Zonal function network frames on the sphere

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

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

Abstract

We introduce a class of zonal function network frames suitable for analyzing data collected at scattered sites on the surface of the unit sphere of a Euclidean space. Our frames consist of zonal function networks and are well localized. The frames belonging to higher and higher scale wavelet spaces have more and more vanishing polynomial moments. The main technique is applicable in the general setting of separable Hilbert spaces, in which context, we study the construction of new frames by perturbing an orthonormal basis. 1

1 Introduction

An important reason for the popularity of neural networks in many applications is their universal approximation property [4, 11, 12]. A neural network with a single hidden layer and a univariate activation function φ evaluates a function of the form \( x \mapsto \sum_{k=1}^{N} a_k \phi(w_k \cdot x + b_k) \), where the coefficients \( a_k \in \mathbb{R} \), the thresholds \( b_k \in \mathbb{R} \), \( x \in \mathbb{R}^s \) denotes, in vector notation, all the inputs to the network, and \( w_k \)'s are the synaptic weights. By a proper choice of these coefficients, thresholds, and weights, one can approximate any continuous target function of \( s \) real variables on any compact subset of \( \mathbb{R}^s \) arbitrarily well. Typically, the target function is unknown; the parameters of the network have to be computed using some training data about the function.

In many applications in geophysics and meteorology, the data is collected over the surface of the earth by satellites or ground stations. One then seeks to find a functional model for the mechanism that generates this data. To make an optimal use of this data, it seems natural to consider networks that evaluate a function of the form \( x \mapsto \sum_{k=1}^{N} a_k \phi(\xi_k \cdot x) \), where the weights \( \xi_k \) are the sites at which the data is collected. In [14], we have called such networks zonal function (ZF) networks. We have studied the approximation properties of such networks, and developed algorithms to construct the coefficients \( a_k \) so that the resulting network provides an optimal approximation to a large class of target functions. A novelty of our approach is that we give explicit formulas for the coefficients. The training of the network thus reduces to simple matrix multiplications; no optimization techniques are required, thereby avoiding all pitfalls such as local minima.

*The research of this author was supported, in part, by grant DMS-9971846 from the National Science Foundation and grant DAAD19-01-1-0001 from the U.S. Army Research Office.
†Research of the authors was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-98-1-0204. The research of JDW was also supported by grant DMS-9971276 from the National Science Foundation.

1991 Mathematics Subject Classification: 65D32, 41A17, 42C10, Accepted for publication in Neural Networks.
One next wishes to decompose the underlying function into “high frequency” and “low frequency” components with the hope of further understanding various properties of the function, for example, the location of singularities. The standard wavelet paradigm of constructing spaces based on an increasing chain of lattices does not apply to this situation. Nevertheless, many authors [5, 6, 8, 21, 26] have derived various nonstationary constructions of wavelets or frames on the sphere. Such constructions fall roughly into three categories: continuous wavelet transforms (CWT), discrete wavelets or frames and biorthogonal wavelets using lifting techniques. Each approach seems to have its advantages and disadvantages.

Since the partial sum of a wavelet expansion may be thought of as a neural network with a wavelet activation function, it is natural to seek to exploit the advantages of both neural network paradigms and wavelet theory paradigms to have an effective analysis of the data. A recent search with WebScience revealed 203 articles under the keywords wavelets and neural networks. Various neural networks are constructed using certain wavelet like activation functions, by Zhang and Benveniste [29], Zhang, Walter, Miao, and Lee [30], Candès [1], Fang and Chow [7] among others. Pati and Krishnaprasad [22] have constructed wavelets using the standard sigmoidal function \( x \mapsto (1 + e^{-x})^{-1} \) in the univariate case. Their wavelets have vanishing integrals, and in general, are meant to mimick the Haar wavelets. In [2], neural networks with multiple hidden layers, and a fixed number of neurons are constructed to approximate the Chui-Wang spline wavelets arbitrarily well. The constructions in [21] provide a multiresolution analysis based on ZF networks, but are not wavelets in the sense of having vanishing (polynomial) moments.

In this paper, we construct a multiresolution analysis based on ZF networks, and describe frames to span the component spaces of this analysis. Our frames are ZF networks themselves, and have an increasing number of vanishing (polynomial) moments. The networks use a general activation function which is not required to have any special wavelet-like properties. The frames enable us to analyse functions based on data collected at scattered sites. In the usual wavelet theory paradigm, one starts with a large data set to get a projection of the target function in a high order scaling space, and then constructs the wavelet coefficients using reconstruction and decomposition formulas. One novelty of our theory is that we construct build-up frames; i.e., we construct higher and higher order frame transforms using more and more data on an as needed basis. Our frames are obtained as a perturbation of the polynomial frames which we constructed in [15]. Thus, they are also suitable for analysing data in the form of spherical harmonic coefficients of the target function. We note that this data is inherently a global data, but our frames are powerful enough to detect local singularities of the function from such data. Naturally, the power of our ZF frames to detect singularities is no better than that of the original polynomial frames. Nevertheless, the ZF networks have a potential advantage in that they can be implemented in parallel, and even in hardware. From a theoretical point of view, we have developed general techniques to develop frames when we know good approximations to a known frame. Our techniques are expected to be applicable more generally, for example, in the case of Gaussian networks and periodic basis function networks.

The paper being very theoretical, technical, and long, we provide a brief tour of the main ideas in Section 2, and discuss a numerical experiment to illustrate the use of our frames for the detection of singularities over an arc in Section 3. In Section 4, we review certain preliminaries regarding spherical harmonics. In Section 5, we describe the construction of multi-resolution analysis, the ZF frames, and their localization properties. In Section 6, we develop frames in the setting of a separable Hilbert space, using perturbations of a complete orthonormal basis of this space. The proofs of the results in Section 5 are given in Section 7.

2 A brief tour

This section being an informal “tour” of the ideas in the paper, the notation in this section may not confirm with that in the rest of the paper.

The classical wavelet analysis begins with the definition of a multi-resolution analysis. Starting with a scaling function \( \varphi : \mathbb{R} \to \mathbb{R} \), satisfying some technical conditions, the multi-resolution analysis consists of spaces \( V_j = \text{closure span} \{ \varphi(2^j \cdot -k) : k \in \mathbb{Z} \} \), \( j \in \mathbb{Z} \). Denoting by \( W_j \) (the wavelet space) the orthogonal complement of \( V_{j-1} \) in \( V_j \), one can construct a mother wavelet \( \psi \) such that \( W_j = \text{closure span} \{ \psi(2^j \cdot -k) : k \in \mathbb{Z} \} \).
\( k \in \mathbb{Z} \), \( j \in \mathbb{Z} \). Moreover, the functions \( \{ \psi(2^j \cdot -k) \} \) form an orthonormal set. The semi-discrete wavelet transform of \( f \in L^2(\mathbb{R}) \) is defined by

\[
\mathcal{F}_\psi f(x) = \int_{\mathbb{R}} f(t) \psi(2^j x - t) dt.
\] (2.1)

In practice, one computes the integral using numerical integration. Some of the desirable properties of the mother wavelet are: compact support, smoothness, and vanishing moments. Unfortunately, these requirements are usually not compatible; for example, one cannot have a compactly supported, nonzero function all of whose moments are zero.

In [3], Chui and Mhaskar developed a new paradigm in which one allows each wavelet space to have its own mother wavelet, and does not require that the translates and dilates of the mother wavelets be orthogonal to each other. This theme has been developed quite a bit and applied to different problems in applied mathematics (see [18] for some of the references). In particular, it was observed that localized polynomial frames, with the number of vanishing moments tending to infinity, can be defined for an effective detection of discontinuities in a function.

In [15], we built upon the work in [16] to define polynomial frames on the sphere \( S^q \) embedded in \( \mathbb{R}^{q+1} \). We take \( V_j \) to be the space of all spherical polynomials of degree at most \( 2^j \), and define \( W_j \) to be the orthogonal complement of \( V_{j-1} \) in \( V_j \). The role of the classical (dilated) mother wavelet \( \psi(2^j \cdot) \) is played by the univariate polynomial

\[
\psi_j(g; t) := \sum_{\ell=0}^{2^j} g\left(\frac{\ell}{2^j}\right) P_{\ell+1}(q; t),
\]

where \( g : [0, \infty) \to [0, \infty) \) is a smooth function such that \( g(t) = 0 \) if \( t \notin [1/2, 1] \), and \( P_{\ell+1}(q; \cdot) \) are the Legendre polynomials, defined in (4.5). The (semi-discrete) frame transform is defined for \( f \in L^2(S^q) \) by

\[
\tau_j(g; f, x) = \int_{S^q} f(y) \psi_j(g; x \cdot y) d\mu_q(y),
\] (2.2)

where \( \mu_q \) is the surface area measure of the sphere. One advantage of our frame transform is that it is free of the choice of coordinates. Since \( \tau_j(g; f) \in W_j \), it is automatically localized in the spherical frequency band \( 2^{j-1}, \ldots, 2^j \). The localization in the space variable depends upon the smoothness of the function \( g \). Taking \( g \) to be a shifted \( B \)-spline of different orders (and hence, corresponding smoothness), Figure 1 shows the graphs of the polynomials \( \psi_j \). It is evident that the smoother the function \( g \), the better localization we get.

![Figure 1](image-url)

Figure 1: The graph of \(|\psi_6(g; t)|/\psi_6(g; 1)\) for dimension=2 and different \( B \)-splines \( g \): (left) Near \( t = 1 \), (right) away from \( t = 1 \). Orders of the spline: Blue: 3, Green: 5, Red: 7.
If the spherical harmonic coefficients \( \hat{f}(\ell, k) \) are known, the operator \( \tau_j(g; f) \) is easy to calculate:

\[
\tau_j(g; f, \mathbf{x}) = \sum_{\ell=0}^{2^j} g \left( \frac{\ell}{2^j} \right) \sum_{k=1}^{d^j_{\ell}} \hat{f}(\ell, k) Y_{\ell,k}(\mathbf{x}),
\]

(2.3)

where the numbers \( d^j_{\ell} \) and the polynomials \( Y_{\ell,k} \) are defined in Section 4.1. In the important case when \( q = 2 \), MATLAB provides built-in functions to calculate the polynomials \( Y_{\ell,k} \), as well as the \( B \)-splines \( g \) (in the spline toolbox). In practice, one would calculate the coefficients by numerical integration. In [13], we proved the existence of positive quadrature formulas, exact for spherical polynomials of high degree, that enables one not only to do the integration accurately, but prove theorems directly based on the discretization. Using our quadrature formulas (cf. Section 4.2), the evaluation of the frame transform takes the form

\[
\tau_{D,j}(g; f, \mathbf{x}) = \sum_{\xi \in C_j} w_\xi f(\xi) \psi_j(g; \mathbf{x} \cdot \xi),
\]

(2.4)

where \( C_j \) is the set of sites \( \{\xi\} \) at which the training data \( \{(\xi, f(\xi))\} \) is collected. The evaluation of the weights \( w_\xi \) involves the solution of a series of ill-conditioned linear or quadratic programming problems as outlined in [13]. Therefore, this technique is likely to be useful in the case when a number of different target signals \( f \) need to be analysed, but the sites at which the data is collected are fixed. In such applications, the calculation of \( w_\xi \) as well as the polynomials \( \psi_j(g; \mathbf{x} \cdot \xi) \) constitute preprocessing, and the analysis itself involves only a matrix-vector multiplication.

One important consequence of this method of evaluating the frame transform based on the data is the following. In classical wavelet analysis, one evaluates \( \int f(t) \varphi(2^j t - k) dt \) for a large value of \( J \), using numerical integration, and then decomposes the resulting coefficients into scaling and wavelet coefficients using a cascade algorithm. In contrast, the frame transforms \( \tau_{D,j}(g; f) \) can be calculated using larger and larger data sets for larger and larger values of \( j \); one does not need to start out with a huge amount of data to obtain a projection onto high degree polynomials. In this sense, our frames are built up.

In this paper, our objective is to construct ZF network frames that approximate \( \tau_j(g; f) \). A straightforward idea is just to approximate the spherical harmonics \( Y_{\ell,k} \) using networks \( p_{j,\ell,k} \) trained on the data \( \{(\xi, Y_{\ell,k}(\xi))\}_{\xi \in C_j} \), and use these approximations in (2.3). This is indeed the idea which we will follow. However, there are many technical details which need to be taken care of. First, one needs to assure that the resulting frame transform must be localized in the same frequency band as \( \tau_j(g; f) \). In particular, the networks must be orthogonal to polynomials of degree \( 2^j - 1 \), and the higher frequencies (which cannot all be zero) should be small. The networks defined in Proposition 5.1 provide such an approximation. Second, the built up nature of the frames implies that the spherical harmonics in different frequency bands are necessarily approximated by networks trained on different data sets. Therefore, one has to ensure that the resulting scaling spaces are nested, so as to obtain a true multi-resolution analysis. This is done, in contrast to the classical wavelet analysis, by building up the scaling spaces from the wavelet spaces as in (5.18) (Theorem 5.1, Proposition 5.2). Third, the quadrature formulas being exact only for polynomials, and not for ZF networks, the analogue of the formula (2.4) requires a lot more theoretical analysis. In our approach, we will use a different approximation to the spherical harmonics for this purpose (Proposition 5.3). Fourth, it should be possible to reconstruct the networks belonging to the modified wavelet spaces in terms of the operators obtained by approximating the polynomial operators. This is ensured in Theorem 5.2 for the wavelet spaces and Theorem 5.3 for the scaling spaces. Finally, the approximation should be good enough to ensure that the space localization properties are preserved. The space frequency localization properties of our frame transform are discussed in Theorem 5.4.

Although our constructions are meant to be mainly of theoretical interest, we illustrate how the frame transform can be computed using data on the 2-dimensional sphere \( S^2 \), using a ZF network with the activation function \( \phi(t) = 4(17 - 8t)^{-1/2} \).

- **Construction of the quadrature formula**

1. Given a set \( C \) of \( N \) distinct sites on the sphere \( S^2 \) where the data is to be collected, we solve the following quadratic programming problem for as high a value of \( n \) as possible: Let \( \mathbf{x} \) be a
column vector in $\mathbb{R}^N$. Minimize $\|x\|_2$, with $x$ subject to the linear constraints
\[
\sum_j Y_{\ell,k}(\xi_j)x_j = 4\pi Y_{0,0}\delta_{\ell,0}, \quad \ell = 0, \ldots, n, \quad 1 \leq k \leq 2\ell + 1
\]
\[
x_j \geq 0, \quad j = 1, \ldots, N.
\]
This optimization program is designed not only to produce positive weights, but also to get them to be as nearly equal as possible. The strategy is to start with a high value of $n$, say $n \approx \sqrt{N} - 1$, and step it down by 1 until we hit a value of $n$ for which we obtain a solution. We assume that $n \geq 16$. Preliminary calculations with small data suggest that we need $N \geq 4n^2$.

2. Set the quadrature weights $w := x$. In the sequel, we will index $w$ by elements of $C$ rather than by $1, \ldots, N$. Find the largest integer $j \geq 4$ such that $2^{j+2} \leq n$. This is the value of $j$ for which the frame transform can be computed.

- **Construction of basic networks**

  1. Calculate for $\ell = 2^{j-1} + 1, \ldots, 2^{j}, \quad k = 1, \ldots, 2\ell + 1$:

\[
p_{j,\ell,k}(x) := \frac{2\ell + 1}{4\ell+1\pi} \sum_{\xi \in C} w_{\xi} Y_{\ell,k}(\xi) \phi(x \cdot \xi).
\]

  2. Find $a_{(\ell,k),(m,s)} (\ell, m = 2^{j-1} + 1, \ldots, 2^j, \quad s = 1, \ldots, 2m + 1, \quad k = 1, \ldots, 2\ell + 1)$ such that for $\nu = 2^{\ell-1} + 1, \ldots, 2^\ell$, $r = 1, \ldots, 2\nu + 1$:

\[
\sum_{m=2^{j-1}+1}^{2^j} \sum_{s=1}^{2m+1} a_{(\ell,k),(m,s)} \sum_{\xi \in C} w_{\xi} p_{j,m,s}(\xi) Y_{\nu,r}(\xi) = \begin{cases} 1, & \text{if } \ell = \nu \text{ and } k = r; \\ 0, & \text{otherwise}, \end{cases}
\]

  3. Define the networks

\[
q_{j,\ell,k}(x) := \sum_{m=2^{j-1}+1}^{2^j} \sum_{s=1}^{2m+1} a_{(\ell,k),(m,s)} p_{j,m,s}(x)
\]
\[
= \frac{2\ell + 1}{4\ell+1\pi} \sum_{\xi \in C} w_{\xi} \left\{ \sum_{m=2^{j-1}+1}^{2^j} \sum_{s=1}^{2m+1} a_{(\ell,k),(m,s)} Y_{m,s}(\xi) \right\} \phi(x \cdot \xi)
\]
\[
=: \sum_{\xi \in C} B_{j,\ell,k} \phi(x \cdot \xi).
\]

- **Analysis of the “training data”** $\{(\xi, f^*(\xi))\}_{\xi \in C}$

  1. Define the discrete coefficients

\[
F_{D,\ell,k} := \sum_{\xi \in C} w_{\xi} f^*(\xi) Y_{\ell,k}(\xi).
\]

  2. With a smooth function $g : [0, \infty) \to [0, \infty)$ such that $g(t) = 0$ if $t \notin [1/2, 1]$, define

\[
\sum_{\xi \in C} \left\{ \sum_{\ell=2^{j-1}+1}^{2^j} \sum_{k=1}^{2\ell+1} g(2^{-j} \ell) F_{D,\ell,k} B_{j,\ell,k} \right\} \phi(x \cdot \xi).
\]

The last network is the desired frame transform.

In the case when the spherical harmonic coefficients of $f$ are known, the construction is substantially simpler. One computes a slightly modified polynomial frame transform (cf. (2.3))

\[
\tau_j'(g; f, \xi) = w_{\xi} \sum_{\ell=2^{j-1}+1}^{2^j} g(2^{-j} \ell) \frac{2\ell + 1}{4\ell+1\pi} \sum_{k=1}^{2\ell+1} \hat{f}(\ell, k) Y_{\ell,k}(\xi), \quad \xi \in C
\]

5
and the desired frame transform is the network

\[ \sum_{\xi \in \mathcal{C}} \tau_j^f(g; f, \xi) \phi(x; \xi). \]

We highlight certain salient features of our constructions. Our networks are easy to train; in fact, there is no training in the classical sense. We have explicit formulas constructing all the approximations, the computational details of which can be taken care of during preprocessing, once the sites are selected. Different target functions can then be analysed by simple matrix vector multiplications using the “training data” directly. Our networks are not merely approximations to existing frames; they are themselves frames in the sense of spanning the wavelet spaces, stability, and vanishing polynomial moments. They are built up frames, in the sense that frame transforms in higher and higher order wavelet spaces are constructed using larger and larger data sets, rather than projecting down from a huge data set to begin with. In principle, we do not require the sites at which the data is collected to be chosen in any particular manner, but equidistributed sites are found to yield more satisfactory numerical results. The attenuation factors based on the function \( g \) help us to adjust the frames to have the desired space localizations. The approximation properties of our frames as well as their ability to detect singularities have been analysed theoretically and quantitatively.

### 3 Numerical Experiments

In this section, we illustrate the capabilities of our polynomial frames to detect singularities in a function, given its spherical harmonic coefficients. For simplicity of computations, we use operators analogous to \( \tau_j^f \) defined in (5.27) below (which are just the operators \( \tau_j(g; f) \) defined in (2.2), but in a more general setting), but giving polynomials of degree at most 50. The coefficients \( g_j, \ell \)'s are obtained as equidistant samples of a shifted \( B \)-spline of order 3, with support \([1/2, 1]\). We consider the function

\[ f(x) = \int_{\pi/4}^{\pi/4} \log(1 - x \cdot y)d\theta(y) + (x \cdot w - 0.95)^2, \]

where \( w = \sqrt{1/3}(1, -1, -1) \), and the integral is the line integral with respect to the spherical coordinate \( \theta \) in the system

\[ y = (\cos \theta(y) \sin \phi(y), \sin \theta(y) \sin \phi(y), \cos \phi(y)). \]

Figure 2 shows the detection of the logarithmic singularity along the correct arc. Part (a) in Figure 3 shows a two-dimensional view of the same, and part (b) shows the simultaneous detection of the “tornado” singularity in the second order derivative at \( w \), which is the center of the ring. To illustrate the effect of the choice of the coefficients \( g_j, \ell \)'s we show in Figure 4 a close-up two dimensional view of the detection of logarithmic singularity using sampled values of a shifted \( B \)-spline with support \([1/2, 1]\) of order 3 (part (a)) and order 1 (part (b)). It appears that the spline of order 1 gives a slightly better localized detection than that of order 3. However, as Figure 5 shows, it is not possible to detect simultaneously the singularity of the second order derivative using the spline of order 1. We remark that the nondetection is only due to the simultaneous presence of a logarithmic singularity. If the tornado singularity alone is present, it can still be detected using a spline of order 1.

### 4 Preliminaries

#### 4.1 Spherical Harmonics

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

\[ \omega_q := \int_{S^q} d\mu_q = \frac{2\pi^{(q+1)/2}}{\Gamma((q + 1)/2)}, \]

(4.1)
where the gamma function is defined, as usual, by

$$
\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt, \quad \Re z > 1.
$$

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

$$
(f, g)_{S^q} := \int_{S^q} f(x)g(x)d\mu_q(x)
$$

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

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

For integer $\ell \geq 0$, the restriction to $S^q$ of a homogeneous harmonic polynomial of degree $\ell$ is called a spherical harmonic of degree $\ell$. Most of the following information is based on [19] and [27, §IV.2], although we use a different notation. The class of all spherical harmonics of degree $\ell$ will be denoted by $H^q_{\ell}$, and the class of all spherical harmonics of degree $\ell \leq n$ will be denoted by $\Pi^q_n$. Of course, $\Pi^q_n = \bigoplus_{\ell=0}^n H^q_{\ell}$, and it comprises the restriction to $S^q$ of all algebraic polynomials in $q + 1$ variables of total degree not exceeding $n$. The dimension of $H^q_{\ell}$ is given by

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

and that of $\Pi^q_n$ is $\sum_{\ell=0}^n d^q_{\ell}$. The spaces $H^q_{\ell}$'s are mutually orthogonal relative to (4.2); also, $L^2(S^q) =$ closure$\{ \bigoplus_{\ell} H^q_{\ell} \}$. Hence, if we choose an orthonormal basis $\{ Y_{\ell,k} : k = 1, \cdots, d^q_{\ell} \}$ for each $H^q_{\ell}$, then the
Figure 3: Two-dimensional view of the detection of logarithmic (left) and tornado (right) singularities

Figure 4: Detection of the logarithmic singularity using splines of order 3 (left) and 1 (right)

set \( \{ Y_{\ell,k} : \ell = 0, 1, \ldots \text{ and } k = 1, \ldots, d^q_\ell \} \) is an orthonormal basis for \( L^2(S^q) \). One has the well-known addition formula [19]:

\[
\sum_{k=1}^{d^q_\ell} Y_{\ell,k}(x)Y_{\ell,k}(y) = \frac{d^q_\ell}{\omega_q} \mathcal{P}_\ell(q + 1; x \cdot y), \quad \ell = 0, 1, \ldots, (4.5)
\]

where \( \mathcal{P}_\ell(q + 1; x) \) is the degree-\( \ell \) Legendre polynomial in \( q + 1 \)-dimensions. (We note that Müller’s \( \omega_q \) and \( N(q, \ell) \) are the same as our \( \omega_{q+1} \) and \( d_{\ell+1}^{q+1} \).)

The Legendre polynomials are normalized so that \( \mathcal{P}_\ell(q + 1; 1) = 1 \), and satisfy the orthogonality relations [19, Lemma 10]

\[
\int_{-1}^{1} \mathcal{P}_\ell(q + 1; x) \mathcal{P}_k(q + 1; x) (1 - x^2)^{\frac{q-1}{2}} dx = \frac{\omega_q}{\omega_{q-1} d_{\ell}^q} \delta_{\ell,k}. \quad (4.6)
\]

They are related to the ultraspherical (Gegenbauer) polynomials \( P_{\ell}^{(q-1)} \) (cf. [28], [19, p. 33]), and the
Jacobi polynomials, $P^{(\alpha,\beta)}_{\ell}$, with $\alpha = \beta = \frac{q}{2} - 1$, via

$$P^{(\frac{q}{2} - \frac{1}{2})}_{\ell}(x) = \left(\frac{\ell + q - 2}{\ell}\right)P_{\ell}(q + 1; x) \quad (q \geq 2)$$

$$P^{(\frac{q}{2} - \frac{3}{2})}_{\ell}(x) = \left(\frac{\ell + \frac{q}{2} - 1}{\ell}\right)P_{\ell}(q + 1; x).$$

When $q = 1$, the Legendre polynomials $P_{\ell}(2; x)$ coincide with the Chebyshev polynomials $T_{\ell}(x)$; the ultraspherical polynomials $P_{\ell}^{(0)}(x) = (2/\ell)T_{\ell}(x)$, if $\ell \geq 1$. For $\ell = 0$, $P_{0}^{(0)}(x) = 1$. From the fact that $\Pi_{n}^{q} = \bigoplus_{\ell=0}^{n} H_{l}^{q}$, and the addition formula (4.5), we obtain that for any $P \in \Pi_{n}^{q}$ and $x \in S^{q}$,

$$P(x) = \sum_{\ell=0}^{n} \frac{d_{\ell}^{q}}{\omega_{q}} \int_{S^{q}} P(y)P_{\ell}(q + 1; x \cdot y)d\mu_{q}(y).$$

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

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

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

Finally, we note that the Funk-Hecke formula [19, Theorem 6] implies the following useful connection between integrals over $S^{q}$ and integrals over $[-1, 1]$ with respect to the weight function $w_{q}$. For any $\phi \in L_{w_{q}}^{1}[-1, 1]$, $y \in S^{q}$, and any $Y_{\ell} \in H_{l}^{q}$, we have

$$\int_{S^{q}} \phi(x \cdot z)Y_{\ell}(z)d\mu_{q}(z) = \omega_{q-1} Y_{\ell}(x) \int_{-1}^{1} \phi(t)P_{\ell}(q + 1; t)w_{q}(t)dx$$
\[ \omega_q \frac{\hat{\phi}(\ell)}{d_\ell^q} Y_\ell(x). \]  \hspace{1cm} (4.13)

We remark that our definition of \( \hat{\phi} \) in (4.13) is chosen so that the Legendre expansion of \( \phi \) has the form \( \sum \hat{\phi}(\ell) P_\ell(q + 1; \cdot) \).

### 4.2 Quadrature formula

Let \( C \) be a finite set of distinct points on \( S^q \). The mesh norm of \( C \) is defined to be

\[ \delta_C := \sup_{x \in S^q} \text{dist}(x, C). \]  \hspace{1cm} (4.14)

The following theorem summarizes the quadrature formula given in [13]. In the sequel, we adopt the following convention regarding constants. The letters \( c, c_1, \cdots \) will denote positive constants depending on the dimension \( q \), and the different norms involved in the formula. Their value will be different at different occurrences, even within the same formula. The symbol \( A \sim B \) will mean \( cA \leq B \leq c_1 A \).

**Theorem 4.1** There exist constants \( \alpha_q \) and \( N_q \) with the following property. Let \( C \) be a finite set of distinct points on \( S^q \), and \( n \) be an integer with \( N_q \leq n \leq \alpha_q \delta_C^{-1} \). Then there exist nonnegative weights \( \{a_\xi\}_{\xi \in C} \), such that for every \( P \in \Pi_n^q \),

\[ \int_{S^q} P(x) d\mu_q(x) = \sum_{\xi \in C} a_\xi P(\xi), \]  \hspace{1cm} (4.15)

and

\[ |\{\xi : a_\xi \neq 0\}| \sim n^q \sim \text{dim}(\Pi_n^q). \]  \hspace{1cm} (4.16)

In [13], we have discussed algorithms to compute \( a_\xi \). The measure that associates the mass \( a_\xi \) with \( \xi \in C \) will be denoted by \( \nu_C \).

### 5 ZF network frames

#### 5.1 Notations and assumptions

The essential ingredients of our frames are: the activation function \( \phi \), the data sets \( C_j \) and certain matrices. We now describe these and introduce some notation.

**The activation function**

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

\[ \hat{\phi}(\ell) \neq 0, \quad \ell = 0, 1, 2, \cdots, \]  \hspace{1cm} (5.1)

where \( \hat{\phi}(\ell) \) is defined in (4.13). In view of the Funk-Hecke formula (4.12), the class of all ZF networks is not dense in \( C(S^q) \) without this condition; in fact, if \( \hat{\phi}(\ell) = 0 \) for some \( \ell \) then this class is orthogonal to \( H_\ell^q \). For integers \( n = 0, 1, \cdots \), we write

\[ E_{n,p}(\phi) := \min \|\phi - P\|_{w_q,p}, \]  \hspace{1cm} (5.2)

where the minimum is taken over all univariate polynomials \( P \) of degree not exceeding \( n \). We will also use the notations

\[ m_N := \min_{0 \leq \ell \leq N} \left| \frac{\hat{\phi}(\ell)}{d_\ell^q} \right|, \quad N = 1, 2, \cdots, \]  \hspace{1cm} (5.3)

\[ N_j := \sum_{\ell=0}^{2^j} d_\ell^q, \quad j = 0, 1, 2, \cdots. \]  \hspace{1cm} (5.3)
We define \( N_{-1} := -1 \). Let integers \( M_j \) be chosen so that
\[
\epsilon_j := \frac{2}{\sqrt{\omega q}} \frac{E_{M_j, \infty}(\phi)}{m_2}, \quad j = 0, 1, 2, \ldots.
\]

We need also to make a couple of more assumptions of a more technical nature. Let
\[
\nu(n, j) := \max(2^n, M_j),
\]
and
\[
\delta_n(\phi) := \frac{2}{\sqrt{\omega q}} \frac{\tilde{E}_{\nu(n, j)}(\phi)}{m_2}.
\]

We assume that
\[
\lim_{j \to \infty} j^{2n/2} \delta_j = 0,
\]
and
\[
\sum_{n=0}^{\infty} (n+1)2^{nq/2} \delta_n < \infty.
\]

We pause in our discussion to note two important examples discussed in \[14\]. First, let \( q \geq 2, 0 < \rho < 1, \) and
\[
\phi_p^G(x) := (1 - 2\rho x + \rho^2)^{-(q-1)/2}, \quad x \in [-1, 1].
\]

Then for every \( n \geq c, \) we have
\[
\frac{E_{4n, \infty}(\phi_p^G)}{m_n} \leq \frac{\tilde{E}_{4n, \infty}(\phi_p^G)}{m_n} \leq cn \rho^n. \tag{5.10}
\]

As a second example, we consider the function
\[
\phi_p^E(x) := \exp(\rho x), \quad x \in [-1, 1], \tag{5.11}
\]
where \( \rho \in \mathbb{R}, \rho \neq 0 \). To underline the fact that, in contrast to the theory of interpolation of functions, we do not require \( \phi \) to be positive definite, we note that \( \rho \) may be any nonzero real number, positive or negative. Let \( t > 0 \) be a fixed number, and for \( n \geq c, \) take \( M = [n(1 + t)] \). Then
\[
\frac{E_{M, \infty}(\phi_p^E)}{m_n} \leq \frac{\tilde{E}_{M, \infty}(\phi_p^E)}{m_n} \leq c(\rho)n^{-tn/2}. \tag{5.12}
\]

Thus, the conditions (5.4), (5.7), and (5.8) are all satisfied in each case with \( M_j = c2^j \) for \( j \geq J_0, \) where \( J_0 \) is some suitably large integer. In these cases we have to start the scaling spaces with the index \( J_0 \) rather than 0.

**The data sets**

Next, we assume that for each \( j, \) there exists a set of points \( C_j \) so that \( M_j + 2^{j+1} \leq \alpha_q \delta_j^{-1} \). Therefore, there exists a quadrature formula (4.15) that is exact for polynomials of degree at most \( M_j + 2^{j+1} \). Let \( \nu_{C_j} \) be the discrete measure occuring in this quadrature formula. The inner product corresponding to \( \nu_{C_j} \) will be denoted by \( \langle \cdot, \cdot \rangle_{j, \delta}, \) and the corresponding norm by \( \| \cdot \|_{j, \delta}. \)

**The matrices**

A matrix \( G \) will be called a scaling matrix if \( g_{j, k} \neq 0 \) for \( k = 0, \ldots, N_j, \) \( j = 0, 1, \ldots, \) and \( g_{j, k} = 0 \) otherwise. Similarly, a matrix \( G \) will be called a frame matrix if \( g_{j, k} \neq 0 \) for \( k = N_{j-1} + 1, \ldots, N_j, \) \( j = 0, 1, 2, \ldots, \) and \( g_{j, k} = 0 \) otherwise. For an integer \( k \) and a matrix \( G, \) we define the matrix \( G^{[k]} \) by
\[
G_{j, t}^{[k]} = \begin{cases} (g_{j, t})^k, & \text{if } g_{j, t} \neq 0, \\ 0, & \text{otherwise}. \end{cases}
\]

11
5.2 Multi-resolution analysis

The building blocks of our multi-resolution analysis are the networks $p_{j;\ell,k}$ defined for $j = 0, 1, \ldots$, $\ell = 0, 1, \ldots, 2^j$, and $k = 1, \ldots, d^j_\ell$ by

$$p_{j;\ell,k}(x) := \frac{d^j_\ell}{\omega_q \phi(\ell)} \int_{\mathbb{S}^q} \phi(x \cdot \xi) Y_{\ell,k}(\xi) d\nu_\xi(\xi).$$

(5.13)

We observe that each of the measures $\nu_\xi$ is a discrete measure, with the number of points in its support being $O(M_j + 2^j)^q$ (see (4.16)). Therefore, each $p_{j;\ell,k}$ is a ZF network, and the number of neurons involved is $O(M_j + 2^j)^q$. Further, for any complex numbers $a_{\ell,k}$,

$$
\sum_{\ell=0}^{2^j} \sum_{k=1}^{d^j_\ell} a_{\ell,k} \sigma_{\ell,k} = \int_{\mathbb{S}^q} \phi(x \cdot \xi) \left\{ \sum_{\ell=0}^{2^j} \frac{d^j_\ell}{\omega_q \phi(\ell)} \sum_{k=1}^{d^j_\ell} a_{\ell,k} Y_{\ell,k}(\xi) \right\} d\nu_\xi(\xi)
$$

can be implemented also using $O(M_j + 2^j)^q$ neurons. The following proposition lists some important properties of these networks. (The proof of this proposition is given in Section 6, along with all other new statements in Section 5.)

**Proposition 5.1** For $j = 0, 1, \ldots, k = 1, \ldots, d^j_\ell$, $\ell = 0, 1, \ldots, 2^j$, we have

$$
\int_{\mathbb{S}^q} p_{j;\ell,k}(x) Y_{m,s}(x) d\mu_q(x) = \delta_{(\ell,k),(m,s)}, \quad s = 1, \ldots, d^j_m, \quad m = 0, \ldots, 2^j.
$$

(5.14)

We have

$$
\max_{0 \leq \ell \leq 2^j, 1 \leq k \leq d^j_\ell} \| p_{j;\ell,k} - Y_{\ell,k} \|_{\mathcal{B}(\mathbb{S}^q)} \leq \frac{2 \ E_{M_j,\infty}(\phi)}{m_{2^j}} =: \epsilon_j,
$$

(5.15)

and hence,

$$
\max_{0 \leq \ell \leq 2^j, 1 \leq k \leq d^j_\ell} \| p_{j;\ell,k} - Y_{\ell,k} \|_{j,2} \leq \sqrt{\omega_q} \epsilon_j.
$$

(5.16)

Further, for $n = 0, 1, \ldots$,

$$
\max_{0 \leq j \leq n, 0 \leq \ell \leq 2^j, 1 \leq k \leq d^j_\ell} \| p_{j;\ell,k} - \text{proj}_{H_{2^j_n}}(p_{j;\ell,k}) \|_{\mathcal{B}(\mathbb{S}^q)} \leq \delta_n,
$$

(5.17)

where $\delta_n$ is defined in (5.6).

Using the set $\{p_{j;\ell,k}\}$, we may define the wavelet spaces $W_j$ and the scaling spaces $V_j$ as follows:

$$
W_j := \text{span}\{p_{j;\ell,k} : k = 1, \ldots, d^j_\ell, \ell = 2^{j-1} + 1, \ldots, 2^j\}, \quad j = 1, 2, \ldots,
$$

$$
V_j := \text{span}\{p_{0;\ell,0} : k = 1, \ldots, d^j_0, \ell = 0, 1\},
$$

$$
V_j := V_0^j + W_1^j + \cdots + W_j^j, \quad j = 1, 2, \ldots.
$$

(5.18)

We also define $V_j^0 := \{0\}$, $W_0^j := V_j^0$.

**Theorem 5.1** For $j = 1, 2, \ldots$, $V_j^j \subseteq V_{j+1}^j$, $V_j^j = V_{j-1}^j + W_j^j$, and $W_j^j$ is orthogonal to $V_{j-1}^j$. Further, there exists $J_0$ such that for $j \geq J_0$, $W_j^j \cap V_{j-1}^j = \{0\}$. Moreover, $\bigcup_{j=0}^\infty V_j^j$ is dense in $L^2(\mathbb{S}^q)$.

Thus, $\{V_j^j\}$ defines a multi-resolution analysis of $L^2(\mathbb{S}^q)$. In contrast to the usual practice, the wavelet spaces $W_j^j$ are required to have an increasing number of vanishing moments, rather than being orthogonal to $V_{j-1}^j$. Our proof will show that the cosine of the angle between $W_j^j$ and $V_{j-1}^j$ tends to 0 as $j \to \infty$.

In the sequel, we will need another basis for the scaling spaces. We define recursively for all integers $j$:

$$
\tilde{p}_{0;\ell,k} := p_{0;\ell,k}, \quad k = 1, \ldots, d^j_\ell, \quad \ell = 0, 1,
$$

$$
\tilde{p}_{j+1;\ell,k} := \begin{cases}
\tilde{p}_{j;\ell,k} + \sum_{m=2^{j+1}}^{2^{j+1}+1} \sum_{s=1}^{d^j_m} \langle \tilde{p}_{j;\ell,k}, Y_{m,s} \rangle Y_{j+1,m,s}, & \text{if } 1 \leq k \leq d^j_\ell, 0 \leq \ell \leq 2^j, j \geq 0, \\
\tilde{p}_{j+1;\ell,k}, & \text{if } k = 1, \ldots, d^j_\ell, 2^j + 1 \leq \ell \leq 2^{j+1}, j \geq 0.
\end{cases}
$$

(5.19)
Proposition 5.2 For \( j = 0, 1, \cdots, k = 1, \cdots, d_j^\ell, \ell = 0, 1, \cdots, 2^j \), we have
\[
\int_{S^q} \tilde{p}_{j,\ell,k}(x) Y_{m,s}(x) d\mu_q(x) = \delta_{(\ell,k),(m,s)}, \quad s = 1, \cdots, d_m^s, \quad m = 0, \cdots, 2^j.
\]
(5.20)

We have
\[
\max_{0 \leq \ell \leq 2^j, 1 \leq k \leq d_{q,\ell}} \| \tilde{p}_{j,\ell,k} - Y_{\ell,k} \|_{�^2,\infty} \leq C_1(j + 1) \delta_j,
\]
(5.21)

where \( C_1 \) is a positive constant, and \( \delta_j \) is defined in (5.6). Hence,
\[
\max_{0 \leq \ell \leq 2^j, 1 \leq k \leq d_{q,\ell}} \| \tilde{p}_{j,\ell,k} - Y_{\ell,k} \|_{j;\infty,2} \leq \sqrt{\omega_q} C_1(j + 1) \delta_j.
\]
(5.22)

Further, for \( n = 0, 1, \cdots, \),
\[
\max_{0 \leq j \leq n, 0 \leq \ell \leq 2^j, 1 \leq k \leq d_{q,\ell}} \| \tilde{p}_{j,\ell,k} - \text{proj}_{2n}(\tilde{p}_{j,\ell,k}) \|_{\infty,2} \leq c(n + 1) \delta_n,
\]
(5.23)

If \( j \) is such that \( \sqrt{\omega_q} C_1 \sqrt{N_j + 1} \delta_j < 1/2 \), there exist \( \tilde{q}_{j,\ell,k} \subset V_j^\infty \) such that
\[
\int_{S^q} \tilde{q}_{j,\ell,k}(x) Y_{m,s}(x) d\nu_q(x) = \delta_{(\ell,k),(m,s)}, \quad s = 1, \cdots, d_m^s, \quad m = 0, \cdots, 2^j,
\]
(5.24)

and
\[
\max_{0 \leq \ell \leq 2^j, 1 \leq k \leq d_{q,\ell}} \| \tilde{q}_{j,\ell,k} - Y_{\ell,k} \|_{j;\infty,2} \leq 4 \sqrt{\omega_q} C_1(j + 1) \delta_j.
\]
(5.25)

The same bound holds also with \( \| \cdot \|_{j;\infty,2} \) replaced by \( \| \cdot \|_{\infty,2} \). Finally, the bound (5.23) holds with \( \tilde{p} \) replaced by \( \tilde{q} \), with a possibly different constant \( c \).

5.3 ZF frames

We now describe the construction and localization properties of our frames. Our constructions are obtained by perturbing the corresponding constructions in [15] for polynomials. First, we review these constructions briefly.

Our scaling spaces in [15] are defined by \( V_j^\Pi := \Pi_j^\Pi, j = 0, 1, \cdots \). The wavelet spaces \( W_j^\Pi \) are defined for \( j = 1, 2, \cdots \) by
\[
W_j^\Pi := \{ P \in V_j^\Pi : \int_{S^q} P(x) R(x) d\mu_q(x) = 0, \quad R \in V_{j-1}^\Pi \}.
\]

We defined the kernel functions for \( x \in S^q, j = 0, 1, \cdots \), by:
\[
K_j^\Pi(G; x, y) := \sum_{\ell=0}^{2^j} g_{j,\ell} \frac{d_{q,\ell}^2}{\omega_q} \mathcal{P}_\ell(x \cdot y) = \sum_{\ell=0}^{2^j} g_{j,\ell} \sum_{k=1}^{d_{q,\ell}^2} Y_{\ell,k}(x) \tilde{Y}_{\ell,k}(y),
\]
(5.26)

and the operators
\[
\tau_j^\Pi(G; f, x) := \int_{S^q} K_j^\Pi(G; x, y) f(y) d\mu_q(y), \quad f \in L^1(S^q).
\]
(5.27)

The operator \( \tau_j \) is analogous to a windowed inverse Fourier transform. One may also think of it as a summability method applied to the middle part of the Fourier series of \( f \). Introducing the matrix \( G \) enabled us to state certain results on the stability, reconstruction, and decomposition of functions in a unified manner.

We define the ZF network frame analogues of these kernels by replacing \( Y_{\ell,k}(x) \) in (5.26) by the ZF network \( p_{j,\ell,k}(x) \). Thus, for a frame matrix \( G \), we define the kernel function for \( x, y \in S^q \) and \( j = 1, 2, \cdots \) by
\[
K_j^Z(G; x, y) := \sum_{\ell=0}^{2^j} g_{j,\ell} \sum_{k=1}^{d_{q,\ell}^2} p_{j,\ell,k}(x) \tilde{Y}_{\ell,k}(y),
\]
(5.28)
and the continuous frame operators

\[ \tau_j^Z(G; f, x) := \int_{\mathbb{S}^q} K_j^Z(G; x, y) f(y) d\mu_q(y), \quad f \in L^1(\mathbb{S}^q). \] (5.29)

To construct the discrete frames, we need the following proposition.

**Proposition 5.3** For \( j = 1, \ldots, \), there exist \( \{q_{j,\ell,k}\} \subseteq W_j, \) \( k = 1, \ldots, d_{\ell}^q, \) \( \ell = 2^{j-1} + 1, \ldots, 2^{j}, \) such that

\[ \int_{\mathbb{S}^q} q_{j,\ell,k}(x) Y_{m,s}(x) d\nu_{j} = \delta_{(\ell,k),(m,s)}, \quad s = 1, \ldots, d_{m}^q, \quad m = 2^{j-1} + 1, \ldots, 2^{j}. \] (5.30)

We have

\[ \max_{1 \leq k \leq d_{\ell}^q} \|q_{j,\ell,k} - Y_{\ell,k}\|_{j;\mathbb{S}^q,2} \leq 4\sqrt{\omega_q} c_j, \quad \ell = 2^{j-1} + 1, \ldots, 2^{j}. \] (5.31)

The same bound holds also with \( \|\cdot\|_{j;\mathbb{S}^q,2} \) replaced by \( \|\cdot\|_{\mathbb{S}^q,2}. \)

For a frame matrix \( G, \) we define the discrete kernel functions for \( x, y \in \mathbb{S}^q \) and \( j = 1, 2, \ldots \) by

\[ K_{D,j}^Z(G; x, y) := \sum_{\ell=0}^{2^j} q_{j,\ell} \sum_{k=1}^{d_{\ell}^q} q_{j,\ell,k}(x) \overline{Y_{\ell,k}(y)}, \] (5.32)

and the discrete frame operators

\[ \tau_{D,j}^Z(G; f, x) := \int_{\mathbb{S}^q} K_j^Z(G; x, y) f(y) d\nu_j(y), \quad f : \mathcal{C}_j \to \mathbb{C}. \] (5.33)

The following theorem describes the reconstruction of elements of \( W_j^Z \) in terms of the continuous and discrete frame operators, and their stability.

**Theorem 5.2** Let \( G \) be a frame matrix, and \( j \geq 1 \) be an integer. If \( f \in W_j^Z, \) then

\[ f(x) = \int_{\mathbb{S}^q} K_j^Z(G; x, y) \tau_j^Z(G^{[-1]}; f, y) d\mu_q(y), \] (5.34)

\[ = \int_{\mathbb{S}^q} K_{D,j}^Z(G; x, y) \tau_{D,j}^Z(G^{[-1]}; f, y) d\nu_j(y). \] (5.35)

Moreover, with

\[ B_{j,2}(G) := \max_{0 \leq \ell \leq 2^j} |g_{j,\ell}|, \] (5.36)

we have

\[ \left\{ \sqrt{3/2 B_{j,2}(G^{[-1]})} \right\}^{-1} \|f\|_{\mathbb{S}^q,2} \leq \|\tau_j^Z(G; f)\|_{\mathbb{S}^q,2} \leq \sqrt{3/2 B_{j,2}(G)} \|f\|_{\mathbb{S}^q,2}, \] (5.37)

\[ \left\{ \sqrt{3} B_{j,2}(G^{[-1]}) \right\}^{-1} \|f\|_{\mathbb{S}^q,2} \leq \|\tau_{D,j}(G; f)\|_{\mathbb{S}^q,2} \leq \sqrt{3} B_{j,2}(G) \|f\|_{\mathbb{S}^q,2}. \] (5.38)

Next, we discuss the scaling functions. Let \( G \) be a scaling matrix. We define the scaling kernel for \( j = 0, 1, \ldots \) by

\[ K_j^Z(G; x, y) := \sum_{\ell=0}^{2^j} g_{j,\ell} \sum_{k=1}^{d_{\ell}^q} Y_{\ell,k}(y) \overline{Y_{j,\ell,k}(x)}, \] (5.39)

and the scaling operator by

\[ \tau_j^Z(G; f) := \int_{\mathbb{S}^q} f(y) K_j^Z(G; x, y) f(y) d\mu_q(y). \] (5.40)
The discrete scaling kernel is defined by

\[ K_{D,j}^z(G; x, y) := \sum_{\ell=0}^{2^j} \sum_{k=1}^{d_j^q} Y_{\ell,k}(y) \hat{q}_{j,\ell,k}(x), \]

where the functions \( \hat{q}_{j,\ell,k} \) are defined in Proposition 5.2, and the discrete scaling transform by

\[ \tau_{D,j}^z(G; f, x) := \int_{\mathbb{S}^q} f(y) K_{D,j}^z(G; x, y) f(y) d\nu_C(y). \]

**Theorem 5.3** Let \( G \) be a scaling matrix, and \( j \geq 0 \) be an integer. If \( f \in V_j^2 \), then

\[ f(x) = \int_{\mathbb{S}^q} K_j^z(G; x, y) \tau_{D,j}^z(G^{[-1]}; f, y) d\mu_q(y), \]

\[ = \int_{\mathbb{S}^q} K_{D,j}^z(G; x, y) \tau_{D,j}^z(G^{[-1]}; f, y) d\nu_C(y). \]

Moreover,

\[ \left\{ \sqrt{3/2} B_{j,2}(G^{[-1]}) \right\}^{-1} \| f \|_{\mathbb{S}^q,2} \leq \| \tau_{D,j}^z(G; f) \|_{\mathbb{S}^q,2} \leq \sqrt{3/2} B_{j,2}(G) \| f \|_{\mathbb{S}^q,2}, \]

\[ \left\{ \sqrt{3} B_{j,2}(G^{[-1]}) \right\}^{-1} \| f \|_{\mathbb{S}^q,2} \leq \| \tau_{D,j}^z(G; f) \|_{j;\mathbb{S}^q,2} \leq \sqrt{3} B_{j,2}(G) \| f \|_{\mathbb{S}^q,2}. \]

Next, we discuss the localization properties of our kernel functions, \( K_j^z(G; x, y) \) and \( K_{D,j}^z(G; x, y) \). Without loss of generality, we may assume that \( y = e_{q+1} := (0, \cdots, 0, 1) \). Following [21], we may measure the frequency localization of a function

\[ \psi := \sum_{\ell=0}^{\infty} \sum_{k=1}^{d_j^q} a_{\ell,k} Y_{\ell,k} \in L^2(\mathbb{S}^q) \]

by

\[ \text{var}_F(\psi) := \left\{ \sum_{\ell=0}^{\infty} \sum_{k=1}^{d_j^q} |a_{\ell,k}|^2 \right\}^{-1} \sum_{\ell=0}^{\infty} \sum_{k=1}^{d_j^q} \ell (\ell + q - 1) |a_{\ell,k}|^2. \]

Obviously, when we take \( \psi \) to be \( K_j^{\Pi}(G, x, e_{q+1}) \), we have the estimate

\[ \text{var}_F(\psi) \leq c4^j. \]

The following proposition asserts that the same bound is valid for the functions \( K_j^z(G, x, e_{q+1}) \) or \( K_{D,j}^z(G, x, e_{q+1}) \).

**Proposition 5.4** Let \( \psi \) be any of the functions \( K_j^z(G, x, e_{q+1}) \) or \( K_{D,j}^z(G, x, e_{q+1}) \), where \( G \) may be a scaling or frame matrix, and

\[ (j + 1)\delta_j \leq c2^{-j(q/2+j)}, \quad j = 0, 1, 2, \cdots. \]

Then (5.48) holds.

A measurement for space localization of \( \psi \in L^2(\mathbb{S}^q) \) is defined by the expression \( \text{var}_S(\psi) \) as in the following formula (5.51). Let

\[ T(\psi) := \| \psi \|_{\mathbb{S}^q,2}^{-2} \int_{\mathbb{S}^q} x|\psi(x)|^2 d\mu_q(x). \]

Then we define (cf. [21])

\[ \text{var}_S(\psi) := \frac{1}{T(\psi)} \left( \frac{\psi}{T(\psi)} \right)^2, \]
where \( \|T(\psi)\| \) is the Euclidean norm of the \( q + 1 \)-dimensional vector \( T(\psi) \). In the case when \( q = 2 \), it is proved in [21] that \( \text{var}_S(\psi) \text{var}_F(\psi) \geq 1 \) for all \( \psi \in L^2(\mathbb{S}^q) \). Rösler and Voit [25] have recently shown that the lower bound of 1 is sharp. In [15], we have studied the localization properties of \( K^1_j(G, x, e_{q+1}) \). The following theorem gives an analogue of our results in [15] in the case of the kernels \( K^Z_j(G, x, e_{q+1}) \) and \( K^Z_{D,j}(G, x, e_{q+1}) \).

**Theorem 5.4** Suppose

\[
E_{M_j, \infty}(\phi) \leq c2^{-(j/2+2j)}, \quad j = 0, 1, 2, \ldots. \tag{5.52}
\]

Let \( G \) be a frame matrix, and

\[
\hat{g}_{j, \ell} := g_{j, \ell}(d \omega_j^{2}/\omega_q)^{1/2} \left\{ \sum_{\ell=0}^{\infty} |g_{j, \ell}|^2(d \omega_j^{2}/\omega_q) \right\}^{-1/2}, \quad j = 0, 1, \ldots, \ell = 0, 1, \ldots. \tag{5.53}
\]

If

\[
\frac{\hat{g}_{j, \ell+1} + \hat{g}_{j, \ell-1}}{2} = \hat{g}_{j, \ell} (1 + O(4^{-j})) \tag{5.54}
\]

then taking \( \psi \) to be any of the functions \( K^Z_j(G, x, e_{q+1}) \) or \( K^Z_{D,j}(G, x, e_{q+1}) \), we have

\[
\text{var}_S(\psi) \leq c4^{-j}, \quad j = 1, 2, \ldots. \tag{5.55}
\]

If we assume, in place of (5.52), that

\[
(j + 1)\delta_j \leq c2^{-(j/2+2j)}, \quad j = 0, 1, 2, \ldots, \tag{5.56}
\]

then the estimate (5.55) remains valid if \( G \) is a scaling matrix, and \( \psi \) is taken to be one of the functions \( K^Z_j(G, x, e_{q+1}) \), or \( K^Z_{D,j}(G, x, e_{q+1}) \).

**Remark.** We remark that the conditions (5.52) and (5.49) are satisfied for both \( \phi^G_\rho \) and \( \phi^E_\rho \) (cf. (5.9), (5.11), (5.10), and (5.12)). We give an example of a frame matrix \( G \) that satisfies the condition (5.54). We write

\[
g_{j, \ell} := \begin{cases} 
    d \omega_j^{1/2} \sin \left( \pi \frac{\ell - 2j - 1}{2j+1} \right), & \text{if } 2^j + 1 \leq \ell \leq 2^j + 1, \ j = 1, 2, \ldots, \\
    0 & \text{otherwise}.
\end{cases}
\]

It is then easy to verify that \( G \) is a frame matrix, and the condition (5.54) is satisfied.

### 6 Perturbation of frames in a Hilbert space

With the exception of Proposition 5.1, Proposition 5.4, and Theorem 5.4, all the remaining new statements in Section 5 can be proved in the setting of a general Hilbert space. Therefore, in this section, we develop a theory in this more general setting, and give the proofs of the results in Section 5 in the next section.

Let \( H \) be a separable Hilbert space, \( \langle \cdot, \cdot \rangle \) be its inner product, \( \| \cdot \| \) be the corresponding norm, and \( \{e_k\} \) be a complete orthonormal basis for \( H \). Let \( \{N_j\}_{j=0}^\infty \) be an increasing sequence of positive integers, with \( N_j \to \infty \) as \( j \to \infty \). We define \( N_{-1} := -1 \). Motivated by [3, 10, 9, 16, 17, 18, 15, 20, 23, 24], we may define frames on \( H \) as follows. Let \( V^E \) denote the span of \( \{e_k : k = 0, \ldots, N_j\} \), and \( W^E_j := V^E_j \ominus V^E_{j-1} \) be the orthogonal complement of \( V^E_{j-1} \) in \( V^E_j \).

For \( f \in H \), we define the transform

\[
\tau^E_j(G; f) := \sum_{\ell=0}^{N_j} g_{j, \ell}(f, e_{\ell}) e_{\ell} \tag{6.1}
\]
If \( G \) is a scaling matrix, then for any \( f \in V^E_j \), it is easy to verify that

\[
 f = \sum_{\ell=0}^{N_j} g_{j,\ell}^{-1} \langle \tau^E_j(G;f), e_\ell \rangle e_\ell = \tau^E_j(G^{-1}; \tau^E_j(G;f)).
\]  

(6.2)

A similar representation is valid for all \( f \in W^F_j \) if \( G \) is a frame matrix. In \([16, 17, 18, 15]\), the representation (6.2) gives a “recovery formula” in terms of certain integrals, and the localized frames are obtained by using quadrature formulas. In this section, our objective is to obtain a decomposition of \( H \) using perturbations of the orthonormal basis \( \{e_k\} \).

Suppose that one has a set of perturbations \( \{f_{j,k}\} \) such that

\[
 \max_{0 \leq k \leq N_j} \|f_{j,k} - e_k\| \leq \epsilon_j, \quad j = 0, 1, 2, \cdots,
\]  

(6.3)

where \( \|\cdot\| \) is the norm on \( H \) and \( \{\epsilon_j\} \) is a sequence converging to zero as \( j \to \infty \). In this section, we will consider the question of obtaining a decomposition \( V^F_0 \subset V^F_1 \subset \cdots \subset H \), where each \( V^F_j \) is a subspace of \( H \) spanned by certain linear combinations of \( \{f_{j,k} : 0 \leq k \leq N_j, 0 \leq \ell \leq j\} \), and each \( V^F_j \) can be expressed in the form \( V^F_{j-1} + W^F_j \), with \( W^F_j \) orthogonal to \( V^F_{j-1} \). This requirement, in contrast to the usual requirement that \( W^F_j \) be orthogonal to \( V^F_{j-1} \), guarantees that the higher the scale of the wavelet spaces, the higher is the number of vanishing moments.

In order to keep the notation simple, as well as for later reference, we formulate the beginning stages of this construction in a somewhat more abstract setting.

**Lemma 6.1** Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space, \( \{u_k\}_{k=0}^M \) be an orthonormal set in \( V \), \( \{v_k\}_{k=0}^M \subset V \), \( \epsilon > 0 \), and

\[
 \|u_k - v_k\| \leq \epsilon, \quad k = 0, \cdots, M,
\]  

(6.4)

where \( \|\cdot\| \) is the norm induced by \( \langle \cdot, \cdot \rangle \). Let \( \langle \cdot, \cdot \rangle' \) be any inner product on \( V \), such that \( \{u_k\}_{k=0}^M \) is orthonormal with respect to \( \langle \cdot, \cdot \rangle' \), and \( \|\cdot\|' \) be the corresponding norm. Further, let \( \|\cdot\|' \) be any pseudo-norm on \( V \). Finally, we assume that \( 4(M + 1)\epsilon^2 < 1 \). Then there exist

\[
 \{w_k\}_{k=0}^M \subset \text{span}\{v_k\}_{k=0}^M,
\]  

(6.5)

such that

\[
 \langle w_k, u_m \rangle = \delta_{k,m}, \quad k, m = 0, \cdots, M.
\]  

(6.6)

Further,

\[
 \|w_k - u_k\|' \leq 2 \left( \max_{0 \leq k \leq M} \|v_k - u_k\|' + \epsilon \right),
\]  

(6.7)

\[
 \|w_k - u_k\| \leq \frac{1}{1 - \sqrt{M + 1}\epsilon} \max_{0 \leq k \leq M} \|v_k - u_k\| + \frac{\sqrt{M + 1}\epsilon}{1 - \sqrt{M + 1}\epsilon} \max_{0 \leq k \leq M} \|u_k\|. \tag{6.8}
\]

In particular, using \( \|\cdot\| \) in place of \( \|\cdot\|' \) in (6.7), we obtain

\[
 \|w_k - u_k\| \leq 4\epsilon.
\]  

(6.9)

**Proof.** Let, in this proof only, \( A \) denote the matrix defined by \( A_{\ell,m} = \langle v_\ell - u_\ell, u_m \rangle \). In view of the Bessel inequality and (6.4), we may estimate the Frobenius norm \( F(A) \) of \( A \) as follows:

\[
 F(A)^2 := \sum_{\ell, m=0}^M |\langle v_\ell - u_\ell, u_m \rangle|^2 \leq \sum_{\ell=0}^M \|v_\ell - u_\ell\|^2 \leq (M + 1)\epsilon^2.
\]  

(6.10)

Since \( 4(M + 1)\epsilon^2 < 1 \), it follows that \( F(A) < 1/2 \), and in particular, the matrix \( I + A \) is invertible, where \( I \) denotes the \((M + 1) \times (M + 1)\) identity matrix. Let, in this proof only, \( S := A(I + A)^{-1} \). Then \( S(I + A) = A; \) i.e., for \( k, m = 0, \cdots, M + 1 \),

\[
 \sum_{\ell=0}^M S_{k,\ell} (\delta_{\ell, m} + \langle v_\ell - u_\ell, u_m \rangle) = \langle v_k - u_k, u_m \rangle.
\]  

(6.11)
Taking into account the fact that $\langle u_\nu, u_m \rangle = \delta_{\nu,m}$, (6.11) can be rearranged in the form

$$\langle v_k - \sum_{\ell=0}^M S_{k,\ell} v_\ell, u_m \rangle = \delta_{k,m}.$$  

We write $w_k := v_k - \sum_{\ell=0}^M S_{k,\ell} v_\ell$ to arrive at (6.6). Further, for each $k = 0, \cdots, M$, we use the Bessel inequality and elementary facts from matrix analysis again to deduce that

$$\sum_{\ell=0}^M |S_{k,\ell}|^2 = \sum_{\ell=0}^M |(A(I + A)^{-1})_{k,\ell}|^2 \leq \frac{1}{(1 - F(A))^2} \sum_{\ell=0}^M |\langle v_k - u_k, u_\ell \rangle|^2 \leq \frac{1}{(1 - F(A))^2} \| v_k - u_k \|^2.$$  

Using (6.4) and (6.10), we conclude that for each $k = 0, \cdots, M$,

$$\sum_{\ell=0}^M |S_{k,\ell}|^2 \leq \frac{\epsilon^2}{(1 - \sqrt{(M+1)\epsilon})^2}. \quad (6.12)$$

In this proof only, let $\eta := \max_{0 \leq k \leq M} \| v_k - u_k \|'$. Using the fact that $\{u_k\}_{k=0}^M$ is orthonormal also with respect to $\langle \cdot, \cdot \rangle'$, we obtain

$$\| w_k - u_k \|' \leq \| v_k - u_k - \sum_{\ell=0}^M S_{k,\ell}(v_\ell - u_\ell) - \sum_{\ell=0}^M S_{k,\ell} u_\ell \|' \leq \| v_k - u_k \|' + \sum_{\ell=0}^M |S_{k,\ell}| \| v_\ell - u_\ell \|' + \left\{ \sum_{\ell=0}^M |S_{k,\ell}|^2 \right\}^{1/2} \leq \eta + \sqrt{M+1} \eta \frac{\epsilon}{1 - \sqrt{(M+1)\epsilon}} + \frac{\epsilon}{1 - \sqrt{(M+1)\epsilon}} = \frac{\eta + \epsilon}{1 - \sqrt{(M+1)\epsilon}} \leq 2(\eta + \epsilon).$$

This proves (6.7). Next, using the definition of $w_k$, we obtain

$$||| w_k - u_k ||| \leq || v_k - u_k || + \sum_{\ell=0}^M |S_{k,\ell}| | || v_\ell - u_\ell || + \sum_{\ell=0}^M |S_{k,\ell}| \cdot || u_\ell || \leq || v_k - u_k || + \max_{0 \leq \ell \leq M} || v_\ell - u_\ell || \sum_{\ell=0}^M |S_{k,\ell}| + \max_{0 \leq \ell \leq M} || u_\ell || \sum_{\ell=0}^M |S_{k,\ell}|.$$  

In view of (6.12), this leads to (6.8).  

We say that the set $\{v_k\}$ is an $\epsilon$-perturbation of the set $\{e_k\}$, and the set $\{w_k\}$ is the $(\cdot,\cdot)$-biorthogonalization of $\{v_k\}$ with respect to $\{e_k\}$.

The next lemma helps us to analyse the behavior of the inner products under perturbations.

**Lemma 6.2** Let $V$ be a vector space, $\langle \cdot, \cdot \rangle$ be an inner product on $V$, and $|| \cdot ||$ be the corresponding norm. Let $u, v, u', v' \in V$, and

$$|| u - u' || + || v - v' || \leq || u' || + || v' ||. \quad (6.13)$$

Then

$$| \langle u, v \rangle - \langle u', v' \rangle | \leq 3 \left( || u' || + || v' || \right) \left( || u - u' || + || v - v' || \right). \quad (6.14)$$
Proof. We observe that for any \(a, b \in V\), \(\max(\|a\|, \|b\|) \leq A\), we have

\[
\left\|a + b \right\|^2 - \|b\|^2 \leq \|a\|^2 + 2\|a, b\| \leq A\|a\| + 2\|a\|\|b\| \leq 3A\|a\|.
\]  

(6.15)

Therefore,

\[
\left\|u + v\right\|^2 - \|u' + v'\|^2 = \left\|u - u' + v - v' + (u' + v')\right\|^2 - \|u' + v'\|^2 \leq 3\left(\|u'\| + \|v'\|\right)\left(\|u - u'\| + \|v - v'\|\right).
\]

The same estimate holds for \(v\) (respectively \(v'\)) replaced by \(-v\) (respectively \(-v'\)), \(iv\) (respectively \(iv'\)), and \(-iv\) (respectively \(-iv'\)). Therefore, the polar identity:

\[
(a, b) = \frac{1}{4} \left\{\|a + b\|^2 - \|a - b\|^2 + i\|a + ib\|^2 - i\|a - ib\|^2\right\}
\]

leads to (6.14). \(\square\)

We will use Lemma 6.2 in the following form:

**Lemma 6.3** Let \(V\) be a vector space, \((\cdot, \cdot)\) be an inner product on \(V\), and \(|\cdot|\) be the corresponding norm. Suppose \(\{u_k\}\) is an orthonormal subset of \(V\), \(\{v_k\}_{k=0}^{L}\) and \(\{w_k\}_{k=0}^{L}\) satisfy

\[
\max_{0 \leq k \leq J} \|v_k - u_k\| \leq \epsilon, \quad \max_{0 \leq k \leq L} \|w_k - u_k\| \leq \eta.
\]  

(6.16)

Further, let \(0 < \alpha \leq 12\), and \(\sqrt{J + 1}\epsilon + \sqrt{L + 1}\eta \leq \alpha/6\). (6.17)

Then for any complex numbers \(\{a_k\}, \{b_l\}\), we have

\[
\left|\sum_{k=0}^{J} a_k v_k - \sum_{k=0}^{J} b_k w_k - \sum_{k=0}^{\min(J, L)} a_k b_k\right| \leq \alpha \left\{\sum_{k=0}^{J} |a_k|^2\right\}^{1/2} \left\{\sum_{l=0}^{L} |b_l|^2\right\}^{1/2}.
\]  

(6.18)

In particular, if \(\sqrt{J + 1}\epsilon \leq \alpha/12\), then

\[
\left\|\sum_{k=0}^{J} a_k v_k\right\|^2 - \sum_{k=0}^{J} |a_k|^2 \leq \alpha \sum_{k=0}^{J} |a_k|^2.
\]  

(6.19)

Proof. Without loss of generality, we may assume that \(\sum_{k=0}^{J} |a_k|^2 = \sum_{k=0}^{L} |b_k|^2 = 1\). In this proof only, we write \(u = \sum_{k=0}^{J} a_k v_k, u' = \sum_{k=0}^{L} a_k u_k, v = \sum_{k=0}^{L} b_k w_k, \) and \(v' = \sum_{k=0}^{L} b_k u_k\). Then

\[
\|u - u'\|^2 \leq \sum_{k=0}^{J} |a_k|^2 \|v_k - u_k\| \leq \epsilon \sum_{k=0}^{J} |a_k| \leq \sqrt{J + 1}\epsilon.
\]  

(6.20)

Similarly, \(\|v - v'\|^2 \leq \sqrt{L + 1}\eta\). In view of (6.17), the condition (6.13) is satisfied. The estimate (6.18) now follows from (6.14) and (6.20). \(\square\)

We now return to the original problem. In the sequel, we assume that

\[
\epsilon_j \sqrt{N_j + 1} < 1/2, \quad j = 0, 1, \ldots
\]  

(6.21)

For each \(j = 0, 1, \ldots\), we may then use Lemma 6.1 to obtain a biorthogonalization of \(\{f_{j,k}\}_{k=0}^{N_j}\) with respect to \(\{\epsilon_k\}_{k=0}^{N_j}\). We denote it by \(\{p_{j,k}\}_{k=0}^{N_j}\). For any pseudo-norm \(|\cdot|\) on \(H\) and integer \(m \geq 0\), we define

\[
E_m(|\cdot|, f) := |f - \text{Proj}_{V_m}(f)|,
\]  

(6.22)

19
Then there exists a constant $C > 0$ and $\{W_j\}$ such that for any fixed integer $m \geq 0$, $E_m(\|\cdot\|, p_{j,k}) : 0 \leq k \leq N_\ell, \ell = 0, \ldots, m \leq 2E_m(\|\cdot\|)$. \hfill (6.24)

The first objective of this section is to prove the following theorem.

**Theorem 6.1** For integer $j = 0, 1, 2, \ldots$, let

$$V_j^F := \text{span}\left\{ \left. \left( \| p_{\ell,k} \| \right)_{k=N_{\ell-1}+1}^{N_{\ell}} \right| \ell \leq 0 \right\}, \quad W_j^F := \text{span}\left\{ \left. p_{j,k} \right| k=N_{j-1}+1 \right\}.$$ \hfill (6.25)

Then for $j = 1, 2, \ldots, V_j^F \subseteq V_{j+1}^F$, $V_j^F = V_j^{F-1} + W_j^F$, and $W_j^F$ is orthogonal to $V_j^{F-1}$. Let

$$\sum (m + 1)\sqrt{N_{m+1} - N_m}E_m(\|\cdot\|) < \infty,$$ \hfill (6.26)

and

$$\lim_{m \to \infty} (m + 1)\sqrt{N_m + 1}E_m(\|\cdot\|) = 0.$$ \hfill (6.27)

Then there exists a $J_0$ such that for $j \geq J_0$, $W_j^F \cap V_{j-1}^F = \{0\}$. Moreover, $\bigcup_{j=0}^\infty V_j$ is dense in $H$.

Most of the statements of the theorem are clear, except perhaps for the density statement and the assertion that $W_j^F \cap V_{j-1}^F = \{0\}$. A straightforward approach to prove the density statement is to obtain an approximation to an arbitrary $f$ from $V_j^F$ for a sufficiently large value of $j$, and replace each of the $e_k$'s occurring in this approximation by the corresponding $p_{\ell,k}$'s. However, for small values of $k$, the error in this replacement may be too large. Therefore, we need to find another basis for $V_j^F$ in order that such an argument will work.

We now define a set $\{\hat{p}_{j,k}\}$ recursively as follows.

$$\hat{p}_{0,k} := p_{0,k}, \quad k = 0, \ldots, N_0,$$ \hfill (6.28)

$$\hat{p}_{j+1,k} := \begin{cases} \hat{p}_{j,k} - \sum_{\ell=N_{j+1}}^{N_j} (\hat{p}_{j,\ell}, e_{\ell}) p_{j+1,\ell}, & \text{if } 0 \leq k \leq N_j, j \geq 0, \\ p_{j+1,k}, & \text{if } N_j + 1 \leq k \leq N_{j+1}, j \geq 0. \end{cases} \hfill (6.29)$$

In the following proposition, we summarize certain properties of $\{\hat{p}_{j,k}\}$.

**Proposition 6.1** For each integer $j \geq 0$, $\{\hat{p}_{j,k}\}$ is a basis for $V_j^F$, and $\langle \hat{p}_{j,k}, e_{\ell} \rangle = \delta_{k,\ell}, k, \ell = 0, \ldots, N_j$. Let $\|\cdot\|$ be a norm on $H$ which satisfies

$$\|f\| \leq A\|f\|, \quad f \in H,$$ \hfill (6.30)

for a constant $A > 0$ (independent of $f$). If

$$\sum (m + 1)\sqrt{N_{m+1} - N_m}E_m(\|\cdot\|) < \infty,$$ \hfill (6.31)

then there exists a constant $C > 0$ such that for $m = 0, 1, \ldots, J \geq m$,

$$\max_{1 \leq k \leq N_m} E_{f} (\|\cdot\|, \hat{p}_{m,k}) \leq C(m+1) \max\{ E_{f}(p_{\ell,k}) : 1 \leq k \leq N_\ell, \ell = 0, 1, \ldots, m \},$$ \hfill (6.32)

and

$$\max_{1 \leq k \leq N_m} \|\hat{p}_{m,k} - e_k\| \leq C(m+1)E_m(\|\cdot\|).$$ \hfill (6.33)

The proof of this proposition involves the solution of a recursive inequality. We summarize this fact in the following lemma.
Lemma 6.4 Let \( \{a_j\}, \{b_j\} \) be sequences of positive real numbers satisfying \( a_0 = b_0 \) and

\[
a_m \leq b_m (1 + \sum_{j=0}^{m-1} a_j), \quad m = 1, 2, \ldots.
\]  

(6.34)

Then for \( m = 1, 2, \ldots \),

\[
a_m \leq b_m \prod_{j=0}^{m-1} (1 + b_j).
\]  

(6.35)

Proof. We prove this lemma by induction. For \( m = 0 \), (6.35) just states the fact that \( a_0 \leq b_0 \). Suppose that (6.35) has been proved for \( 0 \leq j \leq m \) for some value of \( m \). Let for integer \( k \), \( 0 \leq k \leq m \),

\[
T_k := 1 + \sum_{\nu=k}^{m} b_{\nu} \prod_{k=0}^{\nu-1} (1 + b_{k})
\]

where we recall that an empty sum is understood to be equal to 0, and an empty product is understood to be equal to 1. Then

\[
T_k = 1 + b_k + \sum_{\nu=k+1}^{m} b_{\nu} \prod_{k=0}^{\nu-1} (1 + b_{k})
\]

\[= 1 + b_k + (1 + b_k) \sum_{\nu=k+1}^{m} b_{\nu} \prod_{k=0}^{\nu-1} (1 + b_{k})
\]

\[= (1 + b_k) \left\{ 1 + \sum_{\nu=k+1}^{m} b_{\nu} \prod_{k=0}^{\nu-1} (1 + b_{k}) \right\}
\]

\[= (1 + b_k) T_{k+1}.
\]

Consequently,

\[
1 + \sum_{\nu=0}^{m} b_{\nu} \prod_{k=0}^{\nu-1} (1 + b_{k}) = T_0 = \prod_{\nu=0}^{m-1} (1 + b_{\nu}) T_m = \prod_{\nu=0}^{m} (1 + b_{\nu}).
\]

Now, using this fact, the estimate (6.34) with \( m + 1 \) in place of \( m \), and the assumed estimate (6.35) for \( 0 \leq j \leq m \), we get

\[
a_{m+1} \leq b_{m+1} (1 + \sum_{j=0}^{m} a_j)
\]

\[\leq b_{m+1} \left\{ 1 + \sum_{j=0}^{m} b_{j} \prod_{\nu=0}^{j-1} (1 + b_{\nu}) \right\}
\]

\[= b_{m+1} \prod_{j=0}^{m} (1 + b_{j}).
\]

This completes the proof of (6.35).

\[\square\]

Proof of Proposition 6.1. It is elementary induction to see that \( \langle \tilde{p}_{j:k}, e_\ell \rangle = \delta_{k,\ell}, \ell, k = 0, \ldots, N_j \), and hence, that the set \( \{\tilde{p}_{j:k}\}_{k=0}^{N_j} \) is linearly independent. Therefore, we need to show that \( V_{j}^F = \text{span}\{\tilde{p}_{j:k}\}_{k=0}^{N_j} \). This statement is clearly true for \( j = 0 \). Suppose that it is true for some \( j \geq 0 \). Then (6.29) shows that \( \tilde{p}_{j+1:k} \in V_{j+1}^F \) for each \( k = 0, \ldots, N_{j+1} \), and hence, \( \text{span}\{\tilde{p}_{j+1:k}\}_{k=0}^{N_{j+1}} \subseteq V_{j+1}^F \). Further, if \( f \in V_{j+1}^F \), then \( f = v + w \), where \( v \in V_j^F \) and \( w \in \text{span}\{p_{j+1:k}\}_{k=N_j}^{N_{j+1}} = \text{span}\{\tilde{p}_{j+1:k}\}_{k=N_j}^{N_{j+1}} \). Again, in view of (6.29), each \( \tilde{p}_{j:k} \), \( 0 \leq k \leq N_j \) is in \( \text{span}\{\tilde{p}_{j+1:k}\}_{k=0}^{N_{j+1}} \). Therefore, by induction hypothesis,
$v \in \text{span}(\tilde{p}_{j+1:k})_{k=0}^{N_j+1}$, and hence, $f \in \text{span}(\tilde{p}_{j+1:k})_{k=0}^{N_j+1}$. Thus, \(\text{span}(\tilde{p}_{j+1:k})_{k=0}^{N_j+1} = V^F_j\). This proves the first statement in the lemma.

During the remainder of this proof, we consider the norm \(\|\cdot\|\) to be fixed, and omit its mention from the notations. Thus, we will write \(E_J(f)\) rather than \(E_J(\|\cdot\|, f)\). In this proof only, let
\[
\mathcal{E}_{J,j} := \max\{E_J(p_{\ell,k}) : 1 \leq k \leq N_\ell, \ell = 0, 1, \cdots, j\},
\]
and
\[
\eta_{J,j} := \max_{0 \leq k \leq N_j} E_J(\tilde{p}_{j:k}).
\]

Now, let \(m \geq 1, J \geq m\) be any integers. If \(0 \leq j \leq m - 1\) is an integer, and \(0 \leq k \leq N_j\), then
\[
\tilde{p}_{j+1,k} = \tilde{p}_{j:k} - \sum_{\ell = N_j+1}^{N_j+1} (\tilde{p}_{j:k}, e_\ell)p_{\ell+1,k}.
\]
Therefore,
\[
E_J(\tilde{p}_{j+1,k}) \leq E_J(\tilde{p}_{j:k}) + \sum_{\ell = N_j+1}^{N_j+1} |(\tilde{p}_{j:k}, e_\ell)| E_J(p_{\ell+1,k}).
\]
Now, since \(0 \leq k \leq N_j\), \(\text{proj}_{V^F_j}(\tilde{p}_{j:k}) = e_k\), and using the Bessel inequality and (6.30), we deduce that
\[
\sum_{\ell = N_j+1}^{N_j+1} |(\tilde{p}_{j:k}, e_\ell)| = \sum_{\ell = N_j+1}^{N_j+1} |(\tilde{p}_{j:k} - e_k, e_\ell)|
\leq \sqrt{N_j+1 - N_j} \left\{ \sum_{\ell = N_j+1}^{N_j+1} |(\tilde{p}_{j:k} - e_k, e_\ell)|^2 \right\}^{1/2}
\leq \sqrt{N_j+1 - N_j} \left\| \tilde{p}_{j:k} - e_k \right\| \leq A \sqrt{N_j+1 - N_j} E_J(\tilde{p}_{j:k})
\leq A \sqrt{N_j+1 - N_j} \eta_{j,j}.
\]
Consequently, (6.38) implies that for \(0 \leq k \leq N_j\),
\[
E_J(\tilde{p}_{j+1,k}) \leq E_J(\tilde{p}_{j:k}) + A E_{J,m} \sqrt{N_j+1 - N_j} \eta_{j,j} \leq \eta_{J,j} + A E_{J,m} \sqrt{N_j+1 - N_j} \eta_{j,j}.
\]
If \(N_j + 1 \leq k \leq N_{j+1}\), then \(\tilde{p}_{j+1,k} = p_{j+1,k}\), and hence, \(E_J(\tilde{p}_{j+1,k}) = E_J(p_{j+1,k}) \leq E_J_{J,m}\). Thus, for \(1 \leq k \leq N_{j+1}\),
\[
E_J(\tilde{p}_{j+1,k}) \leq \eta_{J,j} + E_{J,m}(1 + A \sqrt{N_j+1 - N_j} \eta_{j,j});
\]
i.e.,
\[
\eta_{J,j+1} \leq \eta_{J,j} + E_{J,m}(1 + A \sqrt{N_j+1 - N_j} \eta_{j,j}).
\]
Taking the sum for \(j = 0, \cdots, m - 1\), we obtain
\[
\eta_{J,m} \leq \eta_{J,0} + E_J_{J,m} \left( m + A \sum_{j=0}^{m-1} \sqrt{N_j+1 - N_j} \eta_{j,j} \right).
\]
Since \(\tilde{p}_{0,k} = p_{0,k}\) for \(k = 1, \cdots, N_0\), we see that \(\eta_{J,0} = E_{J,0} \leq E_{J,m}\). Therefore, for every \(J \geq m\), we get
\[
\eta_{J,m} \leq E_{J,m} \left( m + 1 + A \sum_{j=0}^{m-1} \sqrt{N_j+1 - N_j} \eta_{j,j} \right)
\leq c(m + 1) E_J_{J,m} \left( 1 + \sum_{j=0}^{m-1} \sqrt{N_j+1 - N_j} \eta_{j,j} \right),
\]
(6.39)
where \( c > 0 \) is some constant. We apply this estimate with \( J = m \), and write

\[
a_j := \sqrt{N_{j+1} - N_j} \eta_{j,j}, \quad b_m := c(m + 1)\sqrt{N_{m+1} - N_m} \varepsilon_{m,m}
\]

(6.40) to obtain

\[
a_m \leq b_m (1 + \sum_{j=0}^{m-1} a_j), \quad m = 1, 2, \ldots.
\]

In view of Lemma 6.4, this leads to

\[
a_m \leq b_m \prod_{j=0}^{m-1} (1 + b_j).
\]

(6.41) In view of (6.24), the condition (6.31) implies that

\[
\text{In view of (6.24), the condition (6.31) implies that}
\]

\[
a_m \leq b_m \prod_{j=0}^{m-1} (1 + b_j).
\]

(6.41) Therefore, the infinite product

\[
\prod_{j=0}^{\infty} (1 + b_j)
\]

converges to a positive number. Hence, (6.41) implies

\[
\eta_{m,m} \leq c(m + 1)\varepsilon_{m,m},
\]

which is (6.33). Substituting this estimate back into (6.39) and recalling (6.31) again, we get \( \eta_{j,m} \leq c(m + 1)\varepsilon_{j,m} \), which leads to (6.32).

**Proof of Theorem 6.1** The only statements that are perhaps not obvious are the assertion that \( W_j^F \cap V_{j-1}^F = \{0\} \), and the density statement. In view of (6.27), we may choose an integer \( J \) so that

\[
C(m + 1)\sqrt{N_m + 1} \varepsilon_m(\| \cdot \|) < 1/36, \quad m \geq J,
\]

(6.42) where \( C \) is the constant appearing in (6.33). Now, let \( j \geq J, f \in W_j \cap V_{j-1}, \)

\[
f := \sum_{k=N_{j-1}+1}^{N_j} d_k p_{j,k} = \sum_{k=N_{j-1}+1}^{N_j} d_k \tilde{p}_{j,k} = \sum_{k=0}^{N_{j-1}} h_k \tilde{p}_{j-1,k},
\]

(6.43) and, without loss of generality,

\[
\sum_{k=N_{j-1}+1}^{N_j} |d_k|^2 = \max \left\{ \sum_{k=N_{j-1}+1}^{N_j} |d_k|^2, \sum_{k=0}^{N_{j-1}} |h_k|^2 \right\} = 1.
\]

(6.44) Now we use Lemma 6.3 with \( H \) in place of \( V \), \{\( e_k \)\} in place of \{\( u_k \)\}, \{\( \tilde{p}_{j,k} \)\} in place of \{\( v_k \)\} and \{\( \tilde{p}_{j-1,k} \)\} in place of \{\( w_k \)\}. The estimates (6.33) and (6.42) leads to (6.17) with \( \alpha = 1/3 \). Consequently, (6.18) implies that

\[
\|f\|^2 = \langle f, f \rangle = \left( \sum_{k=0}^{N_j} h_k \tilde{p}_{j-1,k} \right) \left( \sum_{k=N_{j-1}+1}^{N_j} d_k \tilde{p}_{j,k} \right) \leq 1/3,
\]

while (6.19) implies that

\[
\|f\|^2 - 1 = \left| \sum_{k=N_{j-1}+1}^{N_j} d_k \tilde{p}_{j,k} \right|^2 - \left| \sum_{k=N_{j-1}+1}^{N_j} d_k \right|^2 \leq 1/3.
\]

This contradiction proves that \( W_j^F \cap V_{j-1}^F = \{0\} \).

Next, let \( f \in H, \|f\| = 1, \) and \( 0 < \epsilon < 3 \). In view of (6.27) and the fact that \{\( e_k \)\} is a complete orthonormal system, we may find an integer \( J \) such that

\[
C(j + 1)\sqrt{N_j + 1} \varepsilon_j(\| \cdot \|) < \epsilon/3.
\]

(6.45)
and there exists a \( v \in V_j^F \) such that \( \| f - v \| < \epsilon / 3 \). Let \( v = \sum_{k=0}^{N_j} h_k e_k \), and \( w = \sum_{k=0}^{N_j} h_k \tilde{p}_{j;k} \in V_j^F \). Then using (6.33) and (6.45) we deduce that \( \| v - w \| \leq 2\epsilon / 3 \), and hence, \( \| f - w \| < \epsilon \). This proves the density statement.

Next, if \( G \) is a frame matrix, we define the (continuous) frame transform by

\[
\tau_j^F(G; f) := \sum_{k=0}^{N_j} \langle f, g_{j;k} e_k \rangle \tilde{p}_{j;k} = \sum_{k=N_{j-1}+1}^{N_j} \langle f, g_{j;k} e_k \rangle \tilde{p}_{j;k}, \quad j = 0, 1, \cdots.
\] (6.46)

The following theorem summarizes some of the basic properties of the frame transform.

**Theorem 6.2** Let \( G \) be a frame matrix, \( j \geq 0 \) be an integer, and \( 0 < \alpha \leq 12 \). We assume further that

\[
\epsilon_j \sqrt{N_j + 1} \leq \alpha / 48.
\] (6.47)

If \( f \in W_j \), then

\[
f = \sum_{k=N_{j-1}+1}^{N_j} \langle \tau_j^F(G; f), g_{j;k} e_k \rangle \tilde{p}_{j;k} = \tau_j^F(G^{[-1]}; \tau_j^F(G; f)).
\] (6.48)

Further, with

\[
B_{j,2}(G) := \max_{0 \leq \ell \leq N_j} |g_{j,\ell}|,
\] (6.49)

we have for \( j = 0, 1, \cdots, f \in W_j \),

\[
\left\{ \sqrt{1 + \alpha B_{j,2}(G^{[-1]})} \right\}^{-1} \| f \| \leq \| \tau_j(G; f) \| \leq \sqrt{1 + \alpha B_{j,2}(G)} \| f \|.
\] (6.50)

**Proof.** The proof relies upon the following observation. If \( f \in W_j, f = \sum_{k=N_{j-1}+1}^{N_j} a_k \tilde{p}_{j;k} \), then the relations \( \langle p_{j;k}, e_\ell \rangle = \delta_{k,\ell}, \ k, \ell = N_{j-1} + 1, \cdots, N_j \) imply that \( a_k = \langle f, e_k \rangle, k = N_{j-1} + 1, \cdots, N_j \). Thus, we have

\[
f = \sum_{k=N_{j-1}+1}^{N_j} \langle f, e_k \rangle \tilde{p}_{j;k}, \quad f \in W_j, \quad j = 0, 1, 2, \cdots.
\] (6.51)

The representation (6.48) is now obvious. Next, we use Lemma 6.3 with \( \{p_{j;k}\} \) in place of \( \{v_k\} \) and \( \{e_k\} \) in place of \( \{u_k\} \). The estimates (6.47) and (6.9) imply that (6.19) is valid. Using Bessel inequality, and (6.19), we obtain

\[
\| \tau_j(G; f) \|^2 \leq (1 + \alpha) \sum_{k=N_{j-1}+1}^{N_j} |g_{j;k} \langle f, e_k \rangle|^2
\]

\[
\leq (1 + \alpha) B_{j,2}(G)^2 \sum_{k=N_{j-1}+1}^{N_j} |\langle f, e_k \rangle|^2 \leq (1 + \alpha) B_{j,2}(G)^2 \| f \|^2.
\]

This proves the second inequality in (6.50). The first estimate in (6.50) follows from the fact that

\[
f = \tau_j^F(G^{[-1]}; \tau_j^F(G; f)).
\]

In applications, the inner product is typically defined by an integral. Thus, formula (6.48) is an integral inversion formula. In order to obtain a frame representation, one needs a quadrature formula. In our abstract setting, we therefore assume that for every \( j = 0, 1, \cdots \), there is an inner product \( \langle \cdot, \cdot \rangle_j \) such that

\[
\langle e_k, e_\ell \rangle_j = \delta_{k,\ell}, \quad k, \ell = 0, \cdots, N_j.
\] (6.52)
Let $\| \cdot \|_j$ be the norm induced by $\langle \cdot, \cdot \rangle_j$. We assume further that
\[
\max(||p_{j,k} - e_k||, ||p_{j,k} - e_k||_j) \leq \eta_j \leq \frac{1}{2\sqrt{N_j + 1}}, \quad k = 0, \ldots, N_j, \quad j = 0, 1, \ldots. \tag{6.53}
\]

Since $\{e_k\}_{k=N_{j-1}+1}^{N_j}$ is orthonormal with respect to $\langle \cdot, \cdot \rangle_j$, we may obtain a $\langle \cdot, \cdot \rangle_j$-biorthogonalization, $\{q_{j,k}\}_{k=N_{j-1}+1}^{N_j}$, of $\{p_{j,k}\}_{k=N_{j-1}+1}^{N_j}$ with respect to $\{e_k\}_{k=N_{j-1}+1}^{N_j}$. Finally, if $G$ is a frame matrix, we define the discrete frame operator by
\[
\tau_{D,j}^F(G; f) := \sum_{k=0}^{N_j} \langle f, g_{j,k} e_k \rangle q_{j,k} = \sum_{k=N_{j-1}+1}^{N_j} \langle f, g_{j,k} e_k \rangle q_{j,k}, \quad j = 0, 1, \ldots. \tag{6.54}
\]

Analogous to Theorem 6.2, we have the following theorem for the discrete frame operators.

**Theorem 6.3** Let $G$ be a frame matrix, $j \geq 0$ be an integer, and $0 < \alpha < 1$. We assume that $\sqrt{N_j + 1} \eta_j \leq \alpha/48$. If $f \in W_j$, then
\[
f = \sum_{k=N_{j-1}+1}^{N_j} \langle f, e_k \rangle q_{j,k}; \quad f \in W_j, \quad j = 0, 1, \ldots.
\]

The equations (6.55) are now clear. In view of (6.7) and (6.9), we see that
\[
\eta_j \leq N_j, \quad k = N_{j-1} + 1, \ldots, N_j. \tag{6.57}
\]

Therefore, in view of our assumption that $\sqrt{N_j + 1} \eta_j \leq \alpha/48$, we can apply Lemma 6.3 with $W_j$ in place of $V$, $\{e_k\}$ in place of $\{u_k\}$, $\{q_{j,k}\}$ in place of $\{v_k\}$ and $\{w_k\}$ and either of the norms $\| \cdot \|_1$, $\| \cdot \|_2$ (and corresponding inner products). Hence, for $\ell \in W_j$, and complex numbers $\{a_k\}$,
\[
\left| \sum_{k=N_{j-1}+1}^{N_j} a_k q_{j,k} \right|^2 \leq \sum_{k=N_{j-1}+1}^{N_j} |a_k|^2 \leq \alpha \sum_{k=N_{j-1}+1}^{N_j} |a_k|^2. \tag{6.58}
\]

In particular,
\[
(1 - \alpha) \sum_{k=N_{j-1}+1}^{N_j} |a_k|^2 \leq \left| \sum_{k=N_{j-1}+1}^{N_j} a_k q_{j,k} \right|^2 \leq (1 + \alpha) \sum_{k=N_{j-1}+1}^{N_j} |a_k|^2. \tag{6.59}
\]

Hence, for $\ell, m = 1, 2$,
\[
\left\| \tau_{D,j}^F(G; f) \right\|^2_\ell \leq (1 + \alpha) \sum_{k=N_{j-1}+1}^{N_j} |\langle f, g_{j,k} e_k \rangle|^2 \leq B_{j,2}^2 (1 + \alpha) \sum_{k=N_{j-1}+1}^{N_j} |\langle f, e_k \rangle|^2 \leq B_{j,2}^2 \frac{1 + \alpha}{1 - \alpha} \|f\|_m.
\]
This proves the second inequality in (6.56). The first inequality follows by using this part, with the role of \( \ell \) and \( m \) reversed, and (6.55).

Finally, we state the analogue of the above theorems in the case of scaling functions. We assume that (6.30) is valid with \( \| \cdot \| \) as well as \( \| \cdot \| \) on its left hand side, and \( C(j+1)\sqrt{N_j + 1}\xi_j(\| \cdot \| \| \cdot \| < 1/2. \) Then we may obtain a \( \langle \cdot, \cdot \rangle_f \)-biorthogonalization, \( \{ \tilde{q}_{j,k} \}_{k=0}^{N_j} \) of \( \{ \tilde{p}_{j,k} \}_{k=0}^{N_j} \) with respect to \( \{ e_k \}_{k=0}^{N_j}. \) Let \( G \) be a scaling matrix. We define the continuous scaling transform by

\[
\tau^G_f(G; f) := \sum_{k=0}^{N_j} \langle f, g_{j,k}e_k \rangle \tilde{p}_{j,k}, \quad j = 0, 1, \ldots. \tag{6.60}
\]

and similarly, the discrete scaling transform by

\[
\tau^D_f(G; f) := \sum_{k=0}^{N_j} \langle f, g_{j,k}e_k \rangle \tilde{q}_{j,k}, \quad j = 0, 1, \ldots. \tag{6.61}
\]

The following theorem is the analogue of Theorems 6.2 and 6.3.

**Theorem 6.4** Let \( G \) be a scaling matrix, \( 0 < \alpha < 1, \)

\[
\max(\| f \|, \| f \|) \leq A\| f \|, \quad f \in V^F_j, \ j = 0, 1, \ldots, \tag{6.62}
\]

where \( A \) is a constant independent of \( f \) and \( j. \) We assume further that

\[
AC(j+1)\sqrt{N_j + 1}\xi_j(\| \cdot \| \| \cdot \| \leq \alpha/48,
\]

where \( C \) is the constant appearing in (6.33). If \( f \in V_j, \) then

\[
f = \sum_{k=0}^{N_j} \langle \tau^F_f(G; f), g_{j,k}e_k \rangle \tilde{p}_{j,k} = \tau^F_f(G^{-1}; \tau^F_f(G; f))
\]

\[
= \sum_{k=N_j-1}^{N_j} \langle \tau^D_f(G; f), g_{j,k}e_k \rangle \tilde{q}_{j,k} = \tau^D_f(G^{-1}; \tau^D_f(G; f)) \tag{6.63}
\]

We have

\[
\left\{ 1 + \alpha B_{j,2}(G^{-1}) \right\}^{-1} \| f \| \leq \| \tau_j(G; f) \| \leq \sqrt{1 + \alpha B_{j,2}(G)} \| f \|,
\]

and, with the notation as in Theorem 6.3,

\[
\left\{ B_{j,2}(G^{-1}) \sqrt{(1+\alpha)/(1-\alpha)} \right\}^{-1} \| f \| \leq \| \tau_{D,j}(G; f) \| \leq \sqrt{1 + \alpha B_{j,2}(G)} \frac{1}{\sqrt{1-\alpha}} \| f \|.
\]

The proof of this theorem is verbatim the same as that of Theorems 6.2 and 6.3. Therefore, we omit this proof.

\section{Proof of the results in Section 5}

**Proof of Proposition 5.1.** Using the Funk-Hecke formula (4.13), we see that

\[
\int_{S^q} p_{j,k}(x) Y_{m,s}(x) d\mu_q(x) = \frac{d^q}{\omega_q\phi(q)} \int_{S^q} \phi(x \cdot y) Y_{m,s}(x) Y_{\ell,k}(y) d\mu_q(x) d\nu_{C_j}(y)
\]

\[
= \frac{d^q}{\omega_q\phi(q)} \frac{\omega_q\hat{\phi}(m)}{d_m^n} \int_{S^q} Y_{m,s}(y) Y_{\ell,k}(y) d\nu_{C_j}(y)
\]

\[
= \frac{d^q}{\omega_q\phi(q)} \frac{\omega_q\hat{\phi}(m)}{d_m^n} \int_{S^q} Y_{m,s}(y) Y_{\ell,k}(y) d\mu_q(y)
\]

\[
= \delta(\ell,k), (m,s). \]
This proves (5.14).

In view of (4.13), we have

\[
Y_{\ell,k}(x) = \frac{d^q}{\omega_q \phi(\ell)} \int_{S^n} \phi(x \cdot \xi) Y_{\ell,k}(\xi) d\mu_q(\xi).
\]

Writing, in this proof only, \( \sigma = \mu_q - \nu_{C_j} \), we see from (4.15) that for any univariate polynomial \( P \) of degree not exceeding \( M_j \):

\[
\int_{S^n} P(x \cdot \xi) Y_{\ell,k}(\xi) d\sigma(\xi) = 0, \quad \ell = 0, \ldots, 2^j, \ k = 1, \ldots, d^q_n.
\]

Therefore, if \( P \) is the polynomial of best approximation in the supremum norm on \([-1, 1]\) to \( \phi \) from the space of all univariate polynomials of degree not exceeding \( M_j \), we have

\[
|Y_{\ell,k}(x) - p_{j;\ell,k}(x)| = \left| \frac{d^q}{\omega_q \phi(\ell)} \int_{S^n} \phi(x \cdot \xi) Y_{\ell,k}(\xi) d\sigma(\xi) \right|
\]

\[
= \left| \frac{d^q}{\omega_q \phi(\ell)} \int_{S^n} \left( \phi(x \cdot \xi) - P(x \cdot \xi) \right) Y_{\ell,k}(\xi) d\sigma(\xi) \right|
\]

\[
\leq \frac{d^q}{\omega_q \phi(\ell)} \| \phi - P \|_{w_q, \infty} \int_{S^n} |Y_{\ell,k}(\xi)| (d\mu_q(\xi) + d\nu_{C_j}(\xi))
\]

\[
= \frac{d^q}{\omega_q \phi(\ell)} E_{M_j}(\phi) \int_{S^n} |Y_{\ell,k}(\xi)| (d\mu_q(\xi) + d\nu_{C_j}(\xi)). \tag{7.1}
\]

Since

\[
\int_{S^n} |Y_{\ell,k}(\xi)|^2 d\mu_q(\xi) = \int_{S^n} |Y_{\ell,k}(\xi)|^2 d\nu_{C_j}(\xi) = 1,
\]

Cauchy-Schwarz inequality gives

\[
\int_{S^n} |Y_{\ell,k}(\xi)| d\mu_q(\xi) \leq \sqrt{\omega_q}, \quad \int_{S^n} |Y_{\ell,k}(\xi)| d\nu_{C_j}(\xi) \leq \sqrt{\omega_q}. \tag{7.2}
\]

Therefore, (7.1) implies (5.15).

Next, in this proof only, let

\[
s_n(x) := \sum_{\ell=0}^{2^n} \phi(\ell) P_\ell(q + 1; x).
\]

Then it is easy to verify that

\[
\proj_{V_n}(p_{j;\ell,k}, x) = \frac{d^q}{\omega_q \phi(\ell)} \int_{S^n} s_n(x \cdot \xi) Y_{\ell,k}(\xi) d\nu_{C_j}(\xi) \tag{7.3}
\]

If \( 2^n \leq M_j \) then (4.15) and (4.13) imply that \( \proj_{V_n}(p_{j;\ell,k}, x) = Y_{\ell,k}(x) \). Hence, by (5.15),

\[
\| p_{j;\ell,k} - \proj_{V_n}(p_{j;\ell,k}) \|_{S^n, \infty} \leq \epsilon_j \leq \frac{2}{\sqrt{\omega_q}} \frac{E_{M_j}(\phi)}{m_{2^j}}. \tag{7.4}
\]

If \( 2^n > M_j \), then we use (7.2) and Cauchy-Schwarz inequality to conclude that

\[
\| p_{j;\ell,k} - \proj_{V_n}(p_{j;\ell,k}) \|_{S^n, \infty} \leq \frac{1}{\sqrt{\omega_q}} \frac{E_{2^n}(\phi)}{m_{2^j}}.
\]

Along with (7.4), this leads to (5.17). \( \square \)
Next, we use the results in Section 6 with the following choices. We take $H$ to be the space $L^2(\mathbb{S}^q)$, \( \{e_k\} \) to be an enumeration of the \( \{Y_{\ell,k}\} \)'s where the lower degree polynomials come before the higher degree ones. Nevertheless, we find it convenient to keep the double indexed notation in this section. In particular, if $e_\nu$ corresponds to $Y_{\ell,k}$, we will take its perturbation $f_{j,\nu}$ to be $p_{j,\ell,k}$. Proposition 5.1 then shows that \( \{p_{j,\ell,k}\} \) is the $\epsilon_j$-biorthogonalization of itself with respect to \( \{Y_{\ell,k}\} \). The spaces $V_j^Z$ and $W_j^Z$ are then the spaces denoted in Section 6 by $V_j^F$ and $W_j^F$ respectively. In view of (5.16), the quantity $\eta_j$ in (6.53) may be taken to be $\sqrt{\varepsilon_n}$. Next, we take \(|\cdot|\) in (6.22) to be $\|\cdot\|_{2\times2}$. Then (6.30) and (6.62) are satisfied with $\|\cdot\|_{2\times2}$ in place of $\|\cdot\|$ and $\|\cdot\|_{2\times2}$ in place of $\|\cdot\|_2$. Further, (5.17) shows that the quantity denoted by $\varepsilon_n(\|\cdot\|)$ can be bounded from above by $\delta_n$. Therefore, the conditions (5.4), (5.7), (5.8) imply that the hypothesis of Theorem 6.1, Theorem 6.2, and Theorem 6.3 are satisfied, where we may take $\alpha = 1/2$. Theorem 6.1 now implies Theorem 5.1. Proposition 5.2 follows from (5.17), Proposition 6.1, and Lemma 6.1. (We take $\tilde{q}_{j,t,k}$ to be the $\gamma_j$-biorthogonalization of $\tilde{p}_{j,t,k}$ with respect to \( \{Y_{\ell,k}\} \).) Proposition 5.3 follows by taking \( \{q_{j,t,k}\} \) to be the $\eta_j$-biorthogonalization of \( \{p_{j,t,k}\} \) with respect to \( \{Y_{\ell,k}\} \) using the discrete norm on $W_j$ as in Section 6 (cf. (6.57)). Theorem 5.2 follows from Theorem 6.2, and Theorem 6.3. Then 5.3 is a reformulation of Theorem 6.3.

**Proof of Proposition 5.4.** Let $\psi = \sum_{\ell=m}^{\infty} \sum_{k=1}^{d_{\ell}} a_{\ell,k} Y_{\ell,k} \in L^2(\mathbb{S}^q)$. In this proof only, let $b_m := \sum_{\ell=m}^{\infty} \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2$. Then a summation by parts shows that

$$
\sum_{\ell=m}^{\infty} (\ell - 1) \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 = \sum_{\ell=m}^{\infty} (\ell - 1) (b_\ell - b_{\ell+1}) = m(m - 1)b_m + 2 \sum_{\ell=m}^{\infty} \ell b_{\ell+1}.
$$

Therefore, (5.47) shows that for any $j \geq 0$, \( \|\psi\|_{2\times2}^2 \varphi_F(\psi) \)

\begin{align*}
\|\psi\|_{2\times2}^2 \varphi_F(\psi) &= \sum_{\ell=0}^{\infty} \ell (\ell + q - 1) \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 \\
&= \sum_{\ell=0}^{2j-1} \ell (\ell + q - 1) \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 + \sum_{\ell=2}^{\infty} \ell (\ell + q - 1) \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 \\
&\leq c \{4^j \|\psi\|_{2\times2}^2 + \sum_{\ell=2}^{\infty} \ell (\ell - 1) \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 \} \\
&\leq c \{4^j \|\psi\|_{2\times2}^2 + \sum_{\ell=2}^{\infty} \ell b_{\ell+1} \} \\
&= c \{4^j \|\psi\|_{2\times2}^2 + \sum_{\ell=2}^{\infty} \ell b_{\ell+1} \} \\
&\leq c \{4^j \|\psi\|_{2\times2}^2 + \sum_{\nu=j}^{\infty} 4^{\nu}b_{2^{\nu+1}} \}.
\end{align*}

Thus,

$$
\varphi_F(\psi) \leq c \{4^j + \|\psi\|_{2\times2}^2 \sum_{\nu=j}^{\infty} 4^{\nu} \|\psi - \text{proj}_{2\times2}(\psi)\|_{2\times2}^2 \}.
$$

(7.5)

Now, let $G$ be a scaling or frame matrix, and $K_j^H(G; x, e_{q+1}) = \sum_{\ell=0}^{2j} \sum_{k=1}^{d_{\ell}} a_{\ell,k} Y_{\ell,k}(x)$. Then the kernels $K_j^Z(G; x, e_{q+1})$ and $K_j^{Z^2}(G; x, e_{q+1})$ have the form $\sum_{\ell=0}^{2j} \sum_{k=1}^{d_{\ell}} a_{\ell,k} \psi_{\ell,k}(x)$, where $\psi_{\ell,k}$ is one of the functions $p_{j,t,k}$, $q_{j,t,k}$, $\tilde{p}_{j,t,k}$, or $\tilde{q}_{j,t,k}$ as appropriate. In any case our assumptions imply that (cf. (5.15), (5.31), (5.21), (5.25), (6.19))

$$
\|\psi\|_{2\times2}^2 \geq (1/2) \sum_{\ell=0}^{2j} \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2.
$$

(7.6)
Further, using (5.17), (5.23), (6.8), we get

\[
\|\psi - \text{proj}_{\mathcal{H}_{\nu}}(\psi)\|_{\mathcal{S}_{\nu},2}^2 \leq \left(\sum_{\ell=0}^{2^j} \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 \right) \left(\sum_{\ell=0}^{2^j} \sum_{k=1}^{d_{\ell}} |\psi_{\ell,k}|^2 \right)
\]

\[
\leq \left\{ \sum_{\ell=0}^{2^j} \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2 \right\} \left\{ \sum_{\ell=0}^{2^j} \sum_{k=1}^{d_{\ell}} |\psi_{\ell,k}|^2 \right\}
\]

\[
\leq c^2 2^j (\nu + 1)^2 \delta_\nu^2 \sum_{\ell=0}^{2^j} \sum_{k=1}^{d_{\ell}} |a_{\ell,k}|^2
\]

\[
\leq c \|\psi\|_{\mathcal{S}_{\nu},2}^2 2^j (\nu + 1)^2 \delta_\nu^2.
\]

Therefore, (5.49) implies that

\[
\|\psi\|_{\mathcal{S}_{\nu},2}^2 \sum_{\nu=2}^{\infty} 4^\nu \|\psi - \text{proj}_{\mathcal{H}_{\nu}}(\psi)\|_{\mathcal{S}_{\nu},2}^2 \leq c.
\]

Along with (7.5), this implies (5.48). \(\square\)

In order to prove Theorem 5.4, we prove first the following lemma.

**Lemma 7.1** Let \(\psi_1, \psi_2 \in L^2(S^2)\), \(\|\psi_1\|_{\mathcal{S}_{\nu},2} = 1\), and \(\|\psi_1 - \psi_2\|_{\mathcal{S}_{\nu},2} \leq (1/16) |T(\psi_1)|^2\), then

\[
|\text{var}_S(\psi_1) - \text{var}_S(\psi_2)|_{\mathcal{S}_{\nu},2} \leq \frac{16}{|T(\psi_1)|^4} \|\psi_1 - \psi_2\|_{\mathcal{S}_{\nu},2}.
\]

**Proof.** In this proof only, we write \(\| \cdot \|\) in place of \(\| \cdot \|_{\mathcal{S}_{\nu},2}\), and let \(\|\psi_2\|_{\mathcal{S}_{\nu},2} = A_2\), and \(f_2 = \psi_2/A_2\). Then

\[
\|\psi_1 - f_2\| = \frac{\|A_2 \psi_1 - \psi_2\|}{A_2} \leq A_2 \frac{\|\psi_1 - \psi_2\|}{A_2} + \frac{|A_2 - 1| A_2}{A_2}
\]

Since \(|A_2 - 1| \leq \|\psi_1 - \psi_2\|\), it follows that \(\|\psi_1 - f_2\| \leq 2 \|\psi_1 - \psi_2\|\). Hence,

\[
\int_{S^2} \|\psi_1|^2 - |f_2|^2\| d\mu_q
\]

\[
\leq \int_{S^2} (\|\psi_1 - f_2\| \cdot \|\psi_1 + f_2\|) d\mu_q
\]

\[
\leq \|\psi_1 - f_2\| \cdot \|\psi_1 + f_2\| \leq 2 \|\psi_1 - f_2\| \leq 4 \|\psi_1 - \psi_2\|.
\]

It follows from the definitions that

\[
|T(\psi_1) - T(\psi_2)| = |T(\psi_1) - T(f_2)| \leq \int_{S^2} \|\psi_1|^2 - |f_2|^2\| d\mu_q \leq 4 \|\psi_1 - \psi_2\|,
\]

and hence,

\[
|T(\psi_1)|^2 - |T(\psi_2)|^2 \leq 8 \|\psi_1 - \psi_2\|.
\]

So, \(|T(\psi_2)|^2 \geq (1/2)|T(\psi_1)|^2\). Consequently,

\[
|\text{var}_S(\psi_1) - \text{var}_S(\psi_2)| = |\text{var}_S(\psi_1) - \text{var}_S(f_2)|
\]

\[
= \frac{1}{|T(\psi_1)|^2} - \frac{1}{|T(f_2)|^2}
\]

\[
\leq \frac{16}{|T(\psi_1)|^4} \|\psi_1 - \psi_2\|.
\]

This completes the proof.
Proof of Theorem 5.4. We consider the case when $G$ is a frame matrix; the case when $G$ is a scaling matrix is similar. Let $\tilde{G}$ be the matrix defined by $(\tilde{G})_{j,\ell} = \tilde{g}_{j,\ell}$, $\psi_1 = K_{\tilde{G}}^{\mathbb{Z}}(\tilde{G}, x, e_{q+1})$, and $\tilde{\psi}$ be one of the functions $K_{\tilde{G}}^\mathbb{Z}(\tilde{G}, x, e_{q+1})$ or $K_{\tilde{G}}^\mathbb{D}(\tilde{G}, x, e_{q+1})$. Then $\|\psi_1\|_{S^q,2}^2 = 1$. Moreover, using Propositions 5.1 and 5.3, an application of Cauchy-Swarz inequality leads to $\|\psi_1 - \tilde{\psi}\|_{S^q,2} \leq c4^{-j}$. Therefore, (5.55) follows from [15, Theorem 4.1] and Lemma 7.1. For the case of scaling kernel functions, we use Proposition 5.2 to obtain $\|\psi_1 - \tilde{\psi}\|_{S^q,2} \leq c4^{-j}$.  

References


