

# Localized linear polynomial operators and quadrature formulas on the sphere

Q. T. Le Gia\*

School of Mathematics, University of New South Wales,  
Sydney, NSW 2052, Australia  
qlegia@maths.unsw.edu.au

H. N. Mhaskar†

Department of Mathematics, California State University  
Los Angeles, California, 90032, U.S.A.  
hnhaska@calstatela.edu

## Abstract

The purpose of this paper is to construct universal, auto-adaptive, localized, linear, polynomial (-valued) operators based on scattered data on the (hyper-)sphere  $\mathbb{S}^q$  ( $q \geq 2$ ). The approximation and localization properties of our operators are studied theoretically in deterministic as well as probabilistic settings. Numerical experiments are presented to demonstrate their superiority over traditional least squares and discrete Fourier projection polynomial approximations. An essential ingredient in our construction is the construction of quadrature formulas based on scattered data, exact for integrating spherical polynomials of (moderately) high degree. Our formulas are based on scattered sites; i.e., in contrast to such well known formulas as Driscoll–Healy formulas, we need not choose the location of the sites in any particular manner. While the previous attempts to construct such formulas have yielded formulas exact for spherical polynomials of degree at most 18, we are able to construct formulas exact for spherical polynomials of degree 178.

**Keywords:** Quadrature formulas, localized kernels, polynomial quasi-interpolation, learning theory on the sphere.

**AMS classification:** 65D32, 41A10, 41A25

## 1 Introduction

The problem of approximation of functions on the sphere arises in almost all applications involving modeling of data collected on the surface of the earth. More recent applications such as manifold matching and neural networks lead to the approximation of functions on the unit sphere  $\mathbb{S}^q$  embedded in the Euclidean space  $\mathbb{R}^{q+1}$  for integers  $q \geq 3$  as well. Various applications in learning theory, meteorology, cosmology, and geophysics require analysis of *scattered data* collected on the sphere [7, 8, 9]. This means that the data is of the form  $\{(\xi, f(\xi))\}$  for some unknown function  $f : \mathbb{S}^q \rightarrow \mathbb{R}$ , where one has no control on the choice of the sites  $\xi$ .

There are many methods to model such data: spherical splines, radial basis functions (called zonal function networks in this context), etc. However, the most traditional method is to approximate by

---

\*The research of this author was supported by Australian Research Council under its Centres of Excellence Program.

†The research of this author was supported, in part, by grant DMS-0605209 from the National Science Foundation and grant W911NF-04-1-0339 from the U.S. Army Research Office.

spherical polynomials; i.e., restrictions of algebraic polynomials in  $q + 1$  variables to  $\mathbb{S}^q$ . Apart from tradition, some important advantages of polynomials are that they are eigenfunctions of many pseudo-differential operators which arise in practical applications, and that they are infinitely smooth. Unlike in the case of spline approximation with a given degree of the piecewise component polynomials, global polynomial approximation does not exhibit a saturation property [2, Section 2, Chapter 11]; i.e., for an arbitrary sequence  $\delta_n \downarrow 0$ , it is possible to find a continuous function on the sphere, not itself a polynomial, which can be approximated by spherical polynomials of degree at most  $n$  uniformly within  $\delta_n$ ,  $n \geq 1$ . In [20, 19], we have shown how a good polynomial approximation yields also a good zonal function network approximation. In [22], we have shown that the approximation spaces determined by zonal function network approximation are the same as those determined by polynomial approximations.

To illustrate the issues to be discussed in this paper, we consider an example in the case  $q = 1$ , or equivalently, the case of  $2\pi$ -periodic functions on the real line. In this discussion only, let  $f(x) = |\cos x|^{1/4}$ ,  $x \in \mathbb{R}$ . In Figure 1(left), we show the log-plot of the absolute errors between  $f$  and its (trigonometric) Fourier projection of order 31, where the Fourier coefficients are estimated by a 128 point DFT. In Figure 1(right), we show a similar log-plot where the Fourier projection is replaced by a suitable summability operator (described more precisely in (3.1)), yielding again a trigonometric polynomial of order 31. It is clear that our summability operator is far more localized than the Fourier projection; i.e., the error in approximation decreases more rapidly as one goes away from the singularities at  $\pi/2$  and  $3\pi/2$ . The maximum error on  $[3\pi/4, 5\pi/4]$  is 0.0103 for the projection, 0.0028 for our operator. Out of the 2048 points considered for the test, the error by the summability operator is less than  $10^{-3}$  at 38.96% points, the corresponding percentage for the projection is only 4.88%. In contrast to free-knot spline approximation, our summability operator is universal; i.e., its construction (convolution with a kernel) does not require any a priori knowledge about the location of singularities of the target function. It yields a single, globally defined trigonometric polynomial, computed using global data. Nevertheless, it is auto-adaptive, in the sense that the error in approximation on different subintervals adjusts itself according to the smoothness of the target function on these subintervals. In [23, 24], we have given a very detailed analysis of the approximation properties of these operators in the case  $q = 1$ .

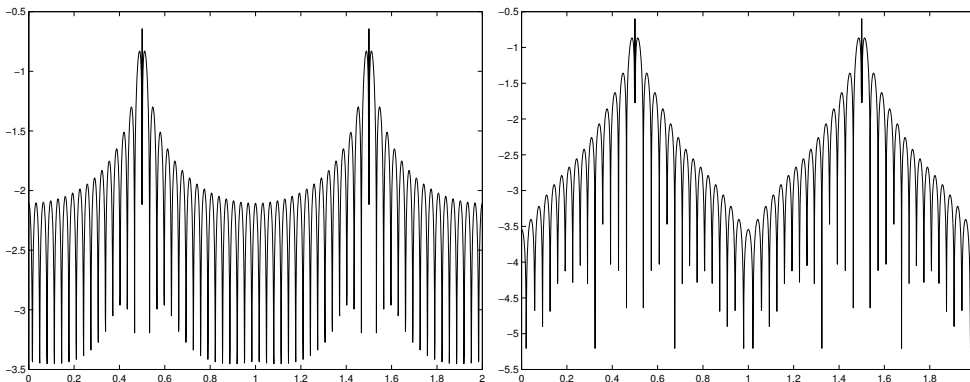


Figure 1: The log-plot of the absolute error between the function  $x \mapsto |\cos x|^{1/4}$ , and (left) its Fourier projection (right) trigonometric polynomial obtained by our summability operator, where the Fourier coefficients are estimated by 128 point FFT. The order of the trigonometric polynomials is 31 in each case. The numbers on the  $x$  axis are in multiples of  $\pi$ , the actual absolute errors are  $10^y$ .

Our computation based on a 128 point DFT implies that the values of the function are available at 128 equidistant points. If only a scattered data is available, the following method is often used (especially in the context of approximation on the sphere) to estimate the values needed for the DFT. For each point  $\xi$ , we consider the nearest point of the form  $2\pi k/128$ , and imagine that the value of  $f$  at this point is  $f(\xi)$ , taking averages in the case of multiplicities, and interpolating in the case of gaps. If we use our summability operator, estimating the Fourier coefficients in this way, then the maximum error on  $[3\pi/4, 5\pi/4]$  is 0.0357, and the proportion of points where the error is less than  $10^{-3}$  is 7.08%. It is clear that a careful construction of quadrature formulas is essential to obtain good approximation results.

The purpose of this paper is to construct universal, auto-adaptive, localized, linear, polynomial (-

valued) operators based on scattered data on  $\mathbb{S}^q$  ( $q \geq 2$ ) and to analyse their approximation properties. An essential ingredient in our construction is the construction of quadrature formulas based on scattered data, exact for integrating spherical polynomials of (moderately) high degree, and satisfying certain technical conditions known as the Marcinkiewicz–Zygmund (M–Z) conditions. Our construction is different from the usual construction of quadrature formulas (designs) studied in numerical analysis, where one has a choice of the placement of nodes. In [21], we had proved the existence of such quadrature formulas for scattered data. However, previous efforts to compute such formulas did not yield exactness beyond degree 18 polynomials. This was a severe limitation on the practical applications of our theoretical constructions. We will show that a very simple idea of solving a system of equations involving a Gram matrix yields surprisingly good results, in particular, quadrature formulas exact for integrating polynomials of degree as high as 178. Gram matrices are typically ill-conditioned. However, we will show both theoretically and numerically that the ones which we use are, in fact, very well conditioned. We will introduce another algorithm of theoretical interest to compute data dependent orthogonal polynomials, and use these to compute the quadrature formulas in a memory efficient manner. To the best of our knowledge, this is the first effort to extend the univariate constructions in Gautschi’s book [11] to a multivariate setting. Considering that computation of classical spherical harmonics is a very delicate task, requiring many tricks based on the special function properties of these polynomials for a stable computation, it is not expected that our computation of data dependent orthogonal polynomials with no such special function properties would be stable. In describing this algorithm, we hope to stimulate further research in this interesting direction. We note that even if this algorithm is not as stable for high degrees as the other algorithm, it yields satisfactory quadrature formulas exact for integrating polynomials of degree 32. Most importantly, our new found ability to compute quadrature formulas for moderately high degrees allows us to offer our operators as a viable, practical method of approximation, even superior to the commonly used methods of least squares and Fourier projection as far as localized approximation is concerned.

An additional problem is when the available values of the target function are noisy. One may assume that the noise is an additive random variable with mean zero. It is also routine in learning theory to assume that the random variables have a bounded range. This assumption is usually satisfied with a high probability even if the random variables do not actually have a bounded range. However, one does not typically know the actual distribution of these random variables. We obtain probabilistic estimates in this setting on the global and local approximations by our operators. To underline the practical utility of our operators, we use them for modelling the MAGSAT data supplied to us by Dr. Thorsten Maier, obtaining results comparable to those obtained by other techniques.

In Section 2, we review certain facts about spherical polynomials, the existence of quadrature formulas to integrate these, a few properties of the quadrature weights, and certain polynomial kernels which we will need throughout the paper. In Section 3, we study the approximation properties of the linear polynomial operators. The new results here are Theorems 3.1 and 3.2. The first parts of these theorems were proved essentially in [18], but not stated in the form given here. In order to apply these operators in practice, one needs quadrature formulas exact for high degree spherical polynomials. Explicit algorithms to construct such formulas are described in Section 4. The new results in this section are Theorems 4.1 and 4.2. Numerical results are presented in Section 5, and the proofs of all new results are given in Section 6. The paper is a result of a long process, involving discussions with a number of mathematicians. In particular, it is our pleasure to acknowledge the support and encouragement of Mahadevan Ganesh, Thorsten Maier, Volker Michel, Dominik Michel, Ian Sloan, and Joe Ward. We are also grateful to the two referees and Fred Hickernell for their many useful suggestions for the improvement of the first draft of this paper.

## 2 Background

In this section, we review some known results regarding spherical polynomials and localized polynomial kernels.

## 2.1 Spherical polynomials

Let  $q \geq 1$  be an integer,  $\mathbb{S}^q$  be the unit sphere embedded in the Euclidean space  $\mathbb{R}^{q+1}$ ; i.e.,  $\mathbb{S}^q := \{(x_1, \dots, x_{q+1}) \in \mathbb{R}^{q+1} : \sum_{k=1}^{q+1} x_k^2 = 1\}$ , and  $\mu_q$  be its Lebesgue surface measure, normalized so that  $\mu_q(\mathbb{S}^q) = 1$ . The surface area of  $\mathbb{S}^q$  is  $\frac{2\pi^{(q+1)/2}}{\Gamma((q+1)/2)}$ . For  $\delta > 0$ , a spherical cap with radius  $\delta$  and center  $\mathbf{x}_0 \in \mathbb{S}^q$  is defined by

$$\mathbb{S}_\delta^q(\mathbf{x}_0) := \{\mathbf{x} \in \mathbb{S}^q : \arccos(\mathbf{x} \cdot \mathbf{x}_0) \leq \delta\}.$$

If  $1 \leq p \leq \infty$ , and  $f : \mathbb{S}^q \rightarrow \mathbb{R}$  is measurable, we write

$$\|f\|_p := \begin{cases} \{\int_{\mathbb{S}^q} |f(\mathbf{x})|^p d\mu_q(\mathbf{x})\}^{1/p}, & \text{if } 1 \leq p < \infty, \\ \text{ess sup}_{\mathbf{x} \in \mathbb{S}^q} |f(\mathbf{x})|, & \text{if } p = \infty. \end{cases}$$

The space of all Lebesgue measurable functions on  $\mathbb{S}^q$  such that  $\|f\|_p < \infty$  will be denoted by  $L^p$ , with the usual convention that two functions are considered equal as elements of this space if they are equal almost everywhere. The symbol  $C(\mathbb{S}^q)$  denotes the class of all continuous, real valued functions on  $\mathbb{S}^q$ , equipped with the norm  $\|\circ\|_\infty$ .

For a real number  $x \geq 0$ , let  $\Pi_x^q$  denote the class of all spherical polynomials of degree at most  $x$ . (This is the same as the class  $\Pi_n^q$ , where  $n$  is the largest integer not exceeding  $x$ . However, our extension of the notation allows us, for example, to use the simpler notation  $\Pi_{n/2}^q$  rather than the more cumbersome notation  $\Pi_{\lfloor n/2 \rfloor}^q$ .) For a fixed integer  $\ell \geq 0$ , the restriction to  $\mathbb{S}^q$  of a homogeneous harmonic polynomial of exact degree  $\ell$  is called a spherical harmonic of degree  $\ell$ . Most of the following information is based on [25], [32, Section IV.2], and [5, Chapter XI], although we use a different notation. The class of all spherical harmonics of degree  $\ell$  will be denoted by  $\mathbf{H}_\ell^q$ . The spaces  $\mathbf{H}_\ell^q$  are mutually orthogonal relative to the inner product of  $L^2$ . For any integer  $n \geq 0$ , we have  $\Pi_n^q = \bigoplus_{\ell=0}^n \mathbf{H}_\ell^q$ . The dimension of  $\mathbf{H}_\ell^q$  is given by

$$d_\ell^q := \dim \mathbf{H}_\ell^q = \begin{cases} \frac{2\ell + q - 1}{\ell + q - 1} \binom{\ell + q - 1}{\ell}, & \text{if } \ell \geq 1, \\ 1, & \text{if } \ell = 0. \end{cases} \quad (2.1)$$

and that of  $\Pi_n^q$  is  $\sum_{\ell=0}^n d_\ell^q = d_n^{q+1}$ . Furthermore,  $L^2 = L^2$ -closure  $\{\bigoplus_{\ell=0}^\infty \mathbf{H}_\ell^q\}$ . Hence, if we choose an orthonormal basis  $\{Y_{\ell,k} : k = 1, \dots, d_\ell^q\}$  for each  $\mathbf{H}_\ell^q$ , then the set  $\{Y_{\ell,k} : \ell = 0, 1, \dots \text{ and } k = 1, \dots, d_\ell^q\}$  is a complete orthonormal basis for  $L^2$ . One has the well-known addition formula [25] and [5, Chapter XI, Theorem 4]:

$$\sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\mathbf{x}) Y_{\ell,k}(\zeta) = \frac{2^{q-1} \Gamma(q/2)^2}{\Gamma(q)} p_\ell(1) p_\ell(\mathbf{x} \cdot \zeta), \quad \ell = 0, 1, \dots, \quad (2.2)$$

where  $p_\ell := p_\ell^{(q/2-1, q/2-1)}$  is the orthonormalized Jacobi polynomial with positive leading coefficient:

$$\int_{-1}^1 p_\ell(t) p_k(t) (1-t^2)^{q/2-1} dt = \begin{cases} 1, & \text{if } \ell = k, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, for  $\mathbf{x} \in \mathbb{S}^q$ ,  $\ell = 0, 1, \dots$ ,

$$\sum_{k=1}^{d_\ell^q} Y_{\ell,k}^2(\mathbf{x}) = \frac{2^{q-1} \Gamma(q/2)^2}{\Gamma(q)} p_\ell(1)^2 = \int_{\mathbb{S}^q} \sum_{k=1}^{d_\ell^q} Y_{\ell,k}^2(\zeta) d\mu_q(\zeta) = d_\ell^q. \quad (2.3)$$

## 2.2 Localized polynomial kernels

Let  $h : [0, \infty) \rightarrow \mathbb{R}$  be a compactly supported function, and  $t > 0$ . We define for  $u \in \mathbb{R}$ ,

$$\Phi_t(h; u) := \frac{2^{q-1} \Gamma(q/2)^2}{\Gamma(q)} \sum_{\ell=0}^{\infty} h(\ell/t) p_\ell(1) p_\ell(u) \quad (2.4)$$

and define  $\Phi_t(h; u) = 0$  if  $t \leq 0$ .

In the sequel, we adopt the following convention regarding constants. The letters  $c, c_1, \dots$  will denote generic, positive constants depending only on the dimension  $q$  and other such fixed quantities in the discussion as the function  $h$ , the different norms involved in the formula, etc.. Their value will be different at different occurrences, even within the same formula. The symbol  $A \sim B$  will mean  $cA \leq B \leq c_1A$ .

The following proposition summarizes some of the important properties of the kernels defined in (2.4).

**Proposition 2.1** *Let  $S \geq q$  be an integer,  $h : [0, \infty) \rightarrow \mathbb{R}$  be a  $S$  times iterated integral of a function of bounded variation,  $h(x) = 1$  for  $x \in [0, 1/2]$ ,  $h(x) = 0$  for  $x > 1$ , and  $h$  be non-increasing. Let  $\mathbf{x} \in \mathbb{S}^q$ . We have for every integer  $n \geq 0$ ,  $\Phi_n(h; \circ \cdot \mathbf{x}) \in \Pi_n^q$ , and*

$$\int \Phi_n(h; \mathbf{x} \cdot \zeta) P(\zeta) d\mu_q(\zeta) = P(\mathbf{x}), \quad P \in \Pi_{n/2}. \quad (2.5)$$

Further,

$$\begin{aligned} \sup_{n \geq 1, \zeta \in \mathbb{S}^q} \int |\Phi_n(h; \zeta \cdot \xi)| d\mu_q(\xi) &= \sup_{n \geq 1} \int |\Phi_n(h; \mathbf{x} \cdot \xi)| d\mu_q(\xi) \\ &= \frac{2\pi^{q/2}}{\Gamma(q/2)} \sup_{n \geq 1} \int_{-1}^1 |\Phi_n(h; u)| (1 - u^2)^{q/2-1} du < \infty, \end{aligned} \quad (2.6)$$

$$\begin{aligned} \int |\Phi_n(h; \mathbf{x} \cdot \xi)|^2 d\mu_q(\xi) &= \frac{2\pi^{q/2}}{\Gamma(q/2)} \int_{-1}^1 |\Phi_n(h; u)|^2 (1 - u^2)^{q/2-1} du \\ &\sim n^q \sim \max_{\xi \in \mathbb{S}^q} |\Phi_n(h; \mathbf{x} \cdot \xi)| = |\Phi_n(h; 1)|. \end{aligned} \quad (2.7)$$

and for every  $\xi \in \mathbb{S}^q$ ,  $\xi \neq \mathbf{x}$ ,

$$|\Phi_n(h; \mathbf{x} \cdot \xi)| \leq cn^q \begin{cases} (n\sqrt{1 - \mathbf{x} \cdot \xi})^{1/2-q/2-S}, & \text{if } 0 \leq \mathbf{x} \cdot \xi < 1, \\ n^{-S}, & \text{if } -1 \leq \mathbf{x} \cdot \xi < 0. \end{cases} \quad (2.8)$$

Except for (2.7), all parts of Proposition 2.1 have been proved and verified repeatedly in [17, 18, 12, 19]. We will sketch a proof of this proposition, mainly to reconcile notations.

PROOF OF PROPOSITION 2.1. The equation (2.5) and the first two equations in (2.6) are clear. The last estimate in (2.6) follows from [17, Lemma 4.6] with following choice of the parameters there:  $\alpha = \beta = q/2 - 1$ ,  $h_\nu = h(\nu/n)$ , where we observe that by a repeated application of the mean value theorem,

$$\sum_{\nu=0}^{\infty} (\nu+1)^s |\Delta^r h(\nu/n)| \leq cn^{s-r+1}, \quad s \in \mathbb{R}, r, n = 1, 2, \dots,$$

where  $\Delta^r$  is the  $r$ -th order forward difference applied with respect to  $\nu$ . Similarly, the estimate (2.8) follows from [17, Lemma 4.10] with same parameters as above,  $S$  in place of  $K$  in [17], and  $y = \mathbf{x} \cdot \xi$  (cf. Appendix to [12]). We prove (2.7). The first equation is a consequence of the rotation invariance of  $\mu_q$ . In view of the addition formula (2.2),

$$\Phi_n(h; \mathbf{x} \cdot \xi) = \sum_{\ell=0}^n h(\ell/n) \sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\mathbf{x}) Y_{\ell,k}(\xi). \quad (2.9)$$

It follows using (2.3) and the facts that  $h(\ell/n) = 1$  for  $\ell \leq n/2$ ,  $0 \leq h(t) \leq 1$  for  $t \in [0, \infty)$ , that

$$\int \Phi_n(h; \mathbf{x} \cdot \xi)^2 d\mu_q(\xi) = \sum_{\ell=0}^n h(\ell/n)^2 \sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\mathbf{x})^2 = \sum_{\ell=0}^n h(\ell/n)^2 d_\ell^q \sim n^q.$$

Similarly, using Schwarz inequality, (2.2), (2.3), and the fact that  $h(\ell/n) \geq 0$ ,

$$\begin{aligned} \Phi_n(h; 1) &= |\Phi_n(h; \mathbf{x} \cdot \mathbf{x})| \leq \sup_{\xi \in \mathbb{S}^q} |\Phi_n(h; \mathbf{x} \cdot \xi)| \\ &\leq \sum_{\ell=0}^n h(\ell/n) \left\{ \sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\mathbf{x})^2 \right\}^{1/2} \left\{ \sum_{k=1}^{d_\ell^q} Y_{\ell,k}(\xi)^2 \right\}^{1/2} = \sum_{\ell=0}^n h(\ell/n) d_\ell^q = \Phi_n(h; 1). \end{aligned}$$

Since  $d_\ell^q \sim \ell^{q-1}$ ,  $0 \leq h(\ell/n) \leq 1$ , and  $h(\ell/n) = 1$  for  $\ell \leq n/2$ , the above two estimates lead to (2.7).  $\square$

In the remainder of this paper,  $h$  will denote a fixed function satisfying the conditions of Proposition 2.1.

### 2.3 Quadrature formulas

Let  $\mathcal{C}$  be a finite set of distinct points on  $\mathbb{S}^q$ . A *quadrature formula* based on  $\mathcal{C}$  has the form  $\mathcal{Q}(f) = \sum_{\xi \in \mathcal{C}} w_\xi f(\xi)$ , where  $w_\xi$ ,  $\xi \in \mathcal{C}$ , are real numbers. For integer  $n \geq 0$ , the formula is exact for degree  $n$  if  $\mathcal{Q}(P) = \int_{\mathbb{S}^q} P d\mu_q$  for all  $P \in \Pi_n^q$ . It is not difficult to verify that if  $\mathcal{Q}_n(f) = \sum w_{\xi_n} f(\xi_n)$  is a sequence of quadrature formulas, with  $\mathcal{Q}_n$  being exact with degree  $n$ , then  $\mathcal{Q}_n(f) \rightarrow \int f d\mu_q$  for every continuous function  $f$  on  $\mathbb{S}^q$  if and only if  $\sum |w_{\xi_n}| \leq c$ , with  $c$  being independent of  $n$ . In the sequel, we will assume tacitly that  $\mathcal{C}$  is one of the members of a nested sequence of finite subsets of  $\mathbb{S}^q$ , whose union is dense in  $\mathbb{S}^q$ . All the constants may depend upon the whole sequence, but not on any individual member of this sequence. Thus, a formula  $\mathcal{Q}$  will be called a bounded variation formula if  $\sum_{\xi \in \mathcal{C}} |w_\xi| \leq c$ , with the understanding that this is an abbreviation for the concept described above with a sequence of quadrature formulas.

**Definition 2.1** *Let  $m \geq 0$  be an integer. The set  $\mathcal{C}$  admits an M-Z quadrature of order  $m$  if there exist weights  $w_\xi$  such that*

$$\int_{\mathbb{S}^q} P(\mathbf{x}) d\mu_q(\mathbf{x}) = \sum_{\xi \in \mathcal{C}} w_\xi P(\xi), \quad P \in \Pi_{2m}^q, \quad (2.10)$$

and

$$\left( \sum_{\xi \in \mathcal{C}} |w_\xi| |P(\xi)|^p \right)^{1/p} \leq c \|P\|_p, \quad P \in \Pi_{2m}^q, \quad 1 \leq p < \infty. \quad (2.11)$$

The weights  $w_\xi$  will be called *M-Z weights of order  $m$* . The condition (2.11) will be referred to as the *M-Z condition*.

If  $\mathcal{C}$  admits an M-Z quadrature of order  $m$ , and  $\{w_\xi\}$  are the weights involved, it is clear from using (2.11) with the polynomial identically equal to 1 in place of  $P$  that  $\sum_{\xi \in \mathcal{C}} |w_\xi| \leq c$ . Further, if  $\zeta \in \mathcal{C}$ , then applying (2.11) with  $p = 2$  and  $\Phi_m(h; \zeta \cdot \circ)$  in place of  $P$ , we obtain for M-Z weights of order  $m$ :

$$|w_\zeta| \Phi_m(h; 1)^2 \leq \sum_{\xi \in \mathcal{C}} |w_\xi| \Phi_m(h; \zeta \cdot \xi)^2 \leq c \int \Phi_m(h; \zeta \cdot \mathbf{x})^2 d\mu_q(\mathbf{x}).$$

The estimate (2.7) now implies that for all M-Z weights  $\{w_\xi\}$  of order  $m$ ,

$$|w_\xi| \leq cm^{-q}, \quad \xi \in \mathcal{C}. \quad (2.12)$$

In [21], we proved that every finite set  $\mathcal{C} \subset \mathbb{S}^q$  admits an M-Z quadrature with an order depending upon how dense the set  $\mathcal{C}$  is. This density is measured in terms of the mesh norm. The mesh norm of  $\mathcal{C}$  with respect to a subset  $K \subseteq \mathbb{S}^q$  is defined to be

$$\delta_{\mathcal{C}}(K) := \sup_{\mathbf{x} \in K} \text{dist}(\mathbf{x}, \mathcal{C}). \quad (2.13)$$

The following theorem summarizes the quadrature formula given in [21].

**Theorem 2.1** *There exists a constant  $\alpha_q$  with the following property. Let  $\mathcal{C}$  be a finite set of distinct points on  $\mathbb{S}^q$ , and  $m$  be an integer with  $m \leq \alpha_q(\delta_{\mathcal{C}}(\mathbb{S}^q))^{-1}$ . Then  $\mathcal{C}$  admits an M-Z quadrature of order  $m$ , and the set  $\{w_\xi\}$  of M-Z weights may be chosen to satisfy*

$$|\{\xi : w_\xi \neq 0\}| \sim m^q \sim \dim(\Pi_{2m}^q). \quad (2.14)$$

### 3 Polynomial operators

For  $t > 0$ , we define the summability operator  $\sigma_t^*$  by the formula

$$\sigma_t^*(h; f, \mathbf{x}) = \int_{\mathbb{S}^q} f(\zeta) \Phi_t(h; \mathbf{x} \cdot \zeta) d\mu_q(\zeta) = \sum_{\ell=0}^{\infty} h(\ell/t) \sum_{k=1}^{d_\ell^q} \hat{f}(\ell, k) Y_{\ell, k}(\mathbf{x}), \quad f \in L^1, \mathbf{x} \in \mathbb{S}^q. \quad (3.1)$$

(It is convenient, and customary in approximation theory, to use the notation  $\sigma_t^*(h; f, \mathbf{x})$  rather than  $\sigma_t^*(h; f)(\mathbf{x})$ .) Although we defined the operator for  $L^1$  to underline the fact that it is a universal operator, we will be interested only in its restriction to  $C(\mathbb{S}^q)$ . If  $f : \mathbb{S}^q \rightarrow \mathbb{R}$  is a continuous function, the degree of approximation of  $f$  from  $\Pi_x^q$  is defined by

$$E_x(f) = \inf_{P \in \Pi_x^q} \|f - P\|_\infty.$$

It is well known [16, 18] that for all integer  $n \geq 1$ , and  $f \in C(\mathbb{S}^q)$ ,

$$E_n(f) \leq \|f - \sigma_n^*(h; f)\|_\infty \leq cE_{n/2}(f). \quad (3.2)$$

Following [18], we now define a discretized version of these operators.

If  $\mathcal{C} \subset \mathbb{S}^q$  is a finite set,  $\mathbf{W} = \{w_\xi\}_{\xi \in \mathcal{C}}$  and  $\mathbf{Z} = \{z_\xi\}_{\xi \in \mathcal{C}}$  are sets of real numbers, we define the polynomial operator

$$\sigma_t(\mathcal{C}, \mathbf{W}; h; \mathbf{Z}, \mathbf{x}) := \sum_{\xi \in \mathcal{C}} w_\xi z_\xi \Phi_t(h; \mathbf{x} \cdot \xi), \quad t \in \mathbb{R}, \mathbf{x} \in \mathbb{S}^q. \quad (3.3)$$

If  $f : \mathbb{S}^q \rightarrow \mathbb{R}$ , and  $z_\xi = f(\xi)$ ,  $\xi \in \mathcal{C}$ , we will write  $\sigma_t(\mathcal{C}, \mathbf{W}; h; f, \mathbf{x})$  in place of  $\sigma_t(\mathcal{C}, \mathbf{W}; h; \mathbf{Z}, \mathbf{x})$ . In [18], we had denoted these operators by  $\sigma_t(\nu; h, f)$ , where  $\nu$  is the measure that associates the mass  $w_\xi$  with  $\xi \in \mathcal{C}$ . In this paper, we prefer to use the slightly expanded notation as in (3.3). If  $n \geq 1$  is an integer,  $\mathcal{C} \subset \mathbb{S}^q$  is a finite set that admits an M-Z quadrature of order  $n$ ,  $\mathbf{W}$  is the set of the corresponding M-Z weights. Then it is shown in [18, Proposition 4.1] that

$$E_n(f) \leq \|f - \sigma_n(\mathcal{C}, \mathbf{W}; h; f)\|_\infty \leq cE_{n/2}(f), \quad f \in C(\mathbb{S}^q). \quad (3.4)$$

In this paper, we will be especially interested in the approximation of functions in the class  $\mathbb{W}_r$ ,  $r > 0$ , comprised of functions  $f \in C(\mathbb{S}^q)$  for which  $E_n(f) = \mathcal{O}(n^{-r})$ ,  $n \geq 1$ . A complete characterization of the classes  $\mathbb{W}_r$  in terms of such constructive properties of its members as the number of partial derivatives and their moduli of smoothness is well known [26, 16]. In view of (3.2),  $f \in \mathbb{W}_r$  if and only if

$$\|f\|_{\mathbb{W}_r} := \|f\|_\infty + \sup_{n \geq 1} 2^{nr} \|\sigma_{2^n}^*(h; f) - \sigma_{2^{n-1}}^*(h; f)\|_\infty < \infty.$$

In practical applications, the data is contaminated with noise. Therefore, we wish to examine the behavior of our operators based on a data of the form  $\{(\xi, f(\xi) + \epsilon_\xi)\}$ , where  $\epsilon_\xi$  are independent random variables with unknown probability distributions, each with mean 0. If the range of these random variables is not bounded, one can still assume that the probability of the variables going out of a sufficiently large interval is small. Hence, it is customary in learning theory to assume that the variables  $\epsilon_\xi$  have a bounded range, so that one may use certain technical inequalities of probability theory, known as Bennett's inequalities; see the proof of Lemma 6.2 below.

In the statements of the theorems below, we use three parameters. The symbol  $M$  denotes the number of points in the data set; we assume that the set admits an  $M$ - $Z$  quadrature of order  $m$ , and the degree  $n$  of the polynomial approximant  $\sigma_n(\mathcal{C}, \mathbf{W}; h; f)$  is determined in terms of  $m$ . For theoretical considerations where one is not concerned about the actual numerical constructions of the quadrature weights, one may imagine a data set  $\mathcal{C}$  with  $M := |\mathcal{C}| \sim \delta_{\mathcal{C}}(\mathbb{S}^q)^{-q}$ , and assume that the weights  $\mathbf{W}$  are as guaranteed by Theorem 2.1. If so, then we may take  $M \sim m^q \sim \delta_{\mathcal{C}}(\mathbb{S}^q)^{-q}$  in the discussion in this section. For example, the estimates (3.7) and (3.8) below can then be expressed in terms of the number of samples respectively as follows:

$$\|f - \sigma_n(\mathcal{C}, \mathbf{W}; h; f)\|_{\infty} \leq cM^{-r/q}, \quad \text{with some } n \sim M^{1/q}, \quad (3.5)$$

and

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{Z}) - f\|_{\infty} \geq c_1 \frac{(\log M)^c}{M^{r/(q+2r)}} \right) \leq c_2 M^{-c}, \quad \text{with some } n \sim (M/\log M)^{1/(2r+q)}. \quad (3.6)$$

**Theorem 3.1** *Suppose that  $m \geq 1$  is an integer,  $\mathcal{C} = \{\xi_j\}_{j=1}^M$  admits an  $M$ - $Z$  quadrature of order  $m$ , and let  $\mathbf{W}$  be the corresponding quadrature weights. Let  $r > 0$ ,  $f \in \mathbb{W}_r$ ,  $\|f\|_{\mathbb{W}_r} = 1$ .*

(a) *For integer  $n \leq m$ , we have*

$$\|f - \sigma_n(\mathcal{C}, \mathbf{W}; h; f)\|_{\infty} \leq cn^{-r}. \quad (3.7)$$

(b) *For  $j = 1, \dots, M$ , let  $\epsilon_j$  be independent random variables with mean 0 and range  $[-1, 1]$ ,  $\mathbf{Z} = \{\epsilon_j + f(\xi_j)\}$ . If  $A > 0$  and  $n \geq 1$  is the greatest integer with  $(A + q)n^{2r+q} \log n \leq c_3 m^q$ ,  $n \leq m$ , then*

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{Z}) - f\|_{\infty} \geq c_1 n^{-r} \right) \leq c_2 n^{-A}. \quad (3.8)$$

Here, the constants  $c_1, c_2, c_3$  are independent of the distribution of the variables  $\epsilon_j$ .

We now turn our attention to local approximation by our operators. In the sequel, if  $K \subseteq \mathbb{S}^q$ ,  $f : K \rightarrow \mathbb{R}$ , then  $\|f\|_{\infty, K} := \sup_{\mathbf{x} \in K} |f(\mathbf{x})|$ . If  $\mathbf{x}_0 \in \mathbb{S}^q$ , a function  $f$  is defined to be  $r$ -smooth at  $\mathbf{x}_0$  if there is a spherical cap  $\mathbb{S}_{\delta}^q(\mathbf{x}_0)$  such that  $f\phi \in \mathbb{W}_r$  for every infinitely differentiable function  $\phi$  supported on  $\mathbb{S}_{\delta}^q(\mathbf{x}_0)$ . We have proved in [18, Theorem 3.3] that  $f$  is  $r$ -smooth at a point  $\mathbf{x}_0$  if and only if there is a cap  $\mathbb{S}_{\delta}^q(\mathbf{x}_0)$  such that

$$\|\sigma_{2^n}^*(h; f) - \sigma_{2^{n-1}}^*(h; f)\|_{\infty, \mathbb{S}_{\delta}^q(\mathbf{x}_0)} = \mathcal{O}(2^{-nr}).$$

Accordingly, if  $K$  is a spherical cap, we may define the class  $\mathbb{W}_r(K)$  to consist of  $f \in C(\mathbb{S}^q)$ , for which

$$\|f\|_{\mathbb{W}_r(K)} := \|f\|_{\infty} + \sup_{n \geq 1} 2^{nr} \|\sigma_{2^n}^*(h; f) - \sigma_{2^{n-1}}^*(h; f)\|_{\infty, K} < \infty.$$

**Theorem 3.2** *Suppose that  $m \geq 1$  is an integer,  $\mathcal{C} = \{\xi_j\}_{j=1}^M$  admits an  $M$ - $Z$  quadrature of order  $m$ , and let  $\mathbf{W}$  be the corresponding quadrature weights. Let  $0 < r \leq S - q$ ,  $K' \subset K$  be concentric spherical caps,  $f \in C(\mathbb{S}^q)$ ,  $\|f\|_{\mathbb{W}_r(K)} = 1$ .*

(a) *For integer  $n$ ,  $1 \leq n \leq m$ ,*

$$\|f - \sigma_n(\mathcal{C}, \mathbf{W}; h; f)\|_{\infty, K'} \leq cn^{-r}. \quad (3.9)$$

(b) *For  $j = 1, \dots, M$ , let  $\epsilon_j$  be independent random variables with mean 0 and range contained in  $[-1, 1]$ , and  $\mathbf{Z} = \{\epsilon_j + f(\xi_j)\}$ . If  $A > 0$  and  $n \geq 1$  is the greatest integer with  $(A + q)n^{2r+q} \log n \leq c_3 m^q$ ,  $n \leq m$ , then*

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{Z}) - f\|_{\infty, K'} \geq c_1 n^{-r} \right) \leq c_2 n^{-A}. \quad (3.10)$$

Here, the constants  $c_1, c_2, c_3$  are independent of the distribution of the variables  $\epsilon_j$ .

## 4 Construction of quadrature formulas

In this section, we describe two algorithms to obtain bounded variation quadrature formulas associated with a given finite set of points  $\mathcal{C} \subset \mathbb{S}^q$ . Both of these constructions can be described in a very general setting. Since this also simplifies the notations and ideas considerably by avoiding the use of real and imaginary parts of a doubly indexed polynomial  $Y_{\ell,k}$ , we will describe the algorithms in this generality.

Let  $\Omega$  be a nonempty set,  $\mu$  be a probability measure on  $\Omega$ ,  $\mathcal{C} \subset \Omega$ ,  $y_1, y_2, \dots$  be a complete orthonormal basis for  $L^2(\Omega, \mu)$ , where  $y_1 \equiv 1$ , and  $V_k$  denote the span of  $y_1, \dots, y_k$ . Let  $\nu$  be another measure on  $\Omega$ , and  $\langle \circ, \circ \rangle$  denote the inner product of  $L^2(\Omega, \nu)$ . For integer  $N \geq 1$ , the Gram matrix  $G_N$  is an  $N \times N$  matrix, defined by  $(G_N)_{\ell,k} = \langle y_\ell, y_k \rangle = (G_N)_{k,\ell}$ ,  $1 \leq k, \ell \leq N$ . We wish to find a weight function  $W$  on  $\Omega$  such that  $\int_\Omega P d\mu = \int_\Omega P W d\nu$  for all  $P \in V_N$  for an integer  $N$  for which  $G_N$  is positive definite.

For the applications to the case of quadrature formulas for the sphere,  $\Omega = \mathbb{S}^q$ ,  $\mu = \mu_q$ , and  $y_k$ 's are the orthogonal spherical harmonics, arranged in a sequence, so that  $y_1 \equiv 1$ , and all polynomials of lower degree are listed before those of a higher degree. To include all polynomials in  $\Pi_n^q$ , we need  $N = d_n^{q+1}$ . There are many possibilities to define the measure  $\nu$ . The simplest is the measure  $\nu^{MC}$  that associates the mass  $1/|\mathcal{C}|$  with each point of  $\mathcal{C}$ . A more sophisticated way to define the measure  $\nu$  is the following. We obtain a partition of  $\mathbb{S}^q$  into a dyadic triangulation such that each triangle contains at least one point of  $\mathcal{C}$ . We choose only one point in each triangle, and hence, assume that each triangle contains exactly one point of  $\mathcal{C}$ . We define the measure  $\nu^{TR}$  to be the measure that associates with each  $\xi \in \mathcal{C}$  the area of the triangle containing  $\xi$ .

One of the simplest ideas to compute the quadrature weights is the following. Let  $N$  be an integer for which  $G_N$  is positive definite. If  $P = \sum_{j=1}^N a_j y_j$ , then  $\int_\Omega P d\mu = a_1$ . Also, the vector  $\mathbf{a} = (a_1, \dots, a_N)^T$  satisfies the matrix equation

$$G_N \mathbf{a} = (\langle P, y_1 \rangle, \dots, \langle P, y_N \rangle)^T,$$

so that

$$\int_\Omega P d\mu = \sum_{k=1}^N (G_N)_{1,k}^{-1} \langle P, y_k \rangle = \langle P, \sum_k (G_N)_{1,k}^{-1} y_k \rangle. \quad (4.1)$$

In the setting of the sphere, this gives the following quadrature formula:

$$\int_{\mathbb{S}^q} P d\mu_q = \sum_{\xi \in \mathcal{C}} P(\xi) \left\{ \nu(\{\xi\}) \sum_{k=1}^N (G_N)_{1,k}^{-1} y_k(\xi) \right\} =: \sum_{\xi \in \mathcal{C}} w_\xi^{LSQ} P(\xi). \quad (4.2)$$

We formulate this as

ALGORITHM LSQ

**Input:** The matrix  $Y = (y_k(\xi))$ ,  $k = 1, \dots, N$  (optional), and the vector  $\mathbf{v} = (\nu(\{\xi\}))$ .

1. Solve  $Y \mathbf{diag}(\mathbf{v}) Y^T \mathbf{b} = (1, 0, \dots, 0)^T$ .
2. Return  $w_\xi^{LSQ} = \nu(\{\xi\}) \sum_{k=1}^N b_k y_k(\xi)$ .

We observe that  $G_N = Y \mathbf{diag}(\mathbf{v}) Y^T$ . It is clear that the matrix  $G_N$  is always positive semi-definite; the assumption that it is positive definite is equivalent to the assumption that no element of  $V_N$  vanishes identically on  $\mathcal{C}$ . If  $\mathcal{C}$  and  $\{\nu(\{\xi\})\}$  satisfy the M-Z inequalities, Theorem 4.1 below shows that  $G_N$  is well conditioned. Assuming that the matrix  $Y$  is input, the time to compute  $G_N$  is  $\mathcal{O}(N^2|\mathcal{C}|)$  and the space requirement is  $\mathcal{O}(N^2)$ . (In the case of the sphere  $\mathbb{S}^q$ , we need  $N = d_n^{q+1} = \mathcal{O}(n^q)$  to compute formulas exact for degree  $n$ .) The vector  $\mathbf{b}$  in Step 1 can be found using such iterative methods as the conjugate residual method. We refer to [6] for a more detailed analysis of this method. Using this approach, the matrix  $Y$  and  $G_N$  need not be stored or precomputed, but the product of the matrix  $G_N$  with an arbitrary residual vector  $\mathbf{r}$  needs to be computed. This observation results in a substantial saving in the time and memory complexity of the algorithm when the results are desired only within a given accuracy. For the unit sphere  $\mathbb{S}^2$ , when  $N = (n+1)^2$ , the product  $G_N \mathbf{r} = Y \mathbf{diag}(\mathbf{v}) Y^T \mathbf{r}$  can be computed within an accuracy  $\epsilon$  using a recent algorithm of Keiner [14] using  $\mathcal{O}(n^2(\log n)^2 + \log(1/\epsilon)|\mathcal{C}|)$  operations, where  $\epsilon$  is the accuracy of the method.

One way to interpret this algorithm is the following. Let  $f : \mathbb{S}^q \rightarrow \mathbb{R}$ , and  $P$  be the solution to the least square problem

$$P = \arg \min \{ \langle f - Q, f - Q \rangle : Q \in V_N \}.$$

If  $\mathbf{f}$  is the vector  $(\langle f, y_j \rangle)$ , then  $P = \sum_j (G_N^{-1} \mathbf{f})_j y_j$ . The quadrature formula with weights  $w_\xi^{LSQ}$  thus offers  $\int_{\mathbb{S}^q} P d\mu_q$  as the approximation to  $\int_{\mathbb{S}^q} f d\mu_q$ . The weights  $w_\xi^{LSQ}$  also satisfy a least square property among all the possible quadrature formulas, as shown in Lemma 6.1(a). We summarize some of the properties of the weights  $w_\xi^{LSQ}$  in the following theorem.

**Theorem 4.1** *Let  $n \geq 1$  be an integer,  $N = d_n^{q+1}$ ,  $\mathcal{C}$  be a finite set of points on  $\mathbb{S}^q$  and  $\nu$  be a measure supported on  $\mathcal{C}$ . Let  $v_\xi := \nu(\{\xi\})$ ,  $\xi \in \mathcal{C}$ , and*

$$c_1 \|P\|_p \leq \left\{ \sum_{\xi \in \mathcal{C}} v_\xi |P(\xi)|^p \right\}^{1/p} \leq c_2 \|P\|_p, \quad P \in \Pi_n^q, \quad 1 \leq p \leq \infty, \quad (4.3)$$

(a) *For the Gram matrix  $G_N$ , the lowest eigenvalue is  $\geq c_1^2$ , and the largest eigenvalue is  $\leq c_2^2$ , where  $c_1, c_2$  are the constants in (4.3) with  $p = 2$ . In particular,  $G_N$  is positive definite. Moreover,  $\sum_{\xi \in \mathcal{C}} |w_\xi^{LSQ}| \leq c$ .*  
(b) *If*

$$\left| \int P^2 d\nu - \int P^2 d\mu_q \right| \leq \frac{c}{n^q} \int P^2 d\mu_q, \quad P \in \Pi_n^q, \quad (4.4)$$

*then  $|w_\xi^{LSQ}| \leq cv_\xi$ ,  $\xi \in \mathcal{C}$ . In particular, the weights  $\{w_\xi^{LSQ}\}$  satisfy the M-Z condition.*

(c) *Let  $M \geq 1$  be an integer,  $\mathcal{C}$  be a independent random sample of  $M$  points chosen from the distribution  $\mu_q$ , and  $A, \eta > 0$ . Let  $v_\xi = 1/M$ ,  $\xi \in \mathcal{C}$ . There exists a constant  $c = c(A)$  such that if  $n \geq 2$  is an integer with  $M \geq cn^q \log n / \eta^2$ , then*

$$\text{Prob} \left( \left| \int P^2 d\nu - \int P^2 d\mu_q \right| \geq \eta \int P^2 d\mu_q, \quad P \in \Pi_n^q \right) \leq c_1 n^{-A}. \quad (4.5)$$

*In particular, if  $M \geq cn^{3q} \log n$ , then the condition (4.4) is satisfied with probability exceeding  $1 - c_1 n^{-A}$ .*

One disadvantage of the algorithm LSQ is that one needs to know the value of  $N$  in advance. We now describe an idea which has the potential to avoid this problem. In the case when  $\Omega$  is a subset of a Euclidean space, and the  $y_j$ 's are polynomials, with  $y_1$  denoting the constant polynomial, one can construct a system  $\{t_k\}$  of orthonormalized polynomials with respect to  $\nu$  using recurrence relations. Recurrence relations for orthogonal polynomials in several variables have been discussed in detail by Dunkl and Xu [4, Chapter 3]. In contrast to the viewpoint in [4], we may depend upon a specific enumeration, but require the recurrence relation to have a specific form described in Theorem 4.2 below. This form allows us to generalize the ideas in Gautschi's book [11, Chapter 2] in our context.

To describe our ideas in general, let  $\Omega \subset \mathbb{R}^{q+1}$ ,  $u_1, u_2, \dots$  be the lexicographic enumeration of the monomials in  $q+1$  variables, so that  $u_1$  is the monomial identically equal to 1, the restrictions of  $u_k$ 's to  $\Omega$  are linearly independent, and  $V_k = \text{span} \{u_1, \dots, u_k\}$ . It is not difficult to see that for every integer  $k \geq 1$ , there is a minimal index  $p(k)$  such that there exists a monomial  $\tilde{f}_k$  of degree 1 with

$$\tilde{f}_k u_{p(k)} = u_{k+1}, \quad k = 1, 2, \dots \quad (4.6)$$

We now let, for each  $k = 1, 2, \dots$ ,  $\{y_1, \dots, y_k\}$  be a basis for  $V_k$  orthonormal with respect to  $\mu$ ,  $N \geq 1$  be an integer for which the Gram matrix  $G_N$  is positive definite, and for each  $k = 1, \dots, N$ ,  $\{t_1, \dots, t_k\}$  be a basis for  $V_k$  orthonormal with respect to  $\nu$ . Clearly, any polynomial  $P \in V_N$  can be written in the form

$$P(\mathbf{x}) = \int P(\zeta) \sum_k t_k(\mathbf{x}) t_k(\zeta) d\nu(\zeta),$$

and consequently, one gets the "quadrature formula"

$$\int P(\mathbf{x}) d\mu(\mathbf{x}) = \int P(\zeta) \left\{ \sum_k \left( \int t_k(\mathbf{x}) d\mu(\mathbf{x}) \right) t_k(\zeta) \right\} d\nu(\zeta). \quad (4.7)$$

In this discussion only, let  $t_k =: \sum_j c_{k,j} y_j$ , and the matrix  $(c_{k,j})$  be denoted by  $C$ . The condition that  $t_1, \dots, t_N$  is an orthonormal system with respect to  $\nu$  is equivalent to the condition that  $CG_N C^T = I$ , where  $I$  is the  $N \times N$  identity matrix. Hence,  $G_N^{-1} = C^T C$ . Moreover,  $\int t_k d\mu = c_{k,1}$  for  $k = 1, \dots, N$ , and hence, we conclude that

$$\sum_k \left( \int t_k(\mathbf{x}) d\mu(\mathbf{x}) \right) t_k = \sum_j \sum_k c_{k,1} c_{k,j} y_j = \sum_j (G_N)_{1,j}^{-1} y_j.$$

Thus, the quadrature weights in (4.7) are the same as those in (4.1).

First, we summarize the various recurrence relations in Theorem 4.2 below, although we will not use all of them. We will denote the (total) degree of  $u_k$  by  $D_k$ , and observe that  $D_k$  is also the degree of  $y_k$  and  $t_k$ ,  $D_j \leq j$ , and  $D_{p(k)} = D_{k+1} - 1$ .

**Theorem 4.2** *There exist real numbers  $s_{k,j}$ ,  $\tilde{r}_{k,j}$ ,  $A_k \geq 0$ , and a linear polynomial  $f_k$ , such that*

$$f_k y_{p(k)} = y_{k+1} - \sum_{\substack{D_{k+1}-2 \leq D_j \leq D_k \\ j \leq k}} \tilde{r}_{k,j} y_j, \quad f_k t_{p(k)} = A_k t_{k+1} - \sum_{\substack{D_{k+1}-2 \leq D_j \leq D_k \\ j \leq k}} s_{k,j} t_j. \quad (4.8)$$

More generally, if  $P$  is any linear polynomial, there exist real numbers  $r_{k,j}(P)$  such that

$$P y_k = \sum_{D_{k-1} \leq D_j \leq D_{k+1}} r_{k,j}(P) y_j, \quad (4.9)$$

We have  $t_k = \sum_j c_{k,j} y_j$ , where

$$A_k c_{k+1,\ell} = \left\{ \sum_{D_{\ell-1} \leq D_m \leq D_{\ell+1}} r_{\ell,m}(f_k) c_{p(k),m} + \sum_{D_{k+1}-2 \leq D_j \leq D_k} s_{k,j} c_{j,\ell} \right\}. \quad (4.10)$$

In the context of the sphere  $\mathbb{S}^q$ , we will compute  $t_k$ 's using (4.8), and compute  $\int t_k d\mu_q$  using a known quadrature formula. The resulting algorithm, Algorithm REC, in the context of the sphere is summarized below. This algorithm is similar to the Stieltjes method in Gautschi's book [11, Section 2.2]. Even though it is feasible to carry out the algorithm for as large an  $N$  as the data allows, and to find this value of  $N$  during run time, it is still desirable from the point of view of numerical stability to limit the largest  $N$  from the outset. Accordingly, in describing the following algorithm, we stipulate that the quadrature formula is to be computed to be exact only for polynomials in  $V_N$  for the largest possible  $N \leq L$  for some integer  $L \geq 1$ . We assume further that we know another quadrature formula (for example, the Driscoll–Healy formula [3]) exact for polynomials in  $V_L$ :

$$\sum_{\zeta \in \mathcal{C}^*} \lambda_\zeta P(\zeta) = \int P d\mu_q, \quad P \in V_L. \quad (4.11)$$

#### ALGORITHM REC

**Input:** An integer  $L$ , the sequence  $p(k)$ ,  $k = 1, \dots, L$ , sets  $\mathcal{C}$ ,  $\mathcal{C}^*$ , weights  $(\lambda_\zeta)_{\zeta \in \mathcal{C}^*}$  so that (4.11) holds, the values  $\{y_j(\xi)\}_{\xi \in \mathcal{C}}$ ,  $\{y_j(\zeta)\}_{\zeta \in \mathcal{C}^*}$  for  $j = 1, 2, 3, 4$ , and the values  $f_k(\xi)$ ,  $f_k(\zeta)$ ,  $k = 1, \dots, L$ .

1. Using Gram–Schmidt procedure, initialize  $t_1, t_2, t_3, t_4$ , both for points in  $\mathcal{C}$  and in  $\mathcal{C}^*$ , and initialize  $N = 4$ .
2. For  $k = 1, \dots, 4$ , let  $\gamma_k = \sum_{\zeta \in \mathcal{C}^*} \lambda_\zeta t_k(\zeta)$ .
3. For each  $\xi \in \mathcal{C}$ , initialize  $w_\xi = \sum_{k=1}^4 \gamma_k t_k(\xi)$ .
4. For  $k = 4, 5, \dots$  (so that the degrees are at least 0 for all polynomials entering in the recursions) and while  $N \leq L$ , repeat steps 5–8 below.

5. For  $j$  with  $D_{k+1} - 2 \leq D_j \leq D_k$ , set

$$s_{k,j} = \langle f_k t_{p(k)}, t_j \rangle.$$

6. Define  $T_{k+1}$  by

$$T_{k+1} = f_k t_{p(k)} - \sum_{D_{k+1}-2 \leq D_j \leq D_k} s_{k,j} t_j.$$

both for points in  $\mathcal{C}$  and points in  $\mathcal{C}^*$ . If  $I_{k+1} = \langle T_{k+1}, T_{k+1} \rangle = 0$ , then stop, and set  $N = k$ . Otherwise, define  $t_{k+1} = T_{k+1}/I_{k+1}^{1/2}$ .

7. Set  $\gamma_{k+1} = \sum_{\zeta \in \mathcal{C}^*} \lambda_\zeta t_{k+1}(\zeta)$ .

8. For each  $\xi \in \mathcal{C}$ ,  $w_\xi = w_\xi + \gamma_{k+1} t_{k+1}(\xi)$ ,  $k = k + 1$ ,  $N = N + 1$ .

In the case of the sphere  $\mathbb{S}^q$ , we take  $L = d_{\tilde{n}}^{q+1}$  for some integer  $\tilde{n} \geq 1$ . The number of  $j$ 's with  $D_{k+1} - 2 \leq D_j \leq D_k$ ,  $1 \leq j, k \leq L$  is  $\mathcal{O}(\tilde{n}^{q-1})$ . In this discussion only, let  $M = |\mathcal{C}| + |\mathcal{C}^*|$ . Consequently, Steps 5 and 6 require  $\mathcal{O}(M\tilde{n}^{q-1})$  operations. Since the remaining two steps in the loop take  $\mathcal{O}(M)$  operations, the loop starting at Step 4 require  $\mathcal{O}(M\tilde{n}^{2q-1})$  operations. Finally, we observe that in implementing the above algorithm, one need not keep the whole matrix  $t_k(\xi)$ ; only the rows corresponding to three degrees are required in any step. In particular, the memory requirement of this algorithm is  $\mathcal{O}(M\tilde{n}^{q-1})$ .

## 5 Numerical experiments

The objective of this section is to demonstrate and supplement the theoretical results presented in Sections 3 and 4.

Our first set of experiments illustrates the algorithms LSQ and REC. The experiments were conducted over a long period of time, many of them long before we started to write the paper. Hence, the normalizations for the spherical polynomials  $Y_{\ell,k}$  are somewhat different in Tables 1 and 2 from the rest of the paper. This is reflected in the sum of the absolute values of the weights, but has no effect on the various results other than scaling.

First, we report on the algorithm LSQ. Each of the experiments in this case was repeated 30 times with data sets chosen randomly from the distribution  $\mu_2$  on  $\mathbb{S}^2$ . To test our algorithms, we computed the computed Gram matrix  $G^{COM}$  given by

$$G_{\ell,m}^{COM} = \sum_{\xi \in \mathcal{C}} w_\xi^{LSQ} y_\ell(\xi) y_m(\xi), \quad \ell, m < \lfloor n/2 \rfloor.$$

The average maximum matrix norm of the difference between  $G^{COM}$  and the identity matrix of the same size indicates the error of the quadrature formulas. The results are shown in Table 1. Based on these results we conjecture that in order to obtain stable quadrature formulas (i.e., with small condition number for the original Gram matrix  $G_N$ ) exact for degree  $n \geq 1$ , one has to use at most  $4d_n^{q+1}$  uniformly distributed points. In contrast, the theoretical guarantee in Theorem 4.1(c) requires  $\mathcal{O}(n^{3q} \log n)$  points.

As can be seen from the table, for a fixed degree  $n$ , the condition number  $\kappa(G_N)$  decreases as the number of points increases. For  $n > 140$  and various sets of randomly generated points on the sphere, we do not obtain good numerical results. This might be due to a defect in the built in numerical procedures used by Matlab in computing the spherical harmonics of high degree at values close to  $-1$  or  $1$ . The situation was much better for the dyadic points; i.e., the centers of the dyadic triangles.

For dyadic points on the sphere, the best result we obtained so far is  $n = 178$  with 131,072 points. As a further verification of this quadrature, we considered the following data. The data are constructed using coefficients  $\{a_{\ell,k}\}$  for spherical polynomials up to degree 90, taken from model MF4 used for modelling the lithospheric field. The model is computed by geophysicists at GeoForschungsZentrum

$M$	$n$	$Error$	$\sum  w_\xi $	$\min w_\xi$	$\max w_\xi$	pos	$\kappa(G_N)$	$\lambda_{\min}$	$\lambda_{\max}$
8192	16	$2.41 * 10^{-15}$	3.5449	$2.29 * 10^{-4}$	$7.88 * 10^{-4}$	8192	2.43	0.607	1.4730
	44	$4.32 * 10^{-15}$	3.5714	$-5.06 * 10^{-4}$	0.0029	8039	37.52	0.078	2.8047
	64	$6.15 * 10^{-15}$	5.5575	-0.00664	0.0073	6068	1695.1	0.003	3.9315
	84	$9.73 * 10^{-12}$	82.152	-0.16274	0.1551	4431	$3.52 * 10^6$	$2.6 * 10^{-6}$	5.4851
16384	44	$4.43 * 10^{-15}$	3.5449	$-9.50 * 10^{-6}$	$8.82 * 10^{-4}$	16382	9.19	0.24036	2.1590
	64	$5.25 * 10^{-15}$	3.5787	$-3.75 * 10^{-4}$	0.0015	16014	51.9	0.05964	2.9150
	84	$7.10 * 10^{-15}$	4.4757	-0.0024	0.0032	13361	944.86	0.00612	3.8457
	100	$1.94 * 10^{-15}$	9.1325	-0.0077	0.0072	10625	11896.1	$4.8 * 10^{-4}$	4.6008
32768	44	$6.02 * 10^{-15}$	3.5449	$3.11 * 10^{-5}$	$2.90 * 10^{-4}$	32768	4.270	0.4157	1.7652
	64	$7.09 * 10^{-15}$	3.5450	$-1.79 * 10^{-5}$	$5.23 * 10^{-4}$	32761	7.977	0.8208	5.2519
	84	$7.71 * 10^{-15}$	3.5574	$-1.43 * 10^{-4}$	$7.92 * 10^{-4}$	32410	42.97	0.0716	2.7967
	100	$7.62 * 10^{-15}$	3.6777	$-4.28 * 10^{-4}$	$9.96 * 10^{-4}$	30819	145.6	0.0250	3.2967

Table 1: The statistics for the experiments with the algorithm LSQ.  $M = |\mathcal{C}|$ ,  $n - 2$  is the degree of spherical polynomials for which exact quadrature formulas were computed,  $N = n^2$ , pos stands for the number of positive weights,  $\kappa(G_N)$ ,  $\lambda_{\min}$ ,  $\lambda_{\max}$  are the condition number, the maximum eigenvalue and the minimum eigenvalue of the matrix  $G_N$  respectively.

Potsdam (Germany) based on CHAMP satellite data. We use those coefficients to construct the samples of a function  $f = \sum_{\ell,k} a_{\ell,k} Y_{\ell,k}$  at the centers of  $8 * 4^7$  dyadic triangles. We then use our pre-computed quadrature based at these centers which can integrate spherical polynomials up to degree 178 to compute the Fourier coefficients  $\hat{a}_{\ell,k}$ . The maximum difference between the vector  $\{\hat{a}_{\ell,k}\}$  and the vector  $\{a_{\ell,k}\}$  was found to be  $6.66 * 10^{-15}$ .

Next, we considered the algorithm REC. In the context of spherical polynomials, the recurrence relations have to be chosen very carefully using the special function properties of the spherical harmonics  $Y_{\ell,k}$ , in order to get stable results [28]. In the present situation, the polynomials  $t_k$  have no special structure. Therefore, it turns out that the algorithm REC is not very stable for high degrees. However, when we took the centers of 8192 dyadic triangles as the quadrature nodes, and used the measure  $\nu^{TR}$  as the starting measure, then we are able to obtain satisfactory quadrature formulas for degree 32. We note an interesting feature here that all the weights obtained by this algorithm are positive. These results are summarized in Table 2 below.

$n$	Error	$\min(w_\xi)$	$\max(w_\xi)$	$\sum w_\xi$
16	$4.196643 * 10^{-14}$	$5.181468 * 10^{-4}$	$2.538441 * 10^{-3}$	12.56637
22	$5.302425 * 10^{-13}$	$5.175583 * 10^{-4}$	$2.543318 * 10^{-3}$	12.56637
32	$9.240386 * 10^{-11}$	$5.154855 * 10^{-4}$	$2.544376 * 10^{-3}$	12.56637
42	$4.434868 * 10^{-8}$	$5.086157 * 10^{-4}$	$2.544141 * 10^{-3}$	12.56637
44	$2.320896 * 10^{-5}$	$5.094771 * 10^{-4}$	$2.562948 * 10^{-3}$	12.56638

Table 2: Quadrature constructed using REC on 8192 dyadic points

Our second set of experiments demonstrates the local approximation properties of the operators  $\sigma_n(\mathcal{C}, \mathbf{W}; h)$  for a smooth function  $h$ . For this purpose, we consider the following benchmark functions, considered by various authors [31, 30, 15, 10], listed in (5.1) below. Using the notation  $\mathbf{x} = (x_1, x_2, x_3)$ , the functions are defined by

$$\begin{aligned}
g_1(\mathbf{x}) &= (x_1 - 0.9)_+^{3/4} + (x_3 - 0.9)_+^{3/4}, \\
g_2(\mathbf{x}) &= [0.01 - (x_1^2 + x_2^2 + (x_3 - 1)^2)]_+ + \exp(x_1 + x_2 + x_3), \\
g_3(\mathbf{x}) &= 1/(101 - 100x_3), \\
g_4(\mathbf{x}) &= 1/(|x_1| + |x_2| + |x_3|),
\end{aligned}$$

$$g_5(\mathbf{x}) = \begin{cases} \cos^2\left(\frac{3\pi}{2}\text{dist}(\mathbf{x}, (-1/2, -1/2, 1/\sqrt{2}))\right), & \text{if } \text{dist}(\mathbf{x}, (-1/2, -1/2, 1/\sqrt{2})) < 1/3, \\ 0, & \text{if } \text{dist}(\mathbf{x}, (-1/2, -1/2, 1/\sqrt{2})) \geq 1/3. \end{cases} \quad (5.1)$$

In order to define the function  $h$ , we recall first that the  $B$  spline  $B_m$  of order  $m$  is defined recursively [1, p. 131] by

$$B_m(x) := \begin{cases} 1, & \text{if } m = 1, 0 < x \leq 1, \\ 0, & \text{if } m = 1, x \in \mathbb{R} \setminus (0, 1], \\ \frac{x}{m-1}B_{m-1}(x) + \frac{m-x}{m-1}B_{m-1}(x-1), & \text{if } m > 1, x \in \mathbb{R}. \end{cases} \quad (5.2)$$

The function  $B_m$  is an  $m-1$  times iterated integral of a function of bounded variation. We will choose  $h$  to be

$$h_m(x) = \sum_{k=-m}^m B_m(2mx - k), \quad (5.3)$$

for different values of  $m$ , in order to illustrate the effect of the smoothness of  $h_m$  on the quality of local approximation. If  $m \geq 3$ , the function  $h_m$  satisfies the conditions in Proposition 2.1 with  $S = m-1$ . We note that the discretized Fourier projection operator  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$  has been called the hyperinterpolation operator [29].

One example of the localization properties of our operators is given in the following table, where we show the error in approximation of  $g_1$  on the whole sphere and on the cap  $K = \mathbb{S}_{0.4510}^2((-1/\sqrt{2}, 0, -1/\sqrt{2})$ . The operators were constructed using Driscoll–Healy quadrature formula [3] based on  $4(n+1)^2$  points, exact for integrating polynomials of degree  $2n$ . The maximum error on the whole sphere, given in Columns 2 and 3, is estimated by the error at 10000 randomly chosen points; that on the cap, given in Columns 4 and 5, is estimated by the error at 1000 randomly chosen points on the cap. It is clear that even though the maximum error on the whole sphere is slightly better for the (discretized) Fourier projection than for our summability operator, the singularities of  $g_1$  continue to dominate the error in the Fourier projection on a cap away from these singularities; the performance of our summability operator is far superior.

n	$S2errh1$	$S2errh5$	$Kerrh1$	$Kerrh5$
63	0.0097	0.0112	$3.4351 * 10^{-4}$	$6.5926 * 10^{-7}$
127	0.0044	0.0055	$8.0596 * 10^{-5}$	$6.5240 * 10^{-8}$
255	0.0033	0.0038	$1.4170 * 10^{-5}$	$1.1816 * 10^{-8}$

Table 3:  $S2errh1 = \max_{\mathbf{x} \in \mathbb{S}^2} |g_1(\mathbf{x}) - \sigma_n(\mathcal{C}, \mathbf{W}; h_1, g_1, \mathbf{x})|$ ,  $S2errh5 = \max_{\mathbf{x} \in \mathbb{S}^2} |g_1(\mathbf{x}) - \sigma_n(\mathcal{C}, \mathbf{W}; h_5, g_1, \mathbf{x})|$ ,  $Kerrh1 = \max_{\mathbf{x} \in K} |g_1(\mathbf{x}) - \sigma_n(\mathcal{C}, \mathbf{W}; h_1, g_1, \mathbf{x})|$ ,  $Kerrh5 = \max_{\mathbf{x} \in K} |g_1(\mathbf{x}) - \sigma_n(\mathcal{C}, \mathbf{W}; h_5, g_1, \mathbf{x})|$ ,  $(\mathcal{C}, \mathbf{W})$  are given by the Driscoll–Healy formulas.

Theorem 3.2 points out another way to demonstrate the superior localization of our summability operator without an a priori knowledge of the locations of the singularities. Since each of the test functions is infinitely differentiable on large caps of different sizes, Theorem 3.2 suggests that the more localized the method, the greater is the probability that the approximation error would be smaller than a given number. To demonstrate also how our ability to construct quadrature formulas based on scattered data helps us to analyse the approximation properties of our summability operators, we took for the set  $\mathcal{C}$  a randomly generated sample of 65536 points. For these points, the weights  $\mathbf{W}$  computed by the algorithm LSQ yield a quadrature formula exact for integrating spherical polynomials of degree 126. We compare three approximation methods, the least square approximation from  $\Pi_{63}^2$ , the approximation given by the operator  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$ , and the approximation given by  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$ . For each function, we computed the absolute value of the difference between the approximate value computed by each of the three methods and the true value of the function at 20,000 randomly chosen points on the sphere. The percentage of points where the value of this difference is less than  $10^{-x}$  is reported in Table 4 below, for  $x = 2 : 10$ . It is very obvious that  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$  gives a far superior performance than the other methods, due to its localization properties.

	$x \rightarrow$	10	9	8	7	6	5	4	3	2
$g_1$	S1	0	0	0.005	0.02	0.42	4.44	39.43	94.79	100
	LS	0	0	0	0.04	0.56	5.32	46.38	95.45	100
	S5	0.02	0.19	1.87	16.89	59.36	68.34	79.01	93.09	99.97
$g_2$	S1	0	0.01	0.09	0.74	7.94	84.99	99.19	99.99	100
	LS	0	0.01	0.15	1.29	13.29	86.09	99.28	100	100
	S5	0.39	3.34	41.95	90.78	94.52	97.19	99.18	99.97	100
$g_3$	S1	0	0.01	0.11	1.26	12.02	91.87	99.86	100	100
	LS	0	0.01	0.10	1.49	16.26	93.12	99.87	100	100
	S5	0.51	5.43	51.08	82.22	91.90	95.79	98.49	99.87	100
$g_4$	S1	0	0	0	0.01	0.18	1.91	18.28	83.81	99.97
	LS	0	0	0.01	0.02	0.25	2.16	21.24	86.43	99.98
	S5	0.01	0.01	0.04	0.36	3.47	17.48	40.98	80.06	99.88
$g_5$	S1	0	0	0.01	0.09	1.12	11.84	88.94	99.45	100
	LS	0	0.01	0.01	0.15	1.42	15.23	90.47	99.75	100
	S5	0.08	0.64	5.73	66.82	83.54	88.74	92.95	96.78	97.64

Table 4: Percentages of error less than  $10^{-x}$  for different functions, LS= Least square, S1= error with  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$ , S5= error with  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$ . For example, for the function  $g_3$ , S5 was less than  $10^{-7}$  for 82.22% of the 20000 randomly selected points, while S1 (respectively, LS) was less than  $10^{-7}$  for 1.26% (respectively, 1.49%) points.

$x \rightarrow$	5	4	3	2	3.0	2.75	2.5	2.25
S1	0.05	0.635	9.93	97.45	0	7.97	92.75	100.00
LS	0	0	0	100	0	0.04	30.93	99.07
S5	0.085	1.015	10.03	97.49	0.24	51.97	99.87	100.00

Table 5: Percentages of error less than  $10^{-x}$  for  $\epsilon = 0.01$ , LS= Least square, S1= error with  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$ , S5= error with  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$ . The random noise in the left half comes from the uniform distribution in  $[-\epsilon, \epsilon]$ , that in the right half from the normal distribution with mean 0, standard deviation  $\epsilon$ .

Next, we illustrate the stability of our operators under noise. Since our operators are linear operators, we assume for this part of the study that the target function  $f$  is the zero function contaminated either by uniform random noise in the range  $[-\epsilon, \epsilon]$ , or a normally distributed random variable with mean 0 and standard deviation  $\epsilon$ . We let  $\mathcal{C}$  be a set of 65536 random points and computed corresponding weights  $\mathbf{W}$  that integrate exactly polynomial up to degree 126. These were used in calculating  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$  and  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$  at each point of a test data set consisting of 20000 random samples from the distribution  $\mu_2$ . For each value of  $\epsilon = 0.1; 0.01; 0.001; 0.0001$ , the experiment is repeated 50 times and the errors are the averaged over the number of repetitions. The percentage of points at which the absolute computed value is less than  $10^{-x}$  is reported in Table 5 in the case when  $\epsilon = 0.01$ . The results for the other values of  $\epsilon$  were consistent with the linearity of the operator. We observe that in each case, both  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$  and  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$  yield better results than the least squared approximation, while  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_5)$  is slightly superior to  $\sigma_{63}(\mathcal{C}, \mathbf{W}; h_1)$ .

Finally, we used our operator  $\sigma_{22}(\mathcal{C}, \mathbf{W}; h_7)$  with the MAGSAT data. Our purpose here is only to test how our methods work on a “real life” data. This data, supplied to us kindly by Dr. Thorsten Maier, measures the magnetic field of the earth in nT as a vector field. It was derived from vectorial MAGSAT morning data that has been processed by Nils Olsen of the Danish Space Research Institute. The measurements are averaged on a longitude-latitude grid with  $\Delta\phi = 4^\circ$  and  $\Delta\theta = 2^\circ$  in geomagnetic coordinates. The radial variations of the MAGSAT satellite have been neglected in the dataset and, therefore, prior to the averaging process, the GSFC(12/83) reference potential model has been subtracted. The data results from one month of measurements, centered at March 21, 1980. We extract the East West component of the vectorial data as a scalar valued function on the sphere. Totally, there are 8190

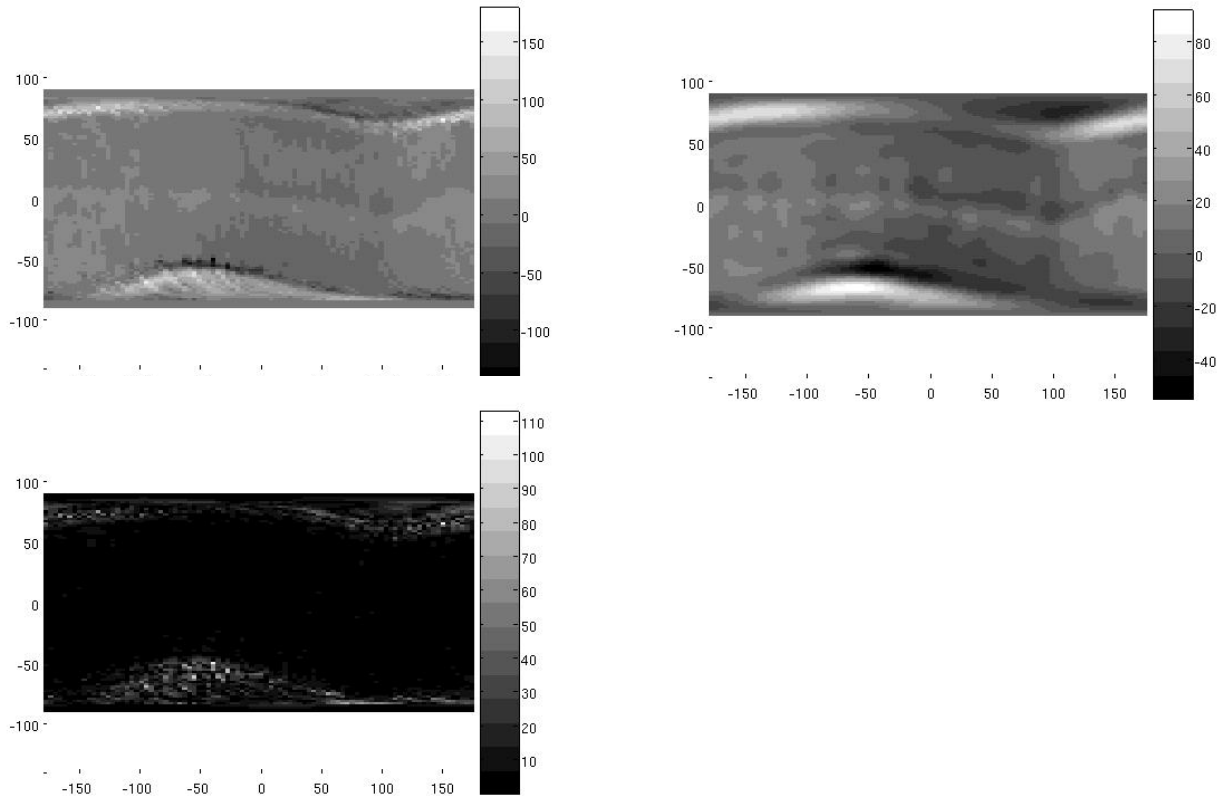


Figure 2: From left to right: The original data, its reconstruction using  $\sigma_{22}(\mathcal{C}, \mathbf{W}; h_7)$ , and the error in the approximation,  $|\sigma_{22}(\mathcal{C}, \mathbf{W}; h_7) - y|$ .

data sites. A quadrature of degree 44 was computed based on those sites. Figure 2 shows the original data, its reconstruction using  $\sigma_{22}(\mathcal{C}, \mathbf{W}; h_7)$ , and the error in the approximation,  $|\sigma_{22}(\mathcal{C}, \mathbf{W}; h_7) - y|$ , as a map in the longitude-latitude plane. As can be seen from the figures, the reconstruction preserves the key features of the original data.

## 6 Proofs

In the interest of organization, we will prove the various new results in the paper in the following order. We will prove Theorem 4.2 first, since its proof does not require any preparation. We will then use Proposition 2.1 to prove Theorems 3.1(a) and 3.2(a). Next, we will prove Lemma 6.1 and use it to prove parts (a) and (b) of Theorem 4.1. The remaining results in this paper involve probabilities. We prove Lemma 6.2 next, estimating the probability that the supremum norm of a sum of random spherical polynomials exceeds a given number. This lemma will be used immediately to prove Theorem 4.1(c). Finally, we will prove Theorems 3.1(b), 3.2(b).

PROOF OF THEOREM 4.2. It is convenient to prove (4.9) first. Since  $Py_k$  is a polynomial of degree  $D_k + 1$ , there exist real numbers  $r_{k,j}(P)$  such that

$$Py_k = \sum_{D_j \leq D_k + 1} r_{k,j}(P) y_j.$$

Since the system  $\{y_k\}$  is orthonormal with respect to  $\mu$ ,

$$r_{k,j}(P) = \int_{\Omega} Py_k y_j d\mu.$$

If  $D_j < D_k - 1$ , then the degree of  $P y_j$  is less than  $D_k$ . Because of the lexicographic ordering where lower degree polynomials appear before the higher degree ones, this implies that  $P y_j \in V_{k-1}$ . Since  $y_k$  is orthogonal to  $V_{k-1}$ , it follows that  $r_{j,k}(P) = 0$  if  $D_j < D_k - 1$ . This completes the proof of (4.9).

We observe that  $y_{p(k)} \in \text{span} \{u_1, \dots, u_{p(k)}\}$ . So, there exists a constant  $\alpha$  such that  $\alpha y_{p(k)} - u_{p(k)} \in V_{p(k)-1}$ . Thus,  $\alpha \tilde{f}_k y_{p(k)} - \tilde{f}_k u_{p(k)} = \alpha \tilde{f}_k y_{p(k)} - u_{k+1}$  is linear combination of terms of the form  $\tilde{f}_k u_j$ ,  $1 \leq j \leq p(k)-1$ . Since  $p(k)$  is the minimal index for which there exists a monomial  $\tilde{f}_k$  with  $\tilde{f}_k u_{p(k)} \in V_{k+1}$ , each of the terms  $\tilde{f}_k u_j$ ,  $1 \leq j \leq p(k)-1$  is in  $V_k$ . It follows that  $\alpha \tilde{f}_k y_{p(k)} - u_{k+1} \in V_k$ . Again, there exists a constant  $\alpha'$  such that  $\alpha' u_{k+1} - y_{k+1} \in V_k$ . Therefore, writing  $f_k = \alpha \alpha' \tilde{f}_k$ , we conclude that  $f_k y_{p(k)} - y_{k+1} = \alpha' (\alpha \tilde{f}_k y_{p(k)} - u_{k+1}) + \alpha' u_{k+1} - y_{k+1} \in V_k$ ; i.e.,  $f_k y_{p(k)} = y_{k+1} - \sum_{D_j \leq D_k} \tilde{r}_{k,j} y_j$ . The

first equation in (4.8) is now proved in view of (4.9), applied with  $p(k)$  in place of  $k$ , and the fact that  $D_{p(k)} = D_{k+1} - 1$ . We note that  $f_k$  is a constant multiple of the monomial  $\tilde{f}_k$ . The second equation in (4.8) is proved in the same way.

Using the second equation in (4.8) and (4.9), we obtain from the definition of  $c_{k,j}$ 's that

$$\begin{aligned} A_k c_{k+1,\ell} &= A_k \int_{\Omega} t_{k+1} y_{\ell} d\mu \\ &= \int_{\Omega} f_k t_{p(k)} y_{\ell} d\mu + \sum_{D_{k+1}-2 \leq D_j \leq D_k} s_{k,j} \int_{\Omega} t_j y_{\ell} d\mu \\ &= \sum_{D_{\ell}-1 \leq D_m \leq D_{\ell}+1} r_{\ell,m}(f_k) \int_{\Omega} t_{p(k)} y_m d\mu + \sum_{D_{k+1}-2 \leq D_j \leq D_k} s_{k,j} c_{j,\ell} \\ &= \sum_{D_{\ell}-1 \leq D_m \leq D_{\ell}+1} r_{\ell,m}(f_k) c_{p(k),m} + \sum_{D_{k+1}-2 \leq D_j \leq D_k} s_{k,j} c_{j,\ell}. \end{aligned}$$

This proves (4.10).  $\square$

Next, we use Proposition 2.1 to prove Theorem 3.1(a) and Theorem 3.2(a).

PROOF OF THEOREM 3.1(a). To prove part (a), we assume without loss of generality that  $n \geq 8$ , and let  $\ell \geq 1$  be the largest integer with  $2^{\ell+2} \leq n$ . In view of (3.4),

$$\begin{aligned} \|f - \sigma_n^*(\mathcal{C}, \mathbf{W}; h; f)\|_{\infty} &\leq c E_{n/2}(f) \leq c E_{2^{\ell+1}}(f) \leq c \|f - \sigma_{2^{\ell+1}}^*(h; f)\|_{\infty} \\ &\leq c \sum_{k=\ell+1}^{\infty} \|\sigma_{2^{k+1}}^*(h; f) - \sigma_{2^k}^*(h; f)\|_{\infty} \leq c 2^{-r\ell} \leq cn^{-r}. \end{aligned}$$

This proves part (a).  $\square$

PROOF OF THEOREM 3.2(a). Let  $K''$  be a spherical cap, concentric with  $K$ ,  $K'$ , and having radius equal to the average of the radii of  $K$ ,  $K'$ . Let  $\psi$  be fixed,  $C^{\infty}$  function that is equal to 1 on  $K''$  and equal to 0 outside of  $K$ . Without loss of generality, we may assume that  $n \geq 8$ , and let  $\ell \geq 1$  be the largest integer such that  $2^{\ell+2} \leq n$ . The direct theorem of approximation theory (cf. [26]) implies that there exists  $P \in \Pi_{2^{\ell}}^q$  such that

$$\|\psi - P\|_{\infty} \leq c 2^{-\ell S}.$$

Therefore, using the definition of  $\|f\|_{\mathbb{W}_r(K)}$ , we conclude that

$$\begin{aligned} E_{2^{\ell+1}}(f\psi) &\leq \|f\psi - P\sigma_{2^{\ell}}^*(h; f)\|_{\infty} \leq \|(f - \sigma_{2^{\ell}}^*(h; f))\psi\|_{\infty} + \|(\psi - P)\sigma_{2^{\ell}}^*(h; f)\|_{\infty} \\ &\leq c \{ \|f - \sigma_{2^{\ell}}^*(h; f)\|_{\infty, K} + 2^{-nS} \|f\|_{\infty} \} \\ &\leq c \left\{ \sum_{k=\ell+1}^{\infty} \|\sigma_{2^{k+1}}^*(h; f) - \sigma_{2^k}^*(h; f)\|_{\infty, K} + 2^{-nS} \|f\|_{\infty} \right\} \leq c 2^{-r\ell}. \end{aligned}$$

In view of (3.4),

$$\begin{aligned} \|f - \sigma_n(\mathcal{C}, \mathbf{W}; h; f\psi)\|_{\infty, K'} &= \|f\psi - \sigma_n(\mathcal{C}, \mathbf{W}; h; f\psi)\|_{\infty, K'} \leq \|f\psi - \sigma_n(\mathcal{C}, \mathbf{W}; h; f\psi)\|_{\infty} \\ &\leq c E_{n/2}(f\psi) \leq E_{2^{\ell+1}}(f\psi) \leq c 2^{-r\ell} \leq cn^{-r}. \end{aligned} \quad (6.1)$$

Since  $1 - \psi(\zeta) = 0$  for  $\zeta \in K''$ , we may use (2.8) to deduce that for  $\mathbf{x} \in K'$ ,

$$\begin{aligned} |\sigma_n(\mathcal{C}, \mathbf{W}; h; (1 - \psi)f, \mathbf{x})| &= \left| \sum_{\xi \in \mathcal{C} \setminus K''} w_\xi f(\xi) (1 - \psi(\xi)) \Phi_n(h; \mathbf{x} \cdot \xi) \right| \\ &\leq \frac{c(K, K', K'')}{n^{S-q}} \|(1 - \psi)f\|_\infty \sum_{\xi \in \mathcal{C}} |w_\xi| \leq \frac{c(K, K', K'')}{n^{S-q}}. \end{aligned}$$

Together with (6.1) and the fact that  $r \leq S - q$ , this implies (3.9).  $\square$

Next, we prove Lemma 6.1, describing certain extremal properties for the weights  $w_\xi^{LSQ}$ . These will be used in the proof of parts (a) and (b) of Theorem 4.1.

**Lemma 6.1** *Let  $n \geq 1$  be an integer,  $N = d_n^{q+1}$ ,  $\mathcal{C}$  be a finite set of points on  $\mathbb{S}^q$  and  $\nu$  be a measure supported on  $\mathcal{C}$ . Let  $v_\xi := \nu(\{\xi\})$ ,  $\xi \in \mathcal{C}$ .*

- (a) *If the Gram matrix is positive definite, then the weights  $w_\xi^{LSQ}$  are solutions of the extremal problem to minimize  $\sum_{\xi \in \mathcal{C}} w_\xi^2 / v_\xi$  subject to the conditions that  $\sum_{\xi \in \mathcal{C}} w_\xi y_\ell(\xi) = \delta_{1,\ell}$ .*  
(b) *If (4.3) holds, there exist real numbers  $W_\xi$ ,  $\xi \in \mathcal{C}$ , such that  $|W_\xi| \leq v_\xi$  for  $\xi \in \mathcal{C}$  and  $\sum_{\xi \in \mathcal{C}} W_\xi P(\xi) = \int P d\mu_q$  for all  $P \in \Pi_n^q$ .*

PROOF. In this proof, we will write  $G$  in place of  $G_N$ . The Lagrange multiplier method to solve the minimization problem sets up parameters  $\lambda_\ell$  and minimizes

$$\sum_{\xi \in \mathcal{C}} w_\xi^2 / v_\xi - 2 \sum_{\ell} \lambda_\ell \left( \sum_{\xi} w_\xi y_\ell(\xi) - \delta_{1,\ell} \right).$$

Setting the gradient (with respect to  $w_\xi$ ) equal to 0, we get  $w_\xi = v_\xi \sum_{\ell} \lambda_\ell y_\ell(\xi)$ . Writing, in this proof only,  $Q = \sum_{\ell} \lambda_\ell y_\ell$ , we see that  $w_\xi = v_\xi Q(\xi)$ . Substituting back in the linear constraints, this reduces to  $\sum_{\xi \in \mathcal{C}} v_\xi Q(\xi) y_\ell(\xi) = \delta_{1,\ell}$ . These conditions determine  $Q$  uniquely; indeed,  $Q = \sum_j G_{1,j}^{-1} y_j$ . This proves part (a).

The part (b) is proved essentially in [21], but since it is not stated in this manner, we sketch a proof again. During this proof, different constants will retain their values. Let  $M = |\mathcal{C}|$ ,  $\mathbb{R}^M$  be equipped with the norm  $\|\mathbf{r}\| = \sum_{\xi \in \mathcal{C}} v_\xi |r_\xi|$ . In this proof only, let  $\mathcal{S}$  be the operator defined on  $\Pi_n^q$  by  $\mathcal{S}(P) = (P(\xi))_{\xi \in \mathcal{C}} \in \mathbb{R}^M$ , and  $\mathbb{V}$  be the range of  $\mathcal{S}$ . The estimate

$$\int |P| d\mu_q \leq c_1 \sum_{\xi \in \mathcal{C}} v_\xi |P(\xi)| \tag{6.2}$$

implies that the operator  $\mathcal{S} : \Pi_n^q \rightarrow \mathbb{V}$  is invertible. We may now define a linear functional on  $\mathbb{V}$  by

$$x^*(\mathbf{r}) = \int \mathcal{S}^{-1}(\mathbf{r}) d\mu_q, \quad \mathbf{r} \in \mathbb{V}.$$

It is clear from (6.2) that the norm of  $x^*$  is bounded above by  $c_1$ . The Hahn–Banach theorem yields a norm preserving extension of this functional to the whole space  $\mathbb{R}^M$ . Identifying this functional with the vector  $(W_\xi)_{\xi \in \mathcal{C}}$ , the extension property implies that  $\sum_{\xi \in \mathcal{C}} W_\xi P(\xi) = \int P d\mu_q$  for all  $P \in \Pi_n^q$ , while the norm preservation property implies that  $|W_\xi| \leq c_1 v_\xi$  for  $\xi \in \mathcal{C}$ .  $\square$

PROOF OF THEOREM 4.1 (a), (b). Let  $N = d_n^{q+1}$ ,  $\mathbf{r} \in \mathbb{R}^N$ , and  $P = \sum_{\ell} r_\ell y_\ell$ . In this proof only, we write  $G$  in place of  $G_N$ . Then

$$\mathbf{r}^T G \mathbf{r} = \sum_{\ell, m} r_\ell \left\{ \sum_{\xi \in \mathcal{C}} v_\xi y_\ell(\xi) y_m(\xi) \right\} r_m = \sum_{\xi \in \mathcal{C}} v_\xi P(\xi)^2,$$

and  $\mathbf{r}^T \mathbf{r} = \|P\|_2^2$ . Therefore, (4.3) with  $p = 2$  implies that  $c_1^2 \mathbf{r}^T \mathbf{r} \leq \mathbf{r}^T G \mathbf{r} \leq c_2^2 \mathbf{r}^T \mathbf{r}$  for all  $\mathbf{r} \in \mathbb{R}^N$ . The statements about the eigenvalues of  $G$  are an immediate consequence of the Raleigh–Ritz theorem [13,

Theorem 4.2.2]. Using Lemma 6.1(b), we obtain weights  $W_\xi$  such that  $\sum_{\xi \in \mathcal{C}} W_\xi y_\ell(\xi) = \delta_{1,\ell}$ ,  $\ell = 1, \dots, N$ , and  $|W_\xi| \leq cv_\xi$ ,  $\xi \in \mathcal{C}$ . During the remainder of this proof, we write  $w_\xi = w_\xi^{LSQ}$ . In view of Lemma 6.1(a), we have

$$\sum_{\xi \in \mathcal{C}} |w_\xi| \leq \left\{ \sum_{\xi} v_\xi \right\}^{1/2} \left\{ \sum_{\xi \in \mathcal{C}} \frac{w_\xi^2}{v_\xi} \right\}^{1/2} \leq \left\{ \sum_{\xi} v_\xi \right\}^{1/2} \left\{ \sum_{\xi \in \mathcal{C}} \frac{W_\xi^2}{v_\xi} \right\}^{1/2} \leq c \sum_{\xi \in \mathcal{C}} v_\xi \leq c_1.$$

This completes the proof of part (a).

In order to prove part (b), we adopt the following notation during this proof only. Let  $I$  denote the  $N \times N$  identity matrix. For any  $N \times N$  matrix  $H$ , let  $\|H\|$  denote  $\sup \|H\mathbf{r}\|$ ,  $\|\mathbf{r}\| = 1$ ,  $\mathbf{r} \in \mathbb{R}^N$ . We note that  $\|H\|$  is the largest singular value of  $H$ . If  $H$  is a symmetric, positive definite matrix, then it is also the largest eigenvalue of  $H$ , and moreover,  $|\mathbf{r}_1^T H \mathbf{r}_2| \leq \|H\| \|\mathbf{r}_1\| \|\mathbf{r}_2\|$ ,  $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^N$ . Using (4.4), it is easy to conclude using the Raleigh–Ritz theorem that  $\|G - I\| \leq cn^{-q}$ ,  $\|G^{-1}\| \leq c$ , and hence,

$$\|G^{-1} - I\| = \|G^{-1}(I - G)\| \leq c\|G^{-1}\| \|G - I\| \leq cn^{-q}.$$

Let  $\mathbf{y}(\mathbf{x})$  denote the vector  $(y_1(\mathbf{x}), \dots, y_N(\mathbf{x}))^T$  for  $\mathbf{x} \in \mathbb{S}^q$ . In view of the addition formula,  $\|\mathbf{y}(\mathbf{x})\|^2$  is independent of  $\mathbf{x}$ , and hence,

$$\|\mathbf{y}(\mathbf{x})\|^2 = \int_{\mathbb{S}^q} \sum_{j=1}^N y_j(\mathbf{x})^2 d\mu_q(\mathbf{x}) = d_n^{q+1} \leq cn^q, \quad \mathbf{x} \in \mathbb{S}^q.$$

Consequently, we have

$$\begin{aligned} \frac{|w_\xi^{LSQ}|}{v_\xi} &= \left| \int \mathbf{y}(\mathbf{x})^T G^{-1} \mathbf{y}(\xi) d\mu_q(\mathbf{x}) \right| \\ &\leq \left| \int \mathbf{y}(\mathbf{x})^T (G^{-1} - I) \mathbf{y}(\xi) d\mu_q(\mathbf{x}) \right| + \left| \int \mathbf{y}(\mathbf{x})^T \mathbf{y}(\xi) d\mu_q(\mathbf{x}) \right| \\ &\leq \|G^{-1} - I\| \int \|\mathbf{y}(\mathbf{x})\| \|\mathbf{y}(\xi)\| d\mu_q + |(1, 0, \dots, 0)^T \mathbf{y}(\xi)| \leq cn^{-q} n^q + c \leq c. \end{aligned}$$

This completes the proof of part (b).  $\square$

The proof of the remaining new results in the paper are based on the following lemma, that gives a recipe for estimating the probabilities involving polynomial valued random variables.

**Lemma 6.2** *Let  $n, M \geq 1$  be integers,  $\{\omega_j\}_{j=1}^M$  be independent random variables, and for  $j = 1, \dots, M$ ,  $Z_j = Z(\omega_j, \circ) \in \Pi_n^q$  have mean equal to 0 according to  $\omega_j$ . Let  $B, R > 0$ ,  $\max_{1 \leq j \leq M, \mathbf{x} \in \mathbb{S}^q} |Z_j(\mathbf{x})| \leq Rn^q$ , and the sum of the variances of  $Z_j$  be bounded by  $Bn^q$  uniformly on  $\mathbb{S}^q$ . If  $A > 0$  and  $12R^2(A+q)n^q \log n \leq B$  then*

$$\text{Prob} \left( \left\| \sum_{j=1}^M Z_j \right\|_\infty \geq \sqrt{12B(A+q)n^q \log n} \right) \leq c_1 n^{-A}. \quad (6.3)$$

Here, the positive constant  $c_1$  is independent of  $M$  and the distributions of  $\omega_j$ .

**PROOF.** The proof depends upon Bennett's inequality [27, p. 192]. In this proof only, we adopt a slightly different meaning for the symbols  $L, V, \eta$ . Let  $L, V, \eta$  be positive numbers, and  $X_j$ ,  $j = 1, \dots, M$ , be independent random variables. According to Bennett's inequality, if the mean of each  $X_j$  is 0, the range of each  $X_j$  is a subset of  $[-L, L]$ , and  $V$  exceeds the sum of the variances of  $X_j$ , then for  $\eta > 0$ ,

$$\text{Prob} \left( \left| \sum_{j=1}^M X_j \right| \geq \eta \right) \leq 2 \exp \left( -\frac{V}{L^2} g(L\eta/V) \right), \quad (6.4)$$

where, in this proof only,  $g(t) := (1+t)\log(1+t) - t$ . We observe that  $g(t) = \int_0^t \int_0^u (1+w)^{-1} dw du$ . Therefore, if  $0 \leq t \leq 1/2$ , then for  $0 \leq w \leq u \leq t$ ,  $(1+w)^{-1} \geq 2/3$ , and hence,  $g(t) \geq t^2/3$ . Consequently, if  $L\eta \leq V/2$ , then

$$\text{Prob} \left( \left| \sum_{j=1}^M X_j \right| \geq \eta \right) \leq 2 \exp(-\eta^2/(3V)). \quad (6.5)$$

Now, let  $\mathbf{x} \in \mathbb{S}^q$ . We apply (6.5) with  $Z_j(\mathbf{x})$  in place of  $X_j$ ,  $Rn^q$  in place of  $L$ ,  $Bn^q$  in place of  $V$ ,  $\eta = \sqrt{3B(A+q)n^q \log n}$ . Our condition on  $n$  ensures that  $L\eta/V \leq 1/2$  with these choices. Therefore,

$$\text{Prob} \left( \left| \sum_{j=1}^M Z_j(\mathbf{x}) \right| \geq \sqrt{3B(A+q)n^q \log n} \right) \leq 2n^{-A-q}. \quad (6.6)$$

Next, in the proof only, let  $P^* = \sum_{j=1}^M Z_j$ ,  $\mathbf{x}^* \in \mathbb{S}^q$  be chosen so that  $|P^*(\mathbf{x}^*)| = \|P^*\|_\infty$ ,  $\mathcal{C} \subset \mathbb{S}^q$  be chosen so that  $|\mathcal{C}| \sim cn^q$  and  $\delta_{\mathcal{C}}(\mathbb{S}^q) \leq 1/(2n)$ . Then we may find  $\xi^* \in \mathcal{C}$  such that  $\text{dist}(\mathbf{x}^*, \xi^*) \leq 1/(2n)$ . Since  $P^* \in \Pi_n^q$ , its restriction to the great circle through  $\mathbf{x}^*$  and  $\xi^*$  is a trigonometric polynomial of order at most  $n$ . In view of the Bernstein inequality for these polynomials [2, Chapter 4, (1.1)],

$$|P^*(\xi^*) - P^*(\mathbf{x}^*)| \leq n \|P^*\|_\infty \text{dist}(\xi^*, \mathbf{x}^*) \leq (1/2) |P^*(\mathbf{x}^*)|.$$

We deduce that

$$\left\| \sum_{j=1}^M Z_j \right\|_\infty \leq 2 \max_{\mathbf{x} \in \mathcal{C}} \left| \sum_{j=1}^M Z_j(\mathbf{x}) \right|.$$

Therefore, the event  $\left\| \sum_{j=1}^M Z_j \right\|_\infty \geq 2\sqrt{3B(A+q)n^q \log n}$  is a subset of the union of the  $|\mathcal{C}|$  events  $|\sum_{j=1}^M Z_j(\mathbf{x})| \geq \sqrt{3B(A+q)n^q \log n}$ ,  $\mathbf{x} \in \mathcal{C}$ . Hence, the estimate (6.6) implies (6.3) with  $c_1 = 2c$ .  $\square$

We are now in a position to prove Theorem 4.1(c).

PROOF OF THEOREM 4.1(c). Let  $\mathbf{x} \in \mathbb{S}^q$ . In this proof only, let  $Z_\xi = \Phi_{4n}(h; \mathbf{x} \cdot \xi) - \int \Phi_{4n}(h; \mathbf{x} \cdot \zeta) d\mu_q(\zeta)$ . Then the mean of each  $Z_\xi$  is 0, and its variance can be estimated by

$$\int Z_\xi^2 d\mu_q(\xi) \leq \int (\Phi_{4n}(h; \mathbf{x} \cdot \xi))^2 d\mu_q(\xi) \leq cn^q.$$

Finally,  $|Z_\xi| \leq cn^q$  for each  $\xi$ . Hence, we may use Lemma 6.2 with  $cM$  in place of  $B$ , and  $c$  in place of  $R$ , to conclude that

$$\text{Prob} \left( \sup_{\mathbf{x} \in \mathbb{S}^q} \left| \frac{1}{M} \sum_{\xi \in \mathcal{C}} \Phi_{4n}(h; \mathbf{x} \cdot \xi) - \int \Phi_{4n}(h; \mathbf{x} \cdot \zeta) d\mu_q(\zeta) \right| \geq c_2 \sqrt{\frac{n^q \log n}{M}} \right) \leq cn^{-A},$$

and with  $M \geq cn^q \log n / \eta^2$ ,

$$\text{Prob} \left( \sup_{\mathbf{x} \in \mathbb{S}^q} \left| \frac{1}{M} \sum_{\xi \in \mathcal{C}} \Phi_{4n}(h; \mathbf{x} \cdot \xi) - \int \Phi_{4n}(h; \mathbf{x} \cdot \zeta) d\mu_q(\zeta) \right| \geq \eta \right) \leq cn^{-A}.$$

Since any  $P \in \Pi_{2n}^q$  can be written in the form

$$P(\zeta) = \int P(\mathbf{x}) \Phi_{4n}(h; \mathbf{x} \cdot \zeta) d\mu_q(\mathbf{x}),$$

we see that with probability exceeding  $1 - cn^{-A}$ ,

$$\left| \frac{1}{M} \sum_{\xi \in \mathcal{C}} P(\xi) - \int P(\zeta) d\mu_q(\zeta) \right|$$

$$\begin{aligned}
&= \left| \frac{1}{M} \sum_{\xi \in \mathcal{C}} \int P(\mathbf{x}) \Phi_{4n}(h; \xi \cdot \mathbf{x}) d\mu_q(\mathbf{x}) \right. \\
&\quad \left. - \int \int P(\mathbf{x}) \Phi_{4n}(h; \mathbf{x} \cdot \zeta) d\mu_q(\mathbf{x}) d\mu_q(\zeta) \right| \\
&\leq \int |P(\mathbf{x})| \left| \frac{1}{M} \sum_{\xi \in \mathcal{C}} \Phi_{4n}(h; \xi \cdot \mathbf{x}) - \int \Phi_{4n}(h; \mathbf{x} \cdot \zeta) d\mu_q(\zeta) \right| d\mu_q(\mathbf{x}) \\
&\leq \eta \int |P(\mathbf{x})| d\mu_q(\mathbf{x}).
\end{aligned}$$

For  $P \in \Pi_n^q$ , we may now apply this estimate with  $P^2 \in \Pi_{2n}^q$ .  $\square$

Another immediate consequence of Lemma 6.2 is the following lemma, describing the probabilistic behavior of the operator  $\sigma_n(\mathcal{C}, \mathbf{W}; h)$ .

**Lemma 6.3** *Suppose that  $m \geq 1$  is an integer,  $\mathcal{C} = \{\xi_j\}_{j=1}^M$  admits an  $M$ -Z quadrature of order  $m$ , and let  $\mathbf{W}$  be the corresponding quadrature weights. Let  $R, V > 0$ , and for  $j = 1, \dots, M$ ,  $\epsilon_j$  be independent random variables with mean 0, variance not exceeding  $V$ , and range  $[-R, R]$ . Let  $g \in C(\mathbb{S}^q)$ ,  $\|g\|_\infty \leq 1$ , and  $\mathbf{E} = \{\epsilon_j g(\xi_j)\}_{\xi_j \in \mathcal{C}}$ . Then for integer  $n \geq 1$  with  $(R^2/V)(A+q)n^q \log n \leq c_3 m^q$ ,*

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{E})\|_\infty \geq c_1 \sqrt{\frac{V(A+q)n^q \log n}{m^q}} \right) \leq c_2 n^{-A}. \quad (6.7)$$

Here, the positive constants  $c_1, c_2, c_3$  depend only on  $q$  but not on  $M$  and the distributions of  $\epsilon_j$ .

PROOF. In this proof only, if  $\xi = \xi_j \in \mathcal{C}$ , we will write  $\epsilon_\xi$  for  $\epsilon_j$  and  $w_\xi$  for the weight in  $\mathbf{W}$  corresponding to  $\xi$ . We use Lemma 6.2 with  $E_\xi = m^q w_\xi \epsilon_\xi g(\xi) \Phi_n(h; \xi \cdot \circ)$ ,  $\xi \in \mathcal{C}$ . We note that the random variable  $\omega_j$  in Lemma 6.2 is  $\epsilon_\xi$  in this case. It is clear that the mean of each  $E_\xi$  is 0. Since (2.12) implies that  $|w_\xi| \leq cm^{-q}$ , (2.7) shows that  $\|E_\xi\|_\infty \leq cRn^q$ . Moreover, for any  $\mathbf{x} \in \mathbb{S}^q$ , the variance of  $E_\xi(\mathbf{x})$  does not exceed  $Vm^{2q} w_\xi^2 \Phi_n(h; \xi \cdot \mathbf{x})^2$ . In view of the fact that  $w_\xi$  are  $M$ -Z quadrature weights, (2.12) and (2.7) imply that

$$\begin{aligned}
\sum_{\xi \in \mathcal{C}} m^{2q} w_\xi^2 \Phi_n(h; \xi \cdot \mathbf{x})^2 &\leq cm^q \sum_{\xi \in \mathcal{C}} |w_\xi| \Phi_n^2(h; \xi \cdot \mathbf{x}) \\
&\leq cm^q \int_{\mathbb{S}^q} \Phi_n^2(h; \zeta \cdot \mathbf{x}) d\mu_q(\zeta) \leq cm^q n^q.
\end{aligned}$$

Thus, we may choose  $B$  in Lemma 6.2 to be  $cVm^q$ . The estimate (6.7) now follows as a simple consequence of (6.3).  $\square$

We are now in a position to prove the probabilistic assertions of Theorems 3.1 and 3.2.

PROOF OF THEOREM 3.1(b). To prove part (b), we use Lemma 6.3 with  $g \equiv 1$ . Since the range of  $\epsilon_j$ 's is contained in  $[-1, 1]$ , we may take  $R = V = 1$ , and obtain from (6.7) that

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{E})\|_\infty \geq c_4 \sqrt{\frac{(A+q)n^q \log n}{m^q}} \right) \leq c_2 n^{-A}.$$

The choice of  $n$  with an appropriate  $c_3$ , ensures that  $(A+q)n^q \log n / m^q \leq n^{-2r}$ . Therefore,

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{E})\|_\infty \geq c_4 n^{-r} \right) \leq c_2 n^{-A}.$$

The estimate (3.8) is now clear in view of (3.4) and the linearity of the operators  $\sigma_n(\mathcal{C}, \mathbf{W}; h)$ .  $\square$

PROOF OF THEOREM 3.2(b). We apply Lemma 6.3 again with  $g \equiv 1$ . As before, we may choose  $R = V = 1$ . The choice of  $n$  with an appropriate  $c_3$ , ensures that  $(A+q)n^q \log n / m^q \leq n^{-2r}$ . Therefore, (6.7) with these choices implies that

$$\text{Prob} \left( \|\sigma_n(\mathcal{C}, \mathbf{W}; h; \mathbf{E})\|_\infty \geq c_4 n^{-r} \right) \leq c_2 n^{-A}.$$

Together with (3.9) and the linearity of the operators  $\sigma_n(\mathcal{C}, \mathbf{W}; h)$ , this leads to (3.10). □

## 7 Conclusion

We have described a construction of linear operators yielding spherical polynomial approximations based on scattered data on a Euclidean sphere. While the operators can be defined for arbitrary continuous functions on the sphere, without any a priori knowledge about the location and nature of its singularities, they are auto-adaptive in the sense that the approximation properties of these globally defined polynomials adapt themselves on the different parts of the sphere according to the smoothness of the target function on these parts. While the theoretical properties of these operators and their localization were studied in [18], a bottleneck in their numerical construction was the construction of quadrature formulas based on scattered data, exact for integrating moderately high degree spherical polynomials. So far, it was possible only to compute quadrature formulas exact at most for degree 18 polynomials. We show that a simple-minded construction involving a Gram matrix is surprisingly well conditioned, and yields the necessary quadrature rules, up to degree 178. Using these newly constructed quadrature formulas, we are able to demonstrate that our constructions yield superior approximation properties to those of more traditional techniques of least squares and Fourier projection, in the sense that the presence of singularities in some parts of the sphere affects the degree of approximation by our operators on other parts far less than in the case of these other traditional techniques. We give probabilistic estimates on the local and global degrees of approximation by our operators in the presence of noise, and demonstrate its use in the modeling of a “real life” data set. We also describe a theoretical algorithm for construction of data dependent multivariate orthogonal polynomials and their use in the construction of quadrature formulas, analogous to the univariate algorithms in the book [11] of Gautschi.

## References

- [1] C. DE BOOR, “A practical guide to splines”, Springer Verlag, New York, 1978.
- [2] R. A. DEVORE AND G. G. LORENTZ, “Constructive approximation”, Springer Verlag, Berlin, 1993.
- [3] J. R. DRISCOLL AND D. M. HEALY, *Computing Fourier transforms and convolutions on the 2-sphere*, Adv. in Applied Math. **15**, 202-250 (1994).
- [4] C. F. DUNKL AND Y. XU, “Orthogonal polynomials of several variables”, Cambridge University Press, Cambridge, 2001.
- [5] A. ERDÉLYI (ED.), W. MAGNUS, F. OBERHETTINGER AND F. G. TRICOMI (RESEARCH ASSOCIATES), “Higher Transcendental Functions, Volume II”, California Institute of Technology, Bateman Manuscript Project, McGraw-Hill Book Company, Inc., New York, Toronto, London, 1953.
- [6] B. FISCHER, *From orthogonal polynomials to iteration schemes for linear systems: CG and CR revisited*, in “Wavelet Analysis and Applications, Proceedings of the international workshop in Delhi, 1999” (P. K. Jain, M. Krishnan, H. N. Mhaskar J. Prestin, and D. Singh Eds.), Narosa Publishing, New Delhi, India, 2001, 225-247.
- [7] W. FREEDEN, M. SCHREINER, AND R. FRANKE, *A Survey on Spherical Spline Approximation*, Surveys Math. Indust. **7** (1997), 29-85.
- [8] W. FREEDEN, T. GERVENS, AND M. SCHREINER, “Constructive approximation on the sphere, with applications to geomathematics”, Clarendon Press, Oxford, 1998.
- [9] W. FREEDEN AND V. MICHEL, “Multiscale potential theory, with applications to geoscience”, Birkhäuser, Boston, 2004.

- [10] M. GANESH AND H. N. MHASKAR, *Matrix-free interpolation on the sphere*, SIAM J. Numer. Analysis **44** (3) (2006), pp. 1314–1331.
- [11] W. GAUTSCHI, “Orthogonal Polynomials: Computation and Approximation”, Oxford University Press, 2004.
- [12] K. HESSE, H. N. MHASKAR, AND I. H. SLOAN, *Quadrature in Besov spaces on the Euclidean sphere*, Accepted for publication in Journal of Complexity.
- [13] R. A. HORN AND C. R. JOHNSON, “Matrix analysis”, Cambridge University Press, 1985.
- [14] J. KEINER, *Fast Spherical Fourier Transforms and Applications*, Diplomarbeit, Universität zu Lübeck, 2005.
- [15] Q. T. LE GIA AND H. N. MHASKAR, *Polynomial operators and local approximation of solutions of pseudo-differential equations on the sphere*, Numer. Math., **103** (2006), 299–322.
- [16] P. I. LIZORKIN AND KH. P. RUSTAMOV, *Nikolskii–Besov spaces on the sphere in connection with approximation theory*, Tr. Mat. Inst. Steklova, **204** (1993), 172–201 (Proc. Steklov Inst. Math., **3** (1994), 149–172).
- [17] H. N. MHASKAR, *Polynomial operators and local smoothness classes on the unit interval*, J. Approx. Theory, **131** (2004), 243–267.
- [18] H. N. MHASKAR, *On the representation of smooth functions on the sphere using finitely many bits*, Appl. Comput. Harmon. Anal., **18** (3) (2005), 215–233.
- [19] H. N. MHASKAR, *Weighted quadrature formulas and approximation by zonal function networks on the sphere*, J. Complexity, **22** (2006), 348–370.
- [20] H. N. MHASKAR, F. J. NARCOWICH, AND J. D. WARD, *Approximation Properties of Zonal Function Networks Using Scattered Data on the Sphere*, Adv. Comput. Math., **11** (1999), 121–137.
- [21] H. N. MHASKAR, F. J. NARCOWICH AND J. D. WARD, *Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature*, Math. Comp. **70** (2001), no. 235, 1113–1130. (Corrigendum: Math. Comp. **71** (2001), 453–454.)
- [22] H. N. MHASKAR, F. NARCOWICH, J. PRESTIN, AND J. D. WARD, *Converse theorems for approximation by SBF networks*, In preparation.
- [23] H. N. MHASKAR AND J. PRESTIN, *On the detection of singularities of a periodic function*, Advances in Computational Mathematics, **12** (2000), 95–131.
- [24] H. N. MHASKAR AND J. PRESTIN, *On local smoothness classes of periodic functions*, Journal of Fourier Analysis and Applications, **11** (3) (2005), 353 - 373.
- [25] C. MÜLLER, “Spherical Harmonics”, Lecture Notes in Mathematics, Vol. 17, Springer Verlag, Berlin, 1966.
- [26] S. PAWELKE, *Über die Approximationsordnung bei Kugelfunktionen und algebraischen Polynomen*, Tôhoku Math. Journ., **24** (1972), 473–486.
- [27] D. POLLARD, *Convergence of stochastic processes*, Springer Verlag, New York, 1984.
- [28] D. POTTS, G. STEIDL, AND M. TASCHE, *Fast algorithms for discrete polynomial transforms*, Math. Comp., **67** (1998), 1577–1590.
- [29] I. H. SLOAN, *Polynomial interpolation and hyperinterpolation over general regions*, J. Approx. Theory **83** (1995), no. 2, 238–254.
- [30] I. H. SLOAN AND A. SOMMARIVA, *Approximation on the sphere using radial basis functions plus polynomials*, Manuscript.

- [31] I. H. SLOAN AND R.S. WOMERSLEY, *Extremal systems of points and numerical integration on the sphere*, Adv. Comput. Math., **21** (2004), 107–125.
- [32] E. M. STEIN AND G. WEISS, “Fourier Analysis on Euclidean Spaces”, Princeton University Press, Princeton, New Jersey, 1971.