z}, \ k \ne 0, \ j = 0, 1, \ldots, s-1.$$

If ϕ satisfies (9.2) and ϕ is piecewise continuous and of bounded variation, then $\mathcal{S}^k$ provides the approximation estimate (9.1) (see, e.g., [DL, Chapter 13]).

We denote by $\mathcal{S}^k(J)$, $k \ge 1$, the restrictions of the spaces $\mathcal{S}^k$ to the interval $J := [-1/2, 1/2]$. The functions $\phi(2^k t - j)$, $j = -\ell + 1 - 2^{k-1}, \ldots, 2^{k-1} - 1$, span $\mathcal{S}^k(J)$. Each function g in $W^s(L_2(J))$ can be extended to **R** with

$$\|g\|_{W^s(L_2(\mathbf{R}))} \le c \|g\|_{W^s(L_2(J))}.$$

It follows therefore that the spaces $\mathcal{S}^k(J)$ provide the approximation property (8.2) and hence Theorem 8.2 applies with $n = 2^k$. The functions R appearing in Theorem 8.2 are of the form

$$R(\mathbf{x}) = \sum_{j=-\ell+1-2^{k-1}}^{2^k-1} \sum_{\omega \in \Omega_{2^k}} c(j, \omega) \phi(2^k \mathbf{x} \cdot \omega - j).$$

There is another representation of the functions in $\mathcal{S}^k(J)$ related to sigmoidal functions. Let

$$(9.3) \qquad \sigma(t) := \sum_{j=0}^{\infty} \phi(t - j).$$

Then the functions $\sigma(2^k t - j)$, $j = -\ell + 1 - 2^{k-1}, \ldots, 2^{k-1} - 1$, also span $\mathcal{S}^k(J)$. The function σ is 0 for t sufficiently large negative and 1 for t sufficiently large positive. However, it is not necessarily monotone (without additional assumptions on ϕ).

COROLLARY 9.1. *Let ϕ satisfy the Strang–Fix conditions* (9.2) *of order s. Then for each function $f \in W^{s+(d-1)/2}(L_2(\mathbf{B}^d_{1/2}))$, there is a function*

$$R(\mathbf{x}) = \sum_{j=-\ell+1+2^{k-1}}^{2^{k-1}-1} \sum_{\omega \in \Omega_{2^k}} c(j,\omega)\sigma(2^k \mathbf{x} \cdot \omega - j)$$

such that

$$\|f - R\|_{L_2(\mathbf{B}^d_{1/2})} \le c\, 2^{-(s+(d-1)/2)k}\|f\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d_{1/2}))}, \quad k = 1,2,\ldots,$$

with c independent of f and k.

For certain choices of ϕ above, we obtain that σ of (9.3) is a sigmoidal function in the terminology of neural networks. We recall that a sigmoidal function is a nonnegative, monotone, univariate function which has limits $= 0$ as $t \to -\infty$ and $= 1$ as $t \to \infty$. To obtain examples of such sigmoidal functions, we can take ϕ to be a B-spline. Let $\phi := N_{0,s}$, where for each $j \in \mathbf{Z}$ and $s = 1,2,\ldots$, $N_{j,s} := s^{-1}M_{j,s}$ is the B-spline of order s (see [DL, Chapter 5]) with breakpoints $\frac{j}{2n},\ldots,\frac{j+s}{2n}$. The function

$$\sigma_s(t) := \sum_{j=-s+1}^{\infty} N_{j,s}(t), \quad t \in \mathbf{R}$$

of (9.3) is a sigmoidal function, and, in the case $s = 1$, it is the unit impulse function $\chi_{[0,\infty)}$. The functions $\sigma_s(t - \frac{j}{2n})$, $j = -n,\ldots,n+s-1$, form a basis for $\mathcal{S}_{n,s}$ the space of all splines of degree $s-1$ defined on J with breakpoints belonging to the set $\{\frac{-n+1}{2n}, \frac{-n+2}{2n}, \ldots, \frac{n-1}{2n}\}$. From Theorem 8.2, we obtain the following.

COROLLARY 9.2. *For any $f \in W^{s+(d-1)/2}(L_2(\mathbf{B}^d_{1/2}))$, there are constants $c(k,\omega)$, $\omega \in \Omega_n$, $k = -n,\ldots,n+s-1$, such that*

$$(9.4) \qquad R(\mathbf{x}) = \sum_{\omega \in \Omega_n} \sum_{k=-n}^{n+s-1} c(k,\omega)\sigma_s\left(\mathbf{x} \cdot \omega - \frac{k}{2n}\right)$$

satisfies

$$\|f - R\|_{L_2(\mathbf{B}^d_{1/2})} \le cn^{-s-(d-1)/2}\|f\|_{W^{s+(d-1)/2}(L_2(\mathbf{B}^d_{1/2}))}$$

with c independent of f and n.

The functions R in (9.4) correspond to the outputs of a feed-forward neural network with $O(n^{d-1})$ nodes of computation. Thus, the corollary shows that such neural networks have computational efficiency comparable with standard methods of approximation like splines and wavelets.

The special case $s = 1$ in Corollary 9.2 is also noteworthy. In this case the function σ is the unit-impulse function and the functions R are piecewise constant. The order of approximation provided by Corollary 9.2 is somewhat surprising. One might expect that such piecewise constants could only provide approximation order 1 while the corollary gives approximation order $(d+1)/2$.

10. Appendix.

A1. Proof of (3.4). Since $\mathcal{P}_n$ is invariant under rotations, it is sufficient to prove that $\langle P(\mathbf{x}), \mathcal{U}_n(x_1)\rangle = 0$ for each $P \in \mathcal{P}_{n-1}$ or that

$$\langle \mathbf{x^m}, \mathcal{U}_n(x_1)\rangle := \int_{\mathbf{B}^d} \mathbf{x^m}\mathcal{U}_n(x_1)\, d\mathbf{x} = 0 \quad \text{when} \quad |\mathbf{m}| \leq n-1.$$

Write

$$\mathbf{B}_{x_1} := \{\mathbf{x}' = (x_2, \ldots, x_d) : x_2^2 + \cdots + x_d^2 \leq 1 - x_1^2\}.$$

We have

$$\langle \mathbf{x^m}, \mathcal{U}_n(x_1)\rangle = \int_I x_1^{m_1}\left(\int_{\mathbf{B}_{x_1}} x_2^{m_2}\ldots x_d^{m_d}\, d\mathbf{x}'\right)\mathcal{U}_n(x_1)\, dx_1.$$

Because of the symmetry it is obvious that the inner integral above is equal to zero if at least one of $m_2, \ldots, m_d$ is odd. Consider the case when all $m_2, \ldots, m_d$ are even. We now change the rectangular coordinates in the inner integral to spherical and find

$$\int_{\mathbf{B}_{x_1}} x_2^{m_2}\ldots x_d^{m_d}\, d\mathbf{x}' = c\int_0^{(1-x_1^2)^{1/2}} r^{m_2+\cdots+m_d+d-2}\, dr = c(1-x_1^2)^{\frac{1}{2}(m_2+\cdots+m_d+d-1)},$$

where c depends on $d, m_2, \ldots, m_d$. Therefore,

$$\langle \mathbf{x}^m, C_n(x_1)\rangle = c\int_I x_1^{m_1}(1-x_1^2)^{\frac{1}{2}(m_2+\cdots+m_d)}\mathcal{U}_n(x_1)(1-x_1^2)^{\frac{d-1}{2}}\, dx_1 = 0$$

since the univariate polynomial $\mathcal{U}_n$ is orthogonal to $\mathcal{P}_{n-1}(I)$ in $L_2(I, w)$ (see (3.3) and (2.1)).   □

A2. Proof of (3.10). We first show that for each $g \in L_1(I, w_{(d-1)/2})$

$$\mathcal{R}(g(\eta\cdot\mathbf{x}); \xi, t) = |B^{d-2}|(1-t^2)^{\frac{d-1}{2}}\int_I g(\cos\theta\cos\psi + u\sin\theta\sin\psi)(1-u^2)^{\frac{d-2}{2}}\, du,$$

(10.1)
where $t =: \cos\theta$, $t \in I$, $\psi \in [0, \pi]$ is the angle between ξ and η $(\cos\psi = \xi\cdot\eta)$, $|B^{d-2}|$ is the volume of the unit ball B^{d-2} in $\mathbf{R}^{d-2}$, $|B^{d-2}| = \frac{2\pi^{d/2-1}}{(d-2)\Gamma(d/2-1)}$, and $\mathcal{R}$ is the Radon transform defined in (3.23). Indeed, it is easily seen that

$$\mathcal{R}(g(\eta\cdot\mathbf{x}); \xi, t) = |B^{d-2}|\int_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} g(t\cos\psi + v\sin\psi)(1-t^2-v^2)^{\frac{d-2}{2}}\, dv.$$

Substituting $v = (1-t^2)^{1/2}u$ in the above integral we get (10.1).

Our second step is to prove that

$$(10.2)\quad \mathcal{R}(C_n^{d/2}(\eta\cdot\mathbf{x}); \xi, t) = |B^{d-2}|\frac{2^{d-1}\Gamma^2(d/2)n!}{\Gamma(n+d)}(1-t^2)^{\frac{d-1}{2}}C_n^{d/2}(t)C_n^{d/2}(\xi\cdot\eta).$$

Indeed, the classical addition theorem for Legendre (Gegenbauer) polynomials can be written as follows (see [E, p. 178]):

$$C_n^\lambda(\cos\theta\cos\psi + \sin\theta\sin\psi\cos\varphi)$$
$$= \sum_{m=0}^n 2^m(2\lambda + 2m - 1)(n-m)!\frac{[(\lambda)_m]^2}{(2\lambda-1)_{n+m+1}}$$
$$\times (\sin\theta)^m C_{n-m}^{\lambda+m}(\cos\theta)(\sin\psi)^m C_{n-m}^{\lambda+m}(\cos\psi)C_m^{\lambda-1/2}(\cos\varphi)$$

and hence, for $\lambda = d/2$, and $u := \cos \varphi$, $u \in I$, we have

$$
C_n^{d/2}(\cos \theta \cos \psi + u \sin \theta \sin \psi)
$$
$$
= \sum_{m=0}^{n} 2^m (d + 2m - 1)(n - m)! \frac{[(d/2)_m]^2}{(d-1)_{n+m+1}}
$$
$$
\times (\sin \theta)^m C_{n-m}^{d/2+m}(\cos \theta)(\sin \psi)^m C_{n-m}^{d/2+m}(\cos \psi) C_m^{(d-1)/2}(u).
$$

We now insert this into (10.1) and use the fact that $C_m^{(d-1)/2}(u)$, $m = 1, 2, \ldots$, are orthogonal to the constants in $L_2(I, w_{(d-1)/2})$ to obtain

$$
\mathcal{R}(C_n^{d/2}(\eta \cdot \mathbf{x}); \xi, t)
$$
$$
= |B^{d-2}|(1 - t^2)^{(d-1)/2} \frac{(d-1)n!}{(d-1)_{n+1}} C_n^{d/2}(\cos \theta) C_n^{d/2}(\cos \psi) \int_I (1 - u^2)^{(d-2)/2} \, du.
$$

This implies (10.2).

We finally use (10.2) to obtain

$$
\int_{\mathbf{B}^d} C_n^{d/2}(\eta \cdot \mathbf{x}) C_n^{d/2}(\xi \cdot \mathbf{x}) \, d\mathbf{x} = \int_I \mathcal{R}(C_n^{d/2}(\eta \cdot \mathbf{x}); \xi, t) C_n^{d/2}(t) \, dt
$$
$$
= |B^{d-2}| \frac{2^{d-1} \Gamma^2(d/2) n!}{\Gamma(n+d)} C_n^{d/2}(\eta \cdot \xi) \int_I [C_n^{d/2}(t)]^2 (1 - t^2)^{(d-1)/2} \, dt
$$
$$
= \gamma_{n,d} C_n^{d/2}(\eta \cdot \xi),
$$

where

$$
\gamma_{n,d} := |B^{d-2}| \frac{2^{d-1} \Gamma^2(\frac{d}{2}) n!}{\Gamma(n+d)} h_{n,d/2}.
$$

Simple calculations show that this is (3.10). See [RK]. $\quad\square$

A3. Proof of (3.16). The following relation between contiguous Gegenbauer polynomials holds (see [E, p. 178], (36)):

$$
(n + \lambda) C_{n+1}^{\lambda - 1} = (\lambda - 1) \left[C_{n+1}^{\lambda} - C_{n-1}^{\lambda} \right], \quad \lambda > 1.
$$

Also, $C_0^\lambda(t) = 1$ and $C_1^\lambda(t) = 2\lambda t$. These identities readily imply

$$
(10.3) \qquad\qquad C_n^\lambda = \sum_{j=0}^{[n/2]} \frac{n - 2j + \lambda - 1}{\lambda - 1} C_{n-2j}^{\lambda - 1}.
$$

Simple calculations show that (10.3) with $\lambda = d/2$ $(d > 2)$ is (3.16). $\quad\square$

A4. Proof of (3.11). Identity (3.11) follows from the fact that $\mathcal{U}_n(\xi \cdot \mathbf{x})$ (as a function of ξ) is a spherical polynomial in $\mathcal{H}_n \oplus \mathcal{H}_{n-2} \oplus \cdots \oplus \mathcal{H}_\epsilon$ and $\nu_n \mathcal{U}_n(\xi \cdot \eta) / \mathcal{U}_n(1)$ is the reproducing kernel for this space (see (3.17)). $\quad\square$

REFERENCES

[A]   R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[BDR]   C. de Boor, R. DeVore, and A. Ron, *Approximation from shift invariant spaces*, Trans. Amer. Math. Soc., 341 (1994), pp. 787–806.

[BS]   C. Bennett and R. Sharpley, *Interpolation of Operators*, Academic Press, New York, 1988.

[DL]   R. DeVore and G. Lorentz, *Constructive Approximation*, Grundlehren Math. Wiss. 303, Springer, Heidelberg, 1993.

[DOP]   R. DeVore, K. Oskolkov, and P. Petrushev, *Approximation by feed-forward neural networks*, Ann. Numer. Math., 4 (1997), pp. 261–287.

[E]   A. Erdelyi, ed., *Higher Transcendental Functions*, Vol. 2, McGraw–Hill, New York, 1953.

[K]   E. Kogbetliantz, *Recherches sur la sommabilité des séries ultraspheriques par la méthode de moyennes arithmetiques*, J. Math. Pures Appl., 9 (1924), pp. 107–187.

[L]   D. Ludwig, *The Radon transform on Euclidean spaces*, Comm. Pure Appl. Math., 19 (1966), pp. 49–81.

[LS]   B. Logan and L. Schepp, *Optimal reconstruction of a function from its projections*, Duke Math. J., 42 (1975), pp. 645–659.

[M]   H. Mhaskar, *Neural networks for optimal approximation of smooth and analytic functions*, Neural Comput., 8 (1996), pp. 164–177.

[MM]   H. Mhaskar and C. Micchelli, *Approximation by superposition of sigmoidal and radial basis functions*, Adv. Appl. Math., 13 (1992), pp. 350–373.

[MM1]   H. Mhaskar and C. Micchelli, *Degree of approximation by neural and translation networks with a single hidden layer*, Adv. Appl. Math., 16 (1995), pp. 151–183.

[O]   K. Oskolkov, *Inequalities of the "large sieve" type and applications to problems of trigonometric approximation*, Anal. Math., 12 (1986), pp. 143–166.

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

[RK]   A. Ramm and A. Katsevich, *The Radon Transform and Local Tomography*, CRC Press, Boca Raton, FL, 1996.

[Se]   R. Seeley, *Spherical harmonics*, Amer. Math. Monthly, 73 (1966), pp. 115–121.

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

[SW]   E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, NJ, 1971.

[Sz]   G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, Providence, RI, 1975.

[Z]   A. Zygmund, *Trigonometric Series*, Vols. I, II, Cambridge University Press, Cambridge, 1977.