Near Optimal Reconstruction of Spherical Harmonic Expansions

Amir Zandieh
MPI-Informatics
azandieh@mpi-inf.mpg.de

Insu Han
Yale University
insu.han@yale.edu

Haim Avron
Tel Aviv University
haimav@tauex.tau.ac.il

Abstract

We propose an algorithm for robust recovery of the spherical harmonic expansion of functions defined on the d-dimensional unit sphere $\mathbb{S}^{d-1}$ using a near-optimal number of function evaluations. We show that for any $f \in L^2(\mathbb{S}^{d-1})$, the number of evaluations of f needed to recover its degree-q spherical harmonic expansion equals the dimension of the space of spherical harmonics of degree at most q up to a logarithmic factor. Moreover, we develop a simple yet efficient algorithm to recover degree-q expansion of f by only evaluating the function on uniformly sampled points on $\mathbb{S}^{d-1}$. Our algorithm is based on the connections between spherical harmonics and Gegenbauer polynomials and leverage score sampling methods. Unlike the prior results on fast spherical harmonic transform, our proposed algorithm works efficiently using a nearly optimal number of samples in any dimension d. We further illustrate the empirical performance of our algorithm on numerical examples.

1 Introduction

Interpolation is the fundamental problem of recovering a function from a finite number of (noisy) observations. To provide accurate and reliable predictions at unobserved points we need to avoid overfitting the target function which is typically achieved through restricting our interpolant to a family of *smooth or structured* functions. In this paper we focus on interpolating square-integrable functions on the d-dimensional unit sphere, with low-degree spherical harmonics. Spherical harmonics are essential in various theoretical and practical applications, including the representation of electromagnetic fields [Wei95], gravitational potential [Wer97], cosmic microwave background radiation [KKS97] and medical imaging [CLL15], as well as modelling of 3D shapes in computer graphics [KFR03].

We begin by observing that any function f in $L^2(\mathbb{S}^{d-1})$, i.e., the family of square-integrable functions defined on the sphere $\mathbb{S}^{d-1}$, can be uniquely decomposed into orthogonal spherical harmonic components. Specifically, if we denote the space of spherical harmonics of degree ℓ in dimension d by $\mathcal{H}_\ell(\mathbb{S}^{d-1})$, then any function $f \in L^2(\mathbb{S}^{d-1})$ has a unique orthogonal expansion $f = \sum_{\ell=0}^{\infty} f_\ell$ with $f_\ell \in \mathcal{H}_\ell(\mathbb{S}^{d-1})$ (see Lemma 6). With this observation, we aim to solve the following problem of finding the best degree $\leq q$ spherical harmonic approximation to f using a minimal number of samples (by essentially treating the higher order terms in f's expansion as noise).

Problem 1 (Informal Version of Problem 2). *For an unknown function $f \in L^2(\mathbb{S}^{d-1})$ and an integer $q \geq 1$, efficiently (both in terms of number of samples from f and computations) learn the first $q + 1$*

spherical harmonic components $\left\{f_\ell \in \mathcal{H}_\ell(\mathbb{S}^{d-1})\right\}_{\ell=0}^q$ *of* f *which minimizes*

$$\int_{\mathbb{S}^{d-1}} \left| \sum_{\ell=0}^{q} f_\ell(w) - f(w) \right|^2 dw. \tag{1}$$

The *angular power spectrum* $\|f_\ell\|_{\mathbb{S}^{d-1}}^2$ of f commonly obeys a power law decay. In fact, for any infinitely differentiable f, $\|f_\ell\|_{\mathbb{S}^{d-1}}^2$ decays asymptotically faster than any rational function of ℓ. Furthermore, for any real analytic f on the sphere, $\|f_\ell\|_{\mathbb{S}^{d-1}}^2$ decays exponentially. Thus, the first $q+1$ spherical harmonic components of f should well approximate f for even modest q, and answering Problem 1 is significantly useful for a wide range of differentiable functions.

1.1 Our Main Results

We reformulate Problem 1 as a least-squares regression and then solve it using techniques from randomized numerical linear algebra. To do so, we first consider an orthonormal projection operator that maps functions in $L^2(\mathbb{S}^{d-1})$ onto the space of bounded-degree spherical harmonics $\bigoplus_{\ell=0}^q \mathcal{H}_\ell(\mathbb{S}^{d-1})$. Specifically, if $\mathcal{K}_d^{(q)}$ is the projection operator that maps any function f with spherical harmonic expansion $f = \sum_{\ell=0}^\infty f_\ell$ with $f_\ell \in \mathcal{H}_\ell(\mathbb{S}^{d-1})$ to $\mathcal{K}_d^{(q)} f = \sum_{\ell=0}^q f_\ell$, Problem 1 can be formulated as

$$\min_{g \in L^2(\mathbb{S}^{d-1})} \int_{\mathbb{S}^{d-1}} \left| \left[\mathcal{K}_d^{(q)} g\right](w) - f(w) \right|^2 dw.$$

However, solving this regression problem with "continuous" cost function is challenging. To resolve this issue, we adopt the approach of [AKM+19] which discretizes the aforementioned regression problem according to the leverage function of the operator $\mathcal{K}_d^{(q)}$. Specifically, if we can randomly draw samples with probability proportional to the leverage function then we can recover degree-q spherical harmonic expansion of f, i.e. $\sum_{\ell=0}^q f_\ell$, with finite number of observations. In particular, by exploiting the connections between spherical harmonics and *Zonal (Gegenbauer) Harmonics* and the fact that zonal harmonics are the reproducing kernels of $\mathcal{H}_\ell(\mathbb{S}^{d-1})$ (Lemma 7), we prove that the leverage function of the operator $\mathcal{K}_d^{(q)}$ is constant. Thus, solving a discrete regression problem with uniformly sampled observations yields near-optimal solution to Problem 1. Our informal results are the following.

Theorem 1 (Informal Version of Theorem 14). *Let* $\beta_{q,d}$ *be the dimension of spherical harmonics of degree at most q, i.e.,* $\beta_{q,d} \equiv \dim\left(\bigoplus_{\ell=0}^q \mathcal{H}_\ell(\mathbb{S}^{d-1})\right)$. *There exists an algorithm that finds a* $(1+\epsilon)$*-approximation to the optimal solution of Problem 1, given* $s = \mathcal{O}(\epsilon^{-2}\beta_{q,d}\log\beta_{q,d})$ *observations of* f *at uniformly sampled points on* $\mathbb{S}^{d-1}$. *Moreover, the algorithm runs in* $\mathcal{O}(s^2 d + s^\omega)$[1] *time.*

We also prove that our bound on the number of required samples is optimal up to a logarithmic factor.

Theorem 2 (Informal Version of Theorem 15). *Any (randomized) algorithm that takes* $s < \beta_{q,d}$ *samples on any input fails with probability greater than* $9/10$, *where* $\beta_{q,d} \equiv \dim\left(\bigoplus_{\ell=0}^q \mathcal{H}_\ell(\mathbb{S}^{d-1})\right)$.

[1] $\omega < 2.3727$ is the exponent of the fast matrix multiplication algorithm [Wil12]

1.2 Related Work

Reconstruction of functions from small number of samples as per Problem 1 has been extensively studies in many areas of science and engineering. Prior results mainly consider reconstructing 1-dimensional functions from a finite number of samples on a finite interval under the assumption that the underlying function is smooth or structured in some sense. Notably, the influential line of work of [SP61, LP61, LP62, XRY01] focuses on reconstructing Fourier-bandlimited functions and the work of [CKPS16, EMM20] consider interpolating Fourier-sparse signals. Recently, [AKM+19] unified the reconstruction methods in dimension $d = 1$ and gave a universal sampling framework for reconstructing nearly all classes of functions with Fourier-based smoothness constraints.

One can view 1-dimensional functions on a finite interval as function on the unit circle $\mathbb{S}^1$. Thus, Problem 1 is indeed a generalization of prior works to reconstruction of functions on $\mathbb{S}^{d-1}$ under the assumption that the *generalized Fourier series* (Lemma 6) of the underlying function only contains bounded-degree spherical harmonics. This degree constraint on spherical harmonic expansions can be viewed as the d-dimensional analog of the Fourier-bandlimited function on circle $\mathbb{S}^1$.

Computing spherical harmonic expansions in dimension $d = 3$ has received considerable attention in physics and applied mathematics communities. The algorithms for this special case of Problem 1 are known in the literature as "fast spherical harmonic transform" [SS00, ST02]. Most notably, [RT06] proposed an algorithm for computing spherical harmonic expansion of degree $\le q$ to precision ϵ using $\mathcal{O}(\beta_{q,3})$ samples and $\mathcal{O}(\beta_{q,3} \log \beta_{q,3} \cdot \log(1/\epsilon))$ time. These fast algorithms were developed based on the fast Fourier transform and fast associated Legendre transform and require access to a (well-conditioned) orthogonal basis of $\mathcal{H}_\ell(\mathbb{S}^{d-1})$, which happened to be the associated Legendre polynomials when $d = 3$. However, it is in general intractable to compute an orthogonal basis for spherical harmonics [MNY06], so unlike our Theorem 1, it is inefficient to extend these prior results to higher d.

2 Mathematical Preliminaries

We denote by $\mathbb{S}^{d-1}$ the unit sphere in d dimension. We use $|\mathbb{S}^{d-1}| = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ to denote the surface area of the sphere $\mathbb{S}^{d-1}$ and $\mathcal{U}(\mathbb{S}^{d-1})$ to denote the uniform probability distribution on $\mathbb{S}^{d-1}$. We denote by $L^2\left(\mathbb{S}^{d-1}\right)$ the set of all square-integrable real-valued functions on the sphere $\mathbb{S}^{d-1}$. Furthermore, for any $f, g \in L^2\left(\mathbb{S}^{d-1}\right)$ we use the following definition of inner product on the unit sphere,

$$\langle f, g \rangle_{\mathbb{S}^{d-1}} \coloneqq \int_{\mathbb{S}^{d-1}} f(w) g(w) dw = \left| \mathbb{S}^{d-1} \right| \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}[f(w) g(w)]. \tag{2}$$

The function space $L^2\left(\mathbb{S}^{d-1}\right)$ is complete with respect to the norm induced by the inner product, i.e. $\|f\|_{\mathbb{S}^{d-1}} \coloneqq \sqrt{\langle f, f \rangle_{\mathbb{S}^{d-1}}}$, so $L^2\left(\mathbb{S}^{d-1}\right)$ is a *Hilbert space*.

We often use the term *quasi-matrix* which informally defines as a "matrix" in which one dimension is finite while the other is infinite. A quasi-matrix can be *tall* (or *wide*) in which there is a finite number of columns (or rows) where each one is a functional operator. For more details and a formal definition, see [SA22].

Our results are profoundly related to the *Spherical Harmonics*, which are special functions defined on $\mathbb{S}^{d-1}$ and are often employed in solving partial differential equations. *Harmonics* are solutions to the Laplace's equation on some domain. Spherical harmonics are the harmonics on a spherical domain, i.e. the solution of Laplace's equation in spherical domains. Formally,

Definition 3 (Spherical Harmonics). *For integers $\ell \geq 0$ and $d \geq 1$, let $\mathcal{P}_\ell(d)$ be the space of degree-ℓ homogeneous polynomials with d variables and real coefficients. Let $\mathcal{H}_\ell(d)$ denote the space of degree-ℓ harmonic polynomials in dimension d, i.e., homogeneous polynomial solutions of Laplace's equation:*

$$\mathcal{H}_\ell(d) := \{P \in \mathcal{P}_\ell(d) : \Delta P = 0\},$$

where $\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}$ is the Laplace operator on $\mathbb{R}^d$. Finally, let $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ be the space of (real) Spherical Harmonics *of order ℓ in dimension d, i.e. restrictions of harmonic polynomials in $\mathcal{H}_\ell(d)$ to the sphere $\mathbb{S}^{d-1}$. The dimension of this space, $\alpha_{\ell,d} \equiv \dim\left(\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)\right)$, is*

$$\alpha_{0,d} = 1, \quad \alpha_{1,d} = d, \quad \alpha_{\ell,d} = \binom{d+\ell-1}{\ell} - \binom{d+\ell-3}{\ell-2} \quad \textit{for } \ell \geq 2.$$

2.1 Gegenbauer Polynomials

The *Gegenbauer* (a.k.a. *ultraspherical*) *polynomial* of degree $\ell \geq 0$ in dimension $d \geq 2$ is given by

$$P_d^\ell(t) := \sum_{j=0}^{\lfloor \ell/2 \rfloor} c_j \cdot t^{\ell-2j} \cdot (1-t^2)^j, \tag{3}$$

where $c_0 = 1$ and $c_{j+1} = -\frac{(\ell-2j)(\ell-2j-1)}{2(j+1)(d-1+2j)} \cdot c_j$ for $j = 0, 1, \ldots, \lfloor \ell/2 \rfloor - 1$. These polynomials satisfy the orthogonality properties on the interval $[-1, 1]$ with respect to the measure $(1-t^2)^{\frac{d-3}{2}}$, i.e.,

$$\int_{-1}^{1} P_d^\ell(t) \cdot P_d^{\ell'}(t) \cdot (1-t^2)^{\frac{d-3}{2}}\, dt = \frac{\left|\mathbb{S}^{d-1}\right|}{\alpha_{\ell,d} \cdot \left|\mathbb{S}^{d-2}\right|} \cdot \mathbb{1}_{\{\ell=\ell'\}}. \tag{4}$$

Zonal Harmonics. The Gegenbauer polynomials naturally provide positive definite dot-product kernels on $\mathbb{S}^{d-1}$ known as *Zonal Harmonics*, which are closely related to the spherical harmonics. The following reproducing property of zonal harmonics plays a crucial role in our analysis.

Lemma 4 (Reproducing Property of Zonal Harmonics). *Let $P_d^\ell(\cdot)$ be the Gengenbauer polynomial of degree ℓ in dimension d. For any $x, y \in \mathbb{S}^{d-1}$:*

$$P_d^\ell(\langle x, y \rangle) = \alpha_{\ell,d} \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle x, w \rangle\right) P_d^\ell\left(\langle y, w \rangle\right)\right],$$

Furthermore, for any $\ell' \neq \ell$:

$$\mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle x, w \rangle\right) \cdot P_d^{\ell'}\left(\langle y, w \rangle\right)\right] = 0.$$

The proof, like most proofs, is deferred to the appendix. The following very useful fact (a.k.a. Addition Theorem) connects Gegenbauer polynomials and spherical harmonics.

Theorem 5 (Addition Theorem). *For every integer $\ell \geq 0$, if $\left\{y_1^\ell, y_2^\ell, \ldots, y_{\alpha_{\ell,d}}^\ell\right\}$ is an orthonormal basis for $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$, then for any $\sigma, w \in \mathbb{S}^{d-1}$ we have*

$$\frac{\alpha_{\ell,d}}{\left|\mathbb{S}^{d-1}\right|} \cdot P_d^\ell\left(\langle \sigma, w \rangle\right) = \sum_{j=1}^{\alpha_{\ell,d}} y_j^\ell(\sigma) \cdot y_j^\ell(w).$$

3 Reconstruction of $L^2\left(\mathbb{S}^{d-1}\right)$ Functions via Spherical Harmonics

In this section we show how to reconstruct any function $f \in L^2\left(\mathbb{S}^{d-1}\right)$ from optimal number of samples via the spherical harmonics. We begin with showing that the spherical harmonics form a complete set of orthonormal functions and thus form an orthonormal basis of the Hilbert space of square-integrable functions on the surface of the sphere $\mathbb{S}^{d-1}$. This is analogous to periodic functions, viewed as functions defined on a circle, which can be expressed as a linear combination of circular functions (sines and cosines) via the Fourier series.

Lemma 6 (Direct Sum Decomposition of $L^2(\mathbb{S}^{d-1})$). *The family of spaces $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ yields a Hilbert space direct sum decomposition $L^2\left(\mathbb{S}^{d-1}\right) = \bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$: the summands are closed and pairwise orthogonal, and that every $f \in L^2\left(\mathbb{S}^{d-1}\right)$ is the sum of a converging series (in the sense of mean-square convergence with the L^2-norm defined in Eq. (2)),*

$$f = \sum_{\ell=0}^{\infty} f_\ell,$$

where $f_\ell \in \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ are uniquely determined functions. Furthermore, given any orthonormal basis $\left\{y_1^\ell, y_2^\ell, \ldots, y_{\alpha_{\ell,d}}^\ell\right\}$ of $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ we have $f_\ell = \sum_{j=1}^{\alpha_{\ell,d}} \langle f, y_j^\ell \rangle_{\mathbb{S}^{d-1}} \cdot y_j^\ell$.

The series expansion in Lemma 6 is the analog of the Fourier expansion of periodic functions, and is known as f's "*generalized Fourier series*" [Pen30] with respect to the Hilbert basis $\left\{y_j^\ell : j \in [\alpha_{\ell,d}], \ell \geq 0\right\}$. We remark that it is in general intractable to compute an orthogonal basis for the space of spherical harmonics [MNY06], which renders the generalized Fourier series expansion in Lemma 6 primarily existential. While finding the generalized Fourier expansion of a function $f \in L^2\left(\mathbb{S}^{d-1}\right)$ is computationally intractable, our goal is to answer the next fundamental question, which is about finding the projection of a function f onto the space of spherical harmonics, i.e., the f_ℓ's in Lemma 6. Concretely, we seek to solve the following problem.

Problem 2. *For an integer $q \geq 0$ and a given input function $f \in L^2\left(\mathbb{S}^{d-1}\right)$ whose decomposition over the Hilbert sum $\bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ is $f = \sum_{\ell=0}^{\infty} f_\ell$ as per Lemma 6, let us define the low-degree expansion of this function as $f^{(q)} := \sum_{\ell=0}^{q} f_\ell$. How efficiently can we learn $f^{(q)} \in \bigoplus_{\ell=0}^{q} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$? More precisely, we want to find a set $\{w_1, w_2, \ldots, w_s\} \subseteq \mathbb{S}^{d-1}$ with minimal cardinality s along with an efficient algorithm that given samples $\{f(w_i)\}_{i=1}^{s}$ can interpolate $f(\cdot)$ with a function $\tilde{f}^{(q)} \in \bigoplus_{\ell=0}^{q} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ such that:*

$$\left\|\tilde{f}^{(q)} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \leq \epsilon \cdot \left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2.$$

To see why learning the low-degree expansion of a function f in Problem 2 makes sense, note that the *angular power spectrum* of f commonly obeys a power law decay of the form $\|f_\ell\|_{\mathbb{S}^{d-1}}^2 \leq \mathcal{O}(\ell^{-s})$, for some $s > 0$, depending on the order of differentiability of f. In particular, the Sobolev inequalities imply that for any infinitely differentiable f, $\|f_\ell\|_{\mathbb{S}^{d-1}}^2$ decays faster than any rational function of ℓ as $\ell \to \infty$. Thus, $f^{(q)}$ should well approximate f for even modest q, and learning the low-degree expansion $f^{(q)}$ is extremely useful for a wide range of differentiable functions.

For ease of notation, we denote the Hilbert space of spherical harmonics of degree at most q by $\mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right) := \bigoplus_{\ell=0}^{q} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$. To answer Problem 2 we exploit the close connection between the spherical harmonics and Gengenbauer polynomials, and in particular the fact that the zonal harmonics are the reproducing kernels of the Hilbert spaces $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$.

Lemma 7 (Reproducing Kernel of $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$). *For every $f \in L^2\left(\mathbb{S}^{d-1}\right)$, if $f = \sum_{\ell=0}^{\infty} f_\ell$ is the unique decomposition of f over $\bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ as per Lemma 6, then f_ℓ is given by*

$$f_\ell(\sigma) = \alpha_{\ell,d} \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[f(w) P_d^\ell\left(\langle \sigma, w \rangle\right)\right] \quad \text{for } \sigma \in \mathbb{S}^{d-1}.$$

Now we define a kernel operator, based on the low-degree Gegenbauer polynomials, which projects functions onto their low-degree spherical harmonic expansion.

Definition 8 (Projection Operator onto $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$). *For any integers $q \geq 0$ and $d \geq 2$, define the kernel operator $\mathcal{K}_d^{(q)} : L^2\left(\mathbb{S}^{d-1}\right) \to L^2\left(\mathbb{S}^{d-1}\right)$ as follows: for $f \in L^2\left(\mathbb{S}^{d-1}\right)$ and $\sigma \in \mathbb{S}^{d-1}$,*

$$\left[\mathcal{K}_d^{(q)} f\right](\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} \left\langle f, P_d^\ell\left(\langle \sigma, \cdot \rangle\right)\right\rangle_{\mathbb{S}^{d-1}} = \sum_{\ell=0}^{q} \alpha_{\ell,d} \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[f(w) P_d^\ell\left(\langle \sigma, w \rangle\right)\right]. \tag{5}$$

This is an integral operator with kernel function $k_{q,d}(\sigma, w) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} \cdot P_d^\ell\left(\langle \sigma, w \rangle\right)$.

Now note that the operator $\mathcal{K}_d^{(q)}$ is self-adjoint and positive semi-definite. Moreover, using the reproducing property of this kernel we can establish that $\mathcal{K}_d^{(q)}$ is a projection operator.

Claim 1. *The operator $\mathcal{K}_d^{(q)}$ defined in Definition 8 satisfies the property $\left(\mathcal{K}_d^{(q)}\right)^2 = \mathcal{K}_d^{(q)}$.*

Furthermore, by the addition theorem (Theorem 5), the operator $\mathcal{K}_d^{(q)}$ is trace-class (i.e., the trace is finite and independent of the choice of basis) because

$$\begin{aligned} \operatorname{trace}\left(\mathcal{K}_d^{(q)}\right) &= \int_{\mathbb{S}^{d-1}} k_{q,d}(w, w)\, dw \\ &= \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} \cdot \int_{\mathbb{S}^{d-1}} P_d^\ell\left(\langle w, w \rangle\right)\, dw \\ &= \sum_{\ell=0}^{q} \alpha_{\ell,d} = \binom{d+q-1}{q} + \binom{d+q-2}{q-1} - 1. \end{aligned} \tag{6}$$

By combining Theorem 5 and Lemma 6, and using the definition of the projection operator $\mathcal{K}_d^{(q)}$, it follows that for any function $f \in L^2\left(\mathbb{S}^{d-1}\right)$ with Hilbert sum decomposition $f = \sum_{\ell=0}^{\infty} f_\ell$, the low-degree component $f^{(q)} = \sum_{\ell=0}^{q} f_\ell \in \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$ can be computed as $f^{(q)} = \mathcal{K}_d^{(q)} f$. Equivalently, in order to learn $f^{(q)}$, it suffices to solve the following least-squares regression problem,

$$\min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \left\| \mathcal{K}_d^{(q)} g - f \right\|_{\mathbb{S}^{d-1}}^2. \tag{7}$$

If g^* is an optimal solution to the above regression problem then $f^{(q)} = \mathcal{K}_d^{(q)} g^*$. In the next claim we show that solving the least squares problem in Eq. (7), even to a coarse approximation, is sufficient to solve our interpolation problem (i.e., Problem 2):

Claim 2. *For any function $f \in L^2\left(\mathbb{S}^{d-1}\right)$, any integer $q \geq 0$, and any $C \geq 1$, if $\tilde{g} \in L^2\left(\mathbb{S}^{d-1}\right)$ is a function that satisfies,*

$$\left\|\mathcal{K}_d^{(q)}\tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 \leq C \cdot \min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2,$$

and if we let $f^{(q)} := \mathcal{K}_d^{(q)} f$, where $\mathcal{K}_d^{(q)}$ is defined as per Definition 8, then the following holds

$$\left\|\mathcal{K}_d^{(q)}\tilde{g} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \leq (C-1) \cdot \left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2.$$

Claim 2 shows that solving the regression problem in Eq. (7) approximately provides a solution to our spherical harmonics interpolation problem (Problem 2). But how can we solve this least-squares problem efficiently? Not only does the problem involve a possibly infinite dimensional parameter vector g, but the objective function also involves the continuous domain on the surface of $\mathbb{S}^{d-1}$.

3.1 Randomized Discretization via Leverage Function Sampling

We solve the continuous regression in Eq. (7) by randomly discretizing the sphere $\mathbb{S}^{d-1}$, thereby reducing our problem to a regression on a finite set of points $w_1, w_2, \ldots, w_s \in \mathbb{S}^{d-1}$. In particular, we propose to sample points on $\mathbb{S}^{d-1}$ with probability proportional to the so-called *leverage function*, a specific distribution that has been widely applied in randomized algorithms for linear algebra problems on discrete matrices [LMP13]. We start with the definition of the leverage function:

Definition 9 (Leverage Function). *For integers $q \geq 0$ and $d > 0$, we define the leverage function of the operator $\mathcal{K}_d^{(q)}$ (see Definition 8) for every $w \in \mathbb{S}^{d-1}$ as follows,*

$$\tau_q(w) := \max_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g\right\|_{\mathbb{S}^{d-1}}^{-2} \cdot \left|\left[\mathcal{K}_d^{(q)} g\right](w)\right|^2 \tag{8}$$

Intuitively, $\tau_q(w)$ is an upper bound of how much a function that is spanned by the eigenfunctions of the operator $\mathcal{K}_d^{(q)}$ can "blow up" at w. The larger the leverage function $\tau_q(w)$ implies the higher the probability we will be required to sample w. This ensures that our sample points well reflect any possibly significant components, or "spikes", of the function. Ultimately, the integral $\int_{\mathbb{S}^{d-1}} \tau_q(w)\, dw$ determines how many samples we require to solve the regression problem Eq. (7) to a given accuracy. It is an already known fact that the leverage function integrates to the rank of the operator $\mathcal{K}_d^{(q)}$ (which turns out to be equal to the dimensionality of the Hilbert space $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$). This will ultimately allow us to achieve a $\widetilde{\mathcal{O}}(\sum_{\ell=0}^{q} \alpha_{\ell,d})$ sample complexity bound for solving the interpolation Problem 2. To express the leverage function as a closed form, we make use of the following lemma that gives a useful alternative characterization of the leverage function.

Lemma 10 (Min Characterization of the Leverage Function). *For any $w \in \mathbb{S}^{d-1}$, let $\tau_q(w)$ be the leverage function (Definition 9) and define $\phi_w \in L^2(\mathbb{S}^{d-1})$ by $\phi_w(\sigma) \equiv \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell(\langle \sigma, w \rangle)$. We have the following minimization characterization of the leverage function:*

$$\tau_q(w) = \left\{ \min_{g \in L^2(\mathbb{S}^{d-1})} \|g\|_{\mathbb{S}^{d-1}}^2, \quad \textit{s.t.} \quad \mathcal{K}_d^{(q)} g = \phi_w \right\}. \tag{9}$$

We prove this lemma in Appendix C. Using the minimization and maximization characterizations of the leverage function we can find upper and lower bounds on this function. Surprisingly, in this case the upper and lower bounds match, so we actually have an exact value for the leverage function.

Lemma 11 (Leverage Function is Constant). *The leverage function given in Definition 9 is equal to* $\tau_q(w) = \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}$ *for every* $w \in \mathbb{S}^{d-1}$.

Proof. First we prove that $\tau_q(w) \le \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}$ using the min-characterization. If we let $\phi_w \in L^2(\mathbb{S}^{d-1})$ be defined as $\phi_w(\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell(\langle \sigma, w\rangle)$, then by Definition 8, for every $\sigma \in \mathbb{S}^{d-1}$ we can write,

$$\begin{aligned}\left[\mathcal{K}_d^{(q)} \phi_w\right](\sigma) &= \sum_{\ell=0}^{q} \alpha_{\ell,d} \cdot \mathbb{E}_{v \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell(\langle \sigma, v\rangle) \cdot \phi_w(v)\right] \\ &= \sum_{\ell=0}^{q} \sum_{\ell'=0}^{q} \frac{\alpha_{\ell,d}\alpha_{\ell',d}}{|\mathbb{S}^{d-1}|} \cdot \mathbb{E}_{v \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell(\langle \sigma, v\rangle) \cdot P_d^{\ell'}(\langle v, w\rangle)\right] \\ &= \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell(\langle \sigma, w\rangle) = \phi_w(\sigma), \end{aligned} \tag{10}$$

where the third line above follows from Lemma 4. Therefore, the test function $g := \phi_w$ satisfies the constraint of the minimization in Eq. (9), i.e., $\mathcal{K}_d^{(q)} g = \phi_w$. Thus, Lemma 10 implies that,

$$\tau_q(w) \le \|g\|_{\mathbb{S}^{d-1}}^2 = \|\phi_w\|_{\mathbb{S}^{d-1}}^2 = \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|},$$

where the equality above follows from Lemma 4 along with Eq. (2). This establishes the upper bound on the leverage function that we sought to prove.

Now, using the maximization characterization of the leverage function in Definition 9, we prove that $\tau_q(w) \ge \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}$. Again, we consider the same test function $g = \phi_w$ and write,

$$\begin{aligned}\left\|\mathcal{K}_d^{(q)} \phi_w\right\|_{\mathbb{S}^{d-1}}^{-2} \cdot \left|\left[\mathcal{K}_d^{(q)} \phi_w\right](w)\right|^2 &= \frac{|\phi_w(w)|^2}{\|\phi_w\|_{\mathbb{S}^{d-1}}^2} \\ &= \frac{\left|\sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell(\langle w, w\rangle)\right|^2}{\sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}} \\ &= \frac{\left|\sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell(1)\right|^2}{\sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}} = \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|},\end{aligned}$$

where the first and second line above follow from Eq. (10) and Lemma 4, respectively. Therefore, the max characterization of the leverage function in Definition 9 implies that,

$$\tau_q(w) \ge \left\|\mathcal{K}_d^{(q)} \phi_w\right\|_{\mathbb{S}^{d-1}}^{-2} \cdot \left|\left[\mathcal{K}_d^{(q)} \phi_w\right](w)\right|^2 = \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}.$$

This completes the proof of Lemma 11 and establishes that $\tau_q(w)$ is uniformly equal to $\sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|}$.

□

The integral of the leverage function, which determines the total samples needed to solve our least-squares regression, is therefore equal to the dimensionality of the Hilbert space $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$.

Corollary 12. *The leverage function defined in Definition 9 integrates to the dimensionality of the Hilbert space $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$, which we denote by $\beta_{q,d}$, i.e.,*

$$\int_{\mathbb{S}^{d-1}} \tau_q(w)\, dw = \dim\left(\mathcal{H}^{(q)}(\mathbb{S}^{d-1})\right) = \sum_{\ell=0}^{q} \alpha_{\ell,d} \equiv \beta_{q,d}.$$

We now show that the leverage function can be used to randomly sample the points on the unit sphere to discretize the regression problem in Eq. (7) and solve it approximately.

Theorem 13 (Approximate Regression via Leverage Function Sampling). *For any $\epsilon, \delta > 0$, let $s = c \cdot \frac{\beta_{q,d}}{\epsilon^2}\left(\log \beta_{q,d} + \delta^{-1}\right)$, for sufficiently large fixed constant c, and let $w_1, w_2, \ldots, w_s$ be i.i.d. uniform samples on $\mathbb{S}^{d-1}$. Define the quasi-matrix $\boldsymbol{P} : \mathbb{R}^s \to L^2(\mathbb{S}^{d-1})$ as follows, for every $v \in \mathbb{R}^d$:*

$$[\boldsymbol{P} \cdot v](\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s \cdot |\mathbb{S}^{d-1}|}} \cdot \sum_{j=1}^{s} v_j \cdot P_d^{\ell}\left(\langle w_j, \sigma\rangle\right) \quad \text{for } \sigma \in \mathbb{S}^{d-1}.$$

Also let $\boldsymbol{f} \in \mathbb{R}^s$ be a vector with $\boldsymbol{f}_j := \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot f(w_j)$ for $j = 1, 2, \ldots, s$ and let $\boldsymbol{P}^*$ be the adjoint of $\boldsymbol{P}$. If $\tilde{g}$ is an optimal solution to the following least-squares problem

$$\tilde{g} \in \underset{g \in L^2\left(\mathbb{S}^{d-1}\right)}{\arg\min} \ \|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2,$$

then with probability at least $1 - \delta$ the following holds,

$$\left\|\mathcal{K}_d^{(q)} \tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 \le (1 + \epsilon) \cdot \min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2.$$

We prove this theorem in Appendix C. Theorem 13 shows that the function $\tilde{g}$ obtained from solving the discretized regression problem provides an approximate solution to Eq. (7).

3.2 Efficient Solution for the Discretized Least-Squares Problem

In this section, we demonstrate how to apply Theorem 13 algorithmically to approximately solve the regression problem of Eq. (7). Specifically, we show how to use the *kernel trick* to solve the randomly discretized least squares problem efficiently.

Theorem 14 (Efficient Spherical Harmonic Interpolation). *Algorithm 1 returns a function $y \in \mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ such that, with probability at least $1 - \delta$:*

$$\left\|y - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \le \epsilon \cdot \left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2,$$

where $f^{(q)} := \mathcal{K}_d^{(q)} f$. Suppose we can compute the Gegenbauer polynomial $P_d^{\ell}(t)$ at every point $t \in [-1, 1]$ in constant time. Algorithm 1 queries the function f at $s = \mathcal{O}\left(\frac{\beta_{q,d}}{\epsilon^2}\left(\log \beta_{q,d} + \delta^{-1}\right)\right)$ points on the sphere $\mathbb{S}^{d-1}$ and runs in $\mathcal{O}(s^2 \cdot d + s^{\omega})$ time. This algorithm evaluates $y(\sigma)$ in $\mathcal{O}(d \cdot s)$ time for any $\sigma \in \mathbb{S}^{d-1}$.

For a proof of this theorem see Appendix D.

Algorithm 1 Efficient Spherical Harmonic Expansion

input: accuracy parameter $\epsilon > 0$, failure probability $\delta \in (0,1)$, integer $q \geq 0$
Set $s = c \cdot \frac{\beta_{q,d}}{\epsilon^2}(\log \beta_{q,d} + \delta^{-1})$ for sufficiently large fixed constant c
Sample i.i.d. random points $w_1, w_2, \dots, w_s$ from a uniform distribution on $\mathbb{S}^{d-1}$
Compute $\boldsymbol{K} \in \mathbb{R}^{s \times s}$ with $\boldsymbol{K}_{i,j} = \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{s} \cdot P_d^{\ell}(\langle w_i, w_j \rangle)$ for $i, j \in [s]$
Compute $\boldsymbol{f} \in \mathbb{R}^s$ with $\boldsymbol{f}_j = \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot f(w_j)$ for $j \in [s]$
Solve the regression by computing $\boldsymbol{z} = \boldsymbol{K}^{\dagger} \boldsymbol{f}$
return $y \in \mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ with $y(\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s \cdot |\mathbb{S}^{d-1}|}} \cdot \sum_{j=1}^{s} \boldsymbol{z}_j \cdot P_d^{\ell}(\langle w_j, \sigma \rangle)$ for $\sigma \in \mathbb{S}^{d-1}$

4 Lower Bound on The Number of Required Observations

We conclude by showing that the dimensionality of the Hilbert space $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ tightly characterizes the sample complexity of Problem 2. Thus, our Theorem 14 is optimal up to a logarithmic factor. The crucial fact that we use for proving the lower bound is that all the eigenvalues of the operator $\mathcal{K}_d^{(q)}$ are equal to one. This fact follows from the addition theorem presented in Theorem 5. By this lemma, if $\left\{ y_1^{\ell}, y_2^{\ell}, \dots, y_{\alpha_{\ell,d}}^{\ell} \right\}$ is an orthonormal basis of $\mathcal{H}_{\ell}\left(\mathbb{S}^{d-1}\right)$, then for any function $f \in L^2\left(\mathbb{S}^{d-1}\right)$,

$$\begin{aligned}
\left[\mathcal{K}_d^{(q)} f\right](\sigma) &= \sum_{\ell=0}^{q} \alpha_{\ell,d} \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^{\ell}(\langle \sigma, w \rangle) \cdot f(w)\right] \\
&= \sum_{\ell=0}^{q} \left|\mathbb{S}^{d-1}\right| \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[\sum_{j=1}^{\alpha_{\ell,d}} y_j^{\ell}(\sigma) \cdot y_j^{\ell}(w) \cdot f(w)\right] \\
&= \sum_{\ell=0}^{q} \sum_{j=1}^{\alpha_{\ell,d}} \langle y_j^{\ell}, f \rangle_{\mathbb{S}^{d-1}} \cdot y_j^{\ell}(\sigma). \qquad (11)
\end{aligned}$$

This shows that all (non-zero) eigenvalues of the operator $\mathcal{K}_d^{(q)}$ are equal to one.

Theorem 15 (Lower Bound). *Consider an error parameter $\epsilon > 0$, and any (possibly randomized) algorithm that solves Problem 2 with probability greater than $1/10$ for any input function f and makes at most r (possibly adaptive) queries on any input. Then $r \geq \beta_{q,d}$.*

We prove this lower bound by describing a distribution on the input functions f on which any deterministic algorithm that takes $r < \beta_{q,d}$ samples on any input fails with probability greater than $9/10$. The theorem then follows by Yao's principle.

Hard Input Distribution. For any integer $\ell \leq q$, consider an orthonormal basis of $\mathcal{H}_{\ell}\left(\mathbb{S}^{d-1}\right)$ and denote it by $\left\{ y_1^{\ell}, y_2^{\ell}, \dots, y_{\alpha_{\ell,d}}^{\ell} \right\}$. Let $\boldsymbol{Y}_{\ell} : \mathbb{R}^{\alpha_{\ell,d}} \to \mathcal{H}_{\ell}\left(\mathbb{S}^{d-1}\right)$ be the quasi-matrix with y_j^{ℓ} as its j^{th} column, i.e., $[\boldsymbol{Y}_{\ell} \cdot u](\sigma) := \sum_{j=1}^{\alpha_{\ell,d}} u_j \cdot y_j^{\ell}(\sigma)$ for any $u \in \mathbb{R}^{\alpha_{\ell,d}}$ and $\sigma \in \mathbb{S}^{d-1}$. Let vectors $v^{(0)} \in \mathbb{R}^{\alpha_{0,d}}, v^{(1)} \in \mathbb{R}^{\alpha_{1,d}}, \dots, v^{(q)} \in \mathbb{R}^{\alpha_{q,d}}$ be independent random vectors with each entry distributed independently as a Gaussian: $v_j^{(\ell)} \sim \mathcal{N}(0,1)$. The random input is defined to be $f := \sum_{\ell=0}^{q} \boldsymbol{Y}_{\ell} \cdot v^{(\ell)}$. In other words, $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_{\ell} \cdot v^{(\ell)}$ is a random linear combination of the eigenfunctions of $\mathcal{K}_d^{(q)}$.

We prove that accurate reconstruction of function f drawn from the above-mentioned hard input distribution yields an accurate reconstruction of the random vectors $v^{(0)}, v^{(1)}, \ldots, v^{(q)}$. Since each $v^{(\ell)}$ is $\alpha_{\ell,d}$-dimensional, this reconstruction requires $\Omega(\sum_{\ell=0}^{q} \alpha_{\ell,d}) = \Omega(\beta_{q,d})$ samples, giving us a lower bound for accurately reconstructing f.

Claim 3. *Given the random input $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)}$ generated as described above, to solve Problem 2, an algorithm must return a function $\tilde{f}^{(q)} \in \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$ such that $\|\tilde{f}^{(q)} - f\|_{\mathbb{S}^{d-1}}^2 = 0$.*

We prove this claim in Appendix E. Now we show that finding an $\tilde{f}^{(q)}$ satisfying the condition of Problem 2 is at least as hard as accurately finding all vectors $v^{(0)}, v^{(1)}, \ldots, v^{(q)}$.

Lemma 16. *If a deterministic algorithm solves Problem 2 with probability at least $1/10$ over our random input distribution $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)}$, then with probability at least $1/10$, the output of the algorithm $\tilde{f}^{(q)}$ satisfies $\boldsymbol{Y}_\ell^* \tilde{f}^{(q)} = v^{(\ell)}$ for all integers $\ell \le q$.*

Finally, we complete the proof of Theorem 15 by arguing that if $\tilde{f}^{(q)}$ is formed using less than $\beta_{q,d}$ queries from f, then $\sum_{\ell=0}^{q} \left\| \boldsymbol{Y}_\ell^* \tilde{f}^{(q)} - v^{(\ell)} \right\|_2^2 > 0$ with good probability. Thus the bound of Lemma 16 cannot hold and so $\tilde{f}^{(q)}$ cannot be a solution to Problem 2 with good probability. Assume for the sake of contradiction that there is a deterministic algorithm which solves Problem 2 with probability at least $1/10$ over the random input $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)}$ that makes $r = \beta_{q,d} - 1$ queries on any input (we can always modify an algorithm that makes fewer queries on some inputs to make exactly $\beta_{\ell,d} - 1$ queries and return the same output). For every $\sigma \in \mathbb{S}^{d-1}$ and integer $\ell \le q$ let the vector $u_\sigma^\ell \in \mathbb{R}^{\alpha_{\ell,d}}$ be defined as $u_\sigma^\ell := \left[y_1^\ell(\sigma), y_2^\ell(\sigma), \ldots, y_{\alpha_{\ell,d}}^\ell(\sigma)\right]$. Also define $\boldsymbol{u}_\sigma \in \mathbb{R}^{\beta_{q,d}}$ as $\boldsymbol{u}_\sigma := \left[u_\sigma^0, u_\sigma^1, \ldots, u_\sigma^q\right]$. Furthermore, define $\boldsymbol{v} \in \mathbb{R}^{\beta_{q,d}}$ as $\boldsymbol{v} := \left(v^{(0)}, v^{(1)}, \ldots, v^{(q)}\right)$. Additionally, define the quasi-matrix $\boldsymbol{Y} := [\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_q]$.

Using the above notations and the definition of the hard input instance f, each query to f is in fact a query to the random vector $\boldsymbol{v}$ in the form of $f(\sigma) = \langle \boldsymbol{u}_\sigma, \boldsymbol{v} \rangle$. Now consider a deterministic function Q, that is given input $\boldsymbol{V} \in \mathbb{R}^{i \times \beta_{q,d}}$ (for any positive integer i) and outputs $Q(\boldsymbol{V}) \in \mathbb{R}^{\beta_{q,d} \times \beta_{q,d}}$ such that $Q(\boldsymbol{V})$ has orthonormal rows with the first i rows spanning the i rows of $\boldsymbol{V}$. If $\sigma_1, \sigma_2, \ldots, \sigma_r \in \mathbb{S}^{d-1}$ denote the points where our algorithm queries the input f, for any integer $i \in [r]$, let:

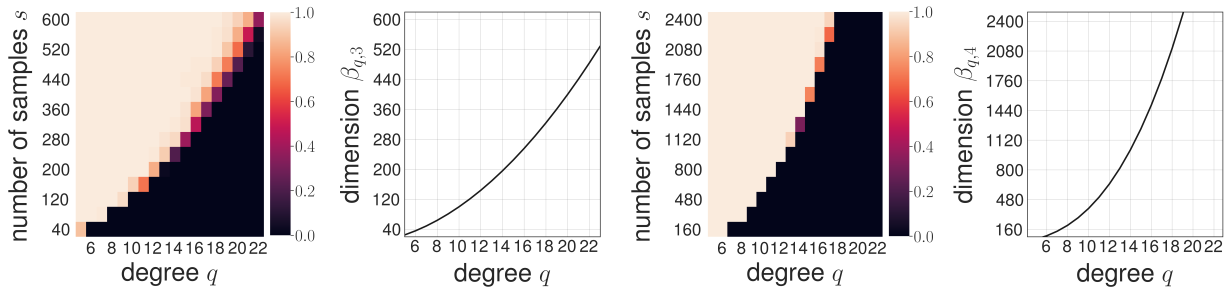
$$\boldsymbol{Q}^i := Q\left(\left[\boldsymbol{u}_{\sigma_1}, \boldsymbol{u}_{\sigma_2}, \ldots, \boldsymbol{u}_{\sigma_i}\right]^\top\right).$$

That is $\boldsymbol{Q}^i$ is an orthonormal matrix whose first i rows span the first i queries of the algorithm. Note that since our algorithm is deterministic, $\boldsymbol{Q}^i$ is a deterministic function of the random input $\boldsymbol{v}$. We have the following claim from [AKM+19]:

Claim 4 (Claim 23 of [AKM+19]). *Conditioned on the queries $f(\sigma_1), f(\sigma_2), \ldots, f(\sigma_r)$ for $r < \beta_{q,d}$, the variable $[\boldsymbol{Q}^r \cdot \boldsymbol{v}](\beta_{q,d})$ is distributed as $\mathcal{N}(0,1)$.*

Now using Claim 4 we can write,

$$\begin{aligned}
\Pr_{\boldsymbol{v}}\left[\sum_{\ell=0}^{q} \left\| v^{(\ell)} - \boldsymbol{Y}_\ell^* \tilde{f}^{(q)} \right\|_2^2 = 0\right] &= \Pr_{\boldsymbol{v}}\left[\boldsymbol{Q}^r \cdot \boldsymbol{v} = \boldsymbol{Q}^r \boldsymbol{Y}^* \tilde{f}^{(q)}\right] \qquad \text{(Since } \boldsymbol{Q}^r \text{ is orthonormal)} \\
&\le \Pr_{\boldsymbol{v}}\left[\left[\boldsymbol{Q}^r \boldsymbol{v}\right](\beta_{q,d}) = \left[\boldsymbol{Q}^r \boldsymbol{Y}^* \tilde{f}^{(q)}\right](\beta_{q,d})\right] \\
&= \mathbb{E}\left[\Pr_{\boldsymbol{v}}\left[\left[\boldsymbol{Q}^r \boldsymbol{v}\right](\beta_{q,d}) = \left[\boldsymbol{Q}^r \boldsymbol{Y}^* \tilde{f}^{(q)}\right](\beta_{q,d}) \middle| f(\sigma_1), \ldots, f(\sigma_r)\right]\right],
\end{aligned}$$



(a) $d = 3$ (b) $d = 4$

Figure 1: (Left) Empirical success probabilities of Algorithm 1 varying the number of samples s and the degree of spherical harmonic expansion q. (Right) The dimension $\beta_{q,d}$ of the Hilbert space $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ as a function of q when (a) $d = 3$ and (b) $d = 4$, respectively.

where the expectation in the last line is taken over the randomness of $f(\sigma_1), \ldots, f(\sigma_r)$. Now note that conditioned on $f(\sigma_1), \ldots, f(\sigma_r)$, the quantity $\left[\boldsymbol{Q}^r \boldsymbol{Y}^* \tilde{f}^{(q)}\right](\beta_{q,d})$ is a fixed vale because the algorithm determines $\tilde{f}^{(q)}$ given the knowledge of the queries $f(\sigma_1), \ldots, f(\sigma_r)$. Furthermore, by Claim 4, $[\boldsymbol{Q}^r \cdot \boldsymbol{v}](\beta_{q,d})$ is a random variable distributed as $\mathcal{N}(0, 1)$, conditioned on $f(\sigma_1), \ldots, f(\sigma_r)$. This implies that,

$$\Pr\left[[\boldsymbol{Q}^r \cdot \boldsymbol{v}](\beta_{q,d}) = \left[\boldsymbol{Q}^r \boldsymbol{Y}^* \tilde{f}^{(q)}\right](\beta_{q,d}) \middle| f(\sigma_1), \ldots, f(\sigma_r) \right] = 0.$$

Thus,

$$\Pr\left[\sum_{\ell=0}^{q} \left\| v^{(\ell)} - \boldsymbol{Y}_\ell^* \tilde{f}^{(q)} \right\|_2^2 = 0 \right] = \mathbb{E}_{f(\sigma_1),\ldots,f(\sigma_r)}[0] = 0.$$

However, we have assumed that this algorithm solves Problem 2 with probability at least $1/10$, and hence, by Lemma 16, $\Pr\left[\sum_{\ell=0}^{q} \|v^{(\ell)} - \boldsymbol{Y}_\ell^* \tilde{f}^{(q)}\|_2^2 = 0\right] \geq 1/10$. This is a contradiction, yielding Theorem 15.

5 Numerical Performance

We conduct experiments for verifying numerical performance of our algorithm. Specifically, for a fixed q, we generate a random function $f(\sigma) = \sum_{\ell=0}^{q} c_\ell P_d^\ell(\langle \sigma, v \rangle)$ where $v \sim \mathcal{U}(\mathbb{S}^{d-1})$ and c_ℓ's are i.i.d. samples from $\mathcal{N}(0, 1)$. Then, f is recovered by running Algorithm 1 with s random evaluations of f on $\mathbb{S}^{d-1}$. Note that $\|\mathcal{K}_d^{(q)} f - f\|_{\mathbb{S}^{d-1}} = 0$ since $f \in \mathcal{H}^{(q)}(\mathbb{S}^{d-1})$, thus, as shown in Theorem 13, Algorithm 1 can recover f "exactly" using $s = \mathcal{O}(\beta_{q,d} \log \beta_{q,d})$ evaluations, where $\beta_{q,d}$ is the dimension of the Hilbert space $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$.

We predict f's value on a random test set on $\mathbb{S}^{d-1}$ and consider the algorithm fails if the testing error is greater than 10^{-12}. We count the number of failures among 100 independent random trials with different choices of $d \in \{3, 4\}$, $q \in \{5, \ldots, 22\}$, and $s \in \{40, \ldots, 2400\}$. The empirical success probabilities for $d = 3$ and 4 are reported in Fig. 1(a) and Fig. 1(b), respectively.

Fig. 1 illustrates that the success probabilities of our algorithm sharply transition to 1 as soon as the number of samples approaches $s \approx \beta_{q,d}$ for a wide range of q and both $d = 3, 4$. These experimental results complement our Theorem 13 along with the lower bound analysis in Section 4 and verify the empirical performance of our algorithm.

Acknowledgements

Haim Avron was partially supported by the Israel Science Foundation (grant no. 1272/17) and by the US-Israel Binational Science Foundation (grant no. 2017698). Amir Zandieh was supported by the Swiss NSF grant No. P2ELP2_195140. Insu Han was supported by TATA DATA Analysis (grant no. 105676).

References

[AH12] Kendall Atkinson and Weimin Han. *Spherical harmonics and approximations on the unit sphere: an introduction*. Springer Science & Business Media, 2012.

[AKM+17] Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. Random Fourier features for kernel ridge regression: Approximation bounds and statistical guarantees. In *International Conference on Machine Learning (ICML)*, 2017.

[AKM+19] Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. A universal sampling method for reconstructing signals with simple fourier transforms. In *Symposium on the Theory of Computing (STOC)*, 2019.

[CKPS16] Xue Chen, Daniel M Kane, Eric Price, and Zhao Song. Fourier-sparse interpolation without a frequency gap. In *Foundations of Computer Science (FOCS)*, 2016.

[CLL15] Pui Tung Choi, Ka Chun Lam, and Lok Ming Lui. FLASH: Fast landmark aligned spherical harmonic parameterization for genus-0 closed brain surfaces. *SIAM Journal on Imaging Sciences*, 2015.

[EMM20] Tamás Erdélyi, Cameron Musco, and Christopher Musco. Fourier sparse leverage scores and approximate kernel learning. *Neural Information Processing Systems (NeurIPS)*, 2020.

[KFR03] Michael Kazhdan, Thomas Funkhouser, and Szymon Rusinkiewicz. Rotation invariant spherical harmonic representation of 3 d shape descriptors. In *Symposium on Geometry Processing*, 2003.

[KKS97] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. Statistics of cosmic microwave background polarization. *Physical Review D*, 1997.

[Lan12] Serge Lang. *Real and functional analysis*. Springer Science & Business Media, 2012.

[LMP13] Mu Li, Gary L Miller, and Richard Peng. Iterative row sampling. In *Foundations of Computer Science (FOCS)*, 2013.

[LP61] Henry J Landau and Henry O Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty?II. *Bell System Technical Journal*, 1961.

[LP62] Henry J Landau and Henry O Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty?III: the dimension of the space of essentially time-and band-limited signals. *Bell System Technical Journal*, 1962.

[MNY06] Ha Quang Minh, Partha Niyogi, and Yuan Yao. Mercer?s theorem, feature maps, and smoothing. In *Conference on Learning Theory (COLT)*, 2006.

[Mor98] Mitsuo Morimoto. *Analytic functionals on the sphere*. American Mathematical Soc., 1998.

[Pen30] WO Pennell. A generalized Fourier series representation of a function. *The American Mathematical Monthly*, 1930.

[RT06] Vladimir Rokhlin and Mark Tygert. Fast algorithms for spherical harmonic expansions. *SIAM Journal on Scientific Computing (SISC)*, 2006.

[SA22] Paz Fink Shustin and Haim Avron. Semi-Infinite Linear Regression and Its Applications. *SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 2022.

[SP61] David Slepian and Henry O Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty?I. *Bell System Technical Journal*, 1961.

[SS00] Paul N Swarztrauber and William F Spotz. Generalized discrete spherical harmonic transforms. *Journal of Computational Physics*, 2000.

[ST02] Reiji Suda and Masayasu Takami. A fast spherical harmonics transform algorithm. *Mathematics of Computation*, 2002.

[Wei95] DR Weimer. Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients. *Journal of Geophysical Research: Space Physics*, 1995.

[Wer97] Robert A Werner. Spherical harmonic coefficients for the potential of a constant-density polyhedron. *Computers & Geosciences*, 1997.

[Wil12] Virginia Vassilevska Williams. Multiplying matrices faster than Coppersmith-Winograd. In *Symposium on the Theory of Computing (STOC)*, 2012.

[XRY01] Hong Xiao, Vladimir Rokhlin, and Norman Yarvin. Prolate spheroidal wavefunctions, quadrature and interpolation. *Inverse problems*, 2001.

A Properties of Gegenbauer Polynomials and Spherical Harmonics

In this section we prove the basic properties of the Gegenbauer Polynomials as well as the Spherical Harmonics and establish the connection between the two. We start by the direct sum decomposition of the Hilbert space $L^2(\mathbb{S}^{d-1})$ in terms of the spherical harmonics,

Lemma 6 (Direct Sum Decomposition of $L^2(\mathbb{S}^{d-1})$). *The family of spaces $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ yields a Hilbert space direct sum decomposition $L^2\left(\mathbb{S}^{d-1}\right) = \bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$: the summands are closed and pairwise orthogonal, and that every $f \in L^2\left(\mathbb{S}^{d-1}\right)$ is the sum of a converging series (in the sense of mean-square convergence with the L^2-norm defined in Eq. (2)),*

$$f = \sum_{\ell=0}^{\infty} f_\ell,$$

where $f_\ell \in \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ are uniquely determined functions. Furthermore, given any orthonormal basis $\left\{y_1^\ell, y_2^\ell, \dots, y_{\alpha_{\ell,d}}^\ell\right\}$ of $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ we have $f_\ell = \sum_{j=1}^{\alpha_{\ell,d}} \langle f, y_j^\ell \rangle_{\mathbb{S}^{d-1}} \cdot y_j^\ell$.

Proof. This is in fact a standard result. For example, see [Lan12] for a proof. □

Now we show that the Gegenbauer polynomials and spherical harmonics are related through the so called addition theorem,

Theorem 5 (Addition Theorem). *For every integer $\ell \geq 0$, if $\left\{y_1^\ell, y_2^\ell, \dots, y_{\alpha_{\ell,d}}^\ell\right\}$ is an orthonormal basis for $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$, then for any $\sigma, w \in \mathbb{S}^{d-1}$ we have*

$$\frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} \cdot P_d^\ell\left(\langle \sigma, w \rangle\right) = \sum_{j=1}^{\alpha_{\ell,d}} y_j^\ell(\sigma) \cdot y_j^\ell(w).$$

Proof. The result can be proven analytically, using the properties of the Poisson kernel in the unit ball. This is classic and the proof can be found in [AH12, Theorem 2.9]. □

Next we show that the Gegenbauer kernels can project any function into the space of their corresponding spherical harmonics,

Lemma 7 (Reproducing Kernel of $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$). *For every $f \in L^2\left(\mathbb{S}^{d-1}\right)$, if $f = \sum_{\ell=0}^{\infty} f_\ell$ is the unique decomposition of f over $\bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ as per Lemma 6, then f_ℓ is given by*

$$f_\ell(\sigma) = \alpha_{\ell,d} \cdot \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[f(w) P_d^\ell\left(\langle \sigma, w \rangle\right)\right] \quad \text{for } \sigma \in \mathbb{S}^{d-1}.$$

Proof. This is a classic textbook result, see [Mor98]. □

Now we prove that the Gegenbauer kernels satisfy the reproducing property for the Hilbert space $\mathcal{H}_\ell(\mathbb{S}^{d-1})$.

Lemma 4 (Reproducing Property of Zonal Harmonics). *Let $P_d^\ell(\cdot)$ be the Gengenbauer polynomial of degree ℓ in dimension d. For any $x, y \in \mathbb{S}^{d-1}$:*

$$P_d^\ell(\langle x,y\rangle) = \alpha_{\ell,d} \cdot \mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle x,w\rangle\right)P_d^\ell\left(\langle y,w\rangle\right)\right],$$

Furthermore, for any $\ell' \neq \ell$:

$$\mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle x,w\rangle\right)\cdot P_d^{\ell'}\left(\langle y,w\rangle\right)\right] = 0.$$

Proof. This result follows directly from the Funk–Hecke formula (See [AH12]). However, we provide another proof here. First note that for every $x \in \mathbb{S}^{d-1}$ the function $P_d^\ell\left(\langle x,\cdot\rangle\right) \in \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$. Therefore the first claim follow by applying Lemma 7 on function $f(\sigma) = P_d^\ell\left(\langle x,\sigma\rangle\right)$ which also satisfies $f_\ell = f$. On the other hand, $P_d^{\ell'}\left(\langle y,\cdot\rangle\right) \in \mathcal{H}_{\ell'}\left(\mathbb{S}^{d-1}\right)$ for every $y \in \mathbb{S}^{d-1}$. Thus, for $\ell' \neq \ell$, using the fact that spherical harmonics are orthogonal spaces of functions, $P_d^{\ell'}\left(\langle y,\cdot\rangle\right) \perp \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$, which gives the second claim.

□

Next we prove that the kernel operator defined in Definition 8 is in fact a projection operator,

Claim 1. *The operator $\mathcal{K}_d^{(q)}$ defined in Definition 8 satisfies the property $\left(\mathcal{K}_d^{(q)}\right)^2 = \mathcal{K}_d^{(q)}$.*

Proof. For every $f \in L^2\left(\mathbb{S}^{d-1}\right)$ and every $\sigma \in \mathbb{S}^{d-1}$, using Definition 8 we have,

$$\begin{aligned}
\left[\left(\mathcal{K}_d^{(q)}\right)^2 f\right](\sigma) &= \sum_{\ell'=0}^{q} \frac{\alpha_{\ell',d}}{|\mathbb{S}^{d-1}|}\left\langle \mathcal{K}_d^{(q)} f, P_d^{\ell'}\left(\langle\sigma,\cdot\rangle\right)\right\rangle_{\mathbb{S}^{d-1}} \\
&= \sum_{\ell=0}^{q}\sum_{\ell'=0}^{q} \alpha_{\ell,d}\alpha_{\ell',d}\cdot\mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^{\ell'}\left(\langle\sigma,w\rangle\right)\cdot\mathbb{E}_{\tau\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle\tau,w\rangle\right)f(\tau)\right]\right] \\
&= \sum_{\ell=0}^{q}\sum_{\ell'=0}^{q} \alpha_{\ell,d}\alpha_{\ell',d}\cdot\mathbb{E}_{\tau\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[f(\tau)\cdot\mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^{\ell'}\left(\langle\sigma,w\rangle\right)P_d^\ell\left(\langle\tau,w\rangle\right)\right]\right] \\
&= \sum_{\ell=0}^{q} \alpha_{\ell,d}\cdot\mathbb{E}_{\tau\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[f(\tau)\cdot P_d^\ell\left(\langle\sigma,\tau\rangle\right)\right] \\
&= \left[\mathcal{K}_d^{(q)} f\right](\sigma),
\end{aligned}$$

where the fourth line above follows from Lemma 4. This proves the claim.

□

B Reducing the Interpolation Problem to a Least-Squares Regression

In this section we show that our spherical harmonic interpolation problem, i.e., Problem 2, can be solved by approximately solving a least-squares problem as claimed in Claim 2. We start by showing that for any function $f \in L^2(\mathbb{S}^{d-1})$, $\mathcal{K}_d^{(q)} f$ gives its low-degree component. More precisely, let $f = \sum_{\ell=0}^{\infty} f_\ell$ the the decomposition of f over the Hilbert sum $\bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$ as per Lemma 6. Now if we let $\mathcal{K}_d^{(q)}$ be the kernel operator from Definition 8 and if $\left\{y_1^\ell, y_2^\ell, \ldots, y_{\alpha_{\ell,d}}^\ell\right\}$ is an orthonormal basis for $\mathcal{H}_\ell\left(\mathbb{S}^{d-1}\right)$, then by Theorem 5 we have,

$$\begin{aligned}
\left[\mathcal{K}_d^{(q)} f\right](\sigma) &= \sum_{\ell=0}^{q} \alpha_{\ell,d} \cdot \mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[f(w) P_d^\ell\left(\langle\sigma, w\rangle\right)\right] \\
&= \sum_{\ell=0}^{q} \left|\mathbb{S}^{d-1}\right| \cdot \mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[f(w)\cdot\sum_{j=1}^{\alpha_{\ell,d}} y_j^\ell(\sigma)\cdot y_j^\ell(w)\right] \\
&= \sum_{\ell=0}^{q}\sum_{j=1}^{\alpha_{\ell,d}} y_j^\ell(\sigma)\cdot\left|\mathbb{S}^{d-1}\right|\cdot\mathbb{E}_{w\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[f(w)\cdot y_j^\ell(w)\right] \\
&= \sum_{\ell=0}^{q}\sum_{j=1}^{\alpha_{\ell,d}} \langle f, y_j^\ell(w)\rangle_{\mathbb{S}^{d-1}}\cdot y_j^\ell(\sigma) \\
&= \sum_{\ell=0}^{q} f_\ell(\sigma) = f^{(q)}(\sigma),
\end{aligned}$$

where the the second line above follows from Theorem 5, the fourth line follows from Eq. (2), and the last line follows from Lemma 6. This proves that the low-degree component $f^{(q)} = \mathcal{K}_d^{(q)} f$.

Claim 2. *For any function $f \in L^2\left(\mathbb{S}^{d-1}\right)$, any integer $q \ge 0$, and any $C \ge 1$, if $\tilde{g} \in L^2\left(\mathbb{S}^{d-1}\right)$ is a function that satisfies,*

$$\left\|\mathcal{K}_d^{(q)}\tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 \le C \cdot \min_{g\in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2,$$

and if we let $f^{(q)} := \mathcal{K}_d^{(q)} f$, where $\mathcal{K}_d^{(q)}$ is defined as per Definition 8, then the following holds

$$\left\|\mathcal{K}_d^{(q)}\tilde{g} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \le (C-1)\cdot\left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2.$$

Proof. First, note that $g^* = f$ is an optimal solution to the least-squares problem in Eq. (7). Thus we have,

$$\min_{g\in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 = \left\|\mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 = \left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2.$$

Next, we can write,

$$\begin{aligned}
\left\|\mathcal{K}_d^{(q)}\tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 &= \left\|\mathcal{K}_d^{(q)}\tilde{g} - \mathcal{K}_d^{(q)} f + \left(\mathcal{K}_d^{(q)} f - f\right)\right\|_{\mathbb{S}^{d-1}}^2 \\
&= \left\|\mathcal{K}_d^{(q)}(\tilde{g} - f) + \left(\mathcal{K}_d^{(q)} f - f\right)\right\|_{\mathbb{S}^{d-1}}^2 \\
&= \left\|\mathcal{K}_d^{(q)}(\tilde{g} - f)\right\|_{\mathbb{S}^{d-1}}^2 + \left\|\mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 \\
&= \left\|\mathcal{K}_d^{(q)}\tilde{g} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 + \left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2,
\end{aligned}$$

where the third line follows from the Pythagorean theorem because $\mathcal{K}_d^{(q)}(\tilde{g} - f) \in \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$ while $\mathcal{K}_d^{(q)} f - f = -\sum_{\ell > q} f_\ell$, thus $\left(\mathcal{K}_d^{(q)} f - f\right) \perp \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$. Combining the two equalities above with inequality $\left\|\mathcal{K}_d^{(q)}\tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 \le C \cdot \min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2$ was given in the statement of the claim proves the Claim 2.

□

C Approximate Regression via Leverage Score Sampling

In this section we ultimately prove our main result of Theorem 13. We start by proving useful properties of the leverage function given in Definition 9. First, we show the fact that the ridge leverage function can be characterized in terms of a least-squares minimization problem, which is crucial for computing the leverage scores distribution. This fact was previously exploited in [AKM+17] and [AKM+19] in the context of Fourier operators.

Lemma 10 (Min Characterization of the Leverage Function). *For any $w \in \mathbb{S}^{d-1}$, let $\tau_q(w)$ be the leverage function (Definition 9) and define $\phi_w \in L^2(\mathbb{S}^{d-1})$ by $\phi_w(\sigma) \equiv \sum_{\ell=0}^q \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell\left(\langle \sigma, w\rangle\right)$. We have the following minimization characterization of the leverage function:*

$$\tau_q(w) = \left\{ \min_{g \in L^2(\mathbb{S}^{d-1})} \|g\|_{\mathbb{S}^{d-1}}^2, \quad \textit{s.t.} \quad \mathcal{K}_d^{(q)} g = \phi_w \right\}. \tag{9}$$

We remark that this lemma is in fact an adaptation and generalization of Theorem 5 of [AKM+19]. We prove this lemma here for the sake of completeness.

Proof. First we show that the right hand side of Eq. (9) is never smaller than the leverage function in Definition 9. Let $g_w^* \in L^2(\mathbb{S}^{d-1})$ be the optimal solution of Eq. (9) for any $w \in \mathbb{S}^{d-1}$. Note that the optimal solution satisfies $\mathcal{K}_d^{(q)} g_w^* = \phi_w$. Thus, for any function $f \in L^2(\mathbb{S}^{d-1})$, using Definition 8, we can write

$$\begin{aligned}
\left|\left[\mathcal{K}_d^{(q)} f\right](w)\right|^2 &= \left|\sum_{\ell=0}^q \alpha_{\ell,d} \cdot \mathbb{E}_{\sigma \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle \sigma, w\rangle\right) \cdot f(\sigma)\right]\right|^2 \\
&= \left|\langle \phi_w, f\rangle_{\mathbb{S}^{d-1}}\right|^2 = \left|\left\langle \mathcal{K}_d^{(q)} g_w^*, f\right\rangle_{\mathbb{S}^{d-1}}\right|^2 \\
&= \left|\left\langle g_w^*, \mathcal{K}_d^{(q)} f\right\rangle_{\mathbb{S}^{d-1}}\right|^2 && \text{(because } \mathcal{K}_d^{(q)} \text{ is self-adjoint)} \\
&\le \|g_w^*\|_{\mathbb{S}^{d-1}}^2 \cdot \left\|\mathcal{K}_d^{(q)} f\right\|_{\mathbb{S}^{d-1}}^2 && \text{(by Cauchy–Schwarz inequality)}
\end{aligned}$$

Therefore, for any $f \in L^2(\mathbb{S}^{d-1})$ with $\left\|\mathcal{K}_d^{(q)} f\right\|_{\mathbb{S}^{d-1}} > 0$, we have

$$\frac{\left|\left[\mathcal{K}_d^{(q)} f\right](w)\right|^2}{\left\|\mathcal{K}_d^{(q)} f\right\|_{\mathbb{S}^{d-1}}^2} \le \|g_w^*\|_{\mathbb{S}^{d-1}}^2. \tag{12}$$

We conclude the proof by showing that the maximum value is attained. First, we show that the optimal solution g_w^* of Eq. (9) satisfies the property that $\mathcal{K}_d^{(q)} g_w^* = g_w^*$. Suppose for the sake of contradiction that $\mathcal{K}_d^{(q)} g_w^* \neq g_w^*$. In this case, Claim 1 implies that,

$$\mathcal{K}_d^{(q)} \left(\mathcal{K}_d^{(q)} g_w^* - g_w^* \right) = \left(\mathcal{K}_d^{(q)} \right)^2 g_w^* - \mathcal{K}_d^{(q)} g_w^* = \mathcal{K}_d^{(q)} g_w^* - \mathcal{K}_d^{(q)} g_w^* = 0.$$

Thus, the function $g = \mathcal{K}_d^{(q)} g_w^*$ satisfies the constraint of the minimization problem in Eq. (9). Now, using the above and the fact that $\mathcal{K}_d^{(q)}$ is self-adjoint we can write,

$$\left\langle \mathcal{K}_d^{(q)} g_w^*, \mathcal{K}_d^{(q)} g_w^* - g_w^* \right\rangle_{\mathbb{S}^{d-1}} = \left\langle g_w^*, \mathcal{K}_d^{(q)} \left(\mathcal{K}_d^{(q)} g_w^* - g_w^* \right) \right\rangle_{\mathbb{S}^{d-1}} = 0.$$

This shows that $\mathcal{K}_d^{(q)} g_w^* \perp \left(\mathcal{K}_d^{(q)} g_w^* - g_w^* \right)$, hence by Pythagorean theorem we have,

$$\|g_w^*\|_{\mathbb{S}^{d-1}}^2 = \left\| \mathcal{K}_d^{(q)} g_w^* \right\|_{\mathbb{S}^{d-1}}^2 + \left\| \mathcal{K}_d^{(q)} g_w^* - g_w^* \right\|_{\mathbb{S}^{d-1}}^2 > \left\| \mathcal{K}_d^{(q)} g_w^* \right\|_{\mathbb{S}^{d-1}}^2 = \|g\|_{\mathbb{S}^{d-1}}^2 ,$$

which is in contrast with the assumption that g_w^* is the optimal solution of Eq. (9). Therefore, our claim that $\mathcal{K}_d^{(\ell)} g_w^* = g_w^*$ holds.

Now, we show that for $f = g_w^*$, the maximum value in inequality Eq. (12) is attained. For any $w \in \mathbb{S}^{d-1}$ we have the following

$$\left[\mathcal{K}_d^{(q)} f \right](w) = \left\langle \mathcal{K}_d^{(q)} g_w^*, f \right\rangle_{\mathbb{S}^{d-1}} = \langle g_w^*, g_w^* \rangle_{\mathbb{S}^{d-1}} = \|g_w^*\|_{\mathbb{S}^{d-1}}^2.$$

On the other hand we have $\left\| \mathcal{K}_d^{(q)} f \right\|_{\mathbb{S}^{d-1}}^2 = \|g_w^*\|_{\mathbb{S}^{d-1}}^2$. Thus, $\left\| \mathcal{K}_d^{(q)} f \right\|_{\mathbb{S}^{d-1}}^{-2} \cdot \left| \left[\mathcal{K}_d^{(q)} f \right](w) \right|^2 = \|g_w^*\|_{\mathbb{S}^{d-1}}^2$ which implies that $\tau_q(w) = \|g_w^*\|_{\mathbb{S}^{d-1}}^2$ and thus proves the lemma.

□

To prove our Theorem 13, we need to use some results about concentration of random operators. In particular we use Lemma 37 from [AKM+19], which is restated bellow,

Lemma 17 (Lemma 37 of [AKM+19]). *Suppose that $\mathcal{H}$ is a separable Hilbert space, and that $\mathcal{B}$ is a fixed self-adjoint Hilbert-Schmidt operator on $\mathcal{H}$. Let $\mathcal{R}$ be a self-adjoint Hilbert-Schmidt random operator that satisfies*

$$\mathbb{E}[\mathcal{R}] = \mathcal{B}, \text{ and } \|\mathcal{R}\|_{op} \leq L$$

Let $\mathcal{M}$ be another self-adjoint trace-class operator such that $\mathbb{E}[\mathcal{R}^2] \preceq \mathcal{M}$. Form the operator sampling estimator

$$\bar{\mathcal{R}}_n := \frac{1}{n} \sum_{k=1}^{n} \mathcal{R}_k$$

where each $\mathcal{R}_k$ is an independent copy of $\mathcal{R}$. Then, for any $t > \sqrt{\|\mathcal{M}\|_{op}/n} + 2L/3n$,

$$\Pr[\|\bar{\mathcal{R}}_n - \mathcal{B}\|_{op} > t] \leq \frac{8 \cdot \text{trace}(\mathcal{M})}{\|\mathcal{M}\|_{op}} \cdot \exp\left(\frac{-nt^2/2}{\|\mathcal{M}\|_{op} + 2Lt/3} \right).$$

Our approach is to apply Lemma 17 to show that the operator $\mathcal{K}_d^{(q)}$ can be well approximated by $\boldsymbol{P}\boldsymbol{P}^*$, where the quasi-matrix $\boldsymbol{P}$ is defined in Theorem 13. In order to prove this formally, we need to define the notion of positive definiteness for self-adjoint operators. We call self-adjoint $A : L^2(\mathbb{S}^{d-1}) \to L^2(\mathbb{S}^{d-1})$ positive semidefinite (or simply positive) and write $A \succeq 0$ if $\langle x, Ax\rangle_{\mathbb{S}^{d-1}} \geq 0$ for all $x \in L^2(\mathbb{S}^{d-1})$. The notation for $A \preceq B$ and $A \succeq B$ follow in the standard way. Now with the notations in place we can prove the following lemma,

Lemma 18 (Approximating $\mathcal{K}_d^{(q)}$ via Leverage Score Sampling). *For any $\delta > 0$ and $\epsilon \in (0, 1/2)$, let $s = \frac{8\beta_{q,d}}{3\epsilon^2} \log \frac{8\beta_{q,d}}{\delta}$, for sufficiently large fixed constant c, and let $w_1, w_2, \ldots, w_s$ be i.i.d. uniform samples from $\mathbb{S}^{d-1}$. Let $\boldsymbol{P} : \mathbb{R}^s \to L^2(\mathbb{S}^{d-1})$ be the quasi-matrix defined as follows, for every $v \in \mathbb{R}^d$ and $\sigma \in \mathbb{S}^{d-1}$:*

$$[\boldsymbol{P} \cdot v](\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s \cdot |\mathbb{S}^{d-1}|}} \cdot \sum_{j=1}^{s} v_j \cdot P_d^\ell \left(\langle w_j, \sigma\rangle\right).$$

Also let $\boldsymbol{P}^*$ be the adjoint of $\boldsymbol{P}$. Then with probability at least $1 - \delta$,

$$(1 - \varepsilon) \cdot \mathcal{K}_d^{(q)} \preceq \boldsymbol{P}\boldsymbol{P}^* \preceq (1 + \varepsilon) \cdot \mathcal{K}_d^{(q)}.$$

Proof. The proof is by invoking Lemma 17. The reason we can invoke this lemma is because $\mathcal{K}_d^{(q)}$ is a self adjoint trace-class orthonormal projection operator, by Claim 1 and Eq. (6), thus this operator is Hilbert-Schmidt. Furthermore, the Hilbert space that we care about is $L^2(\mathbb{S}^{d-1})$ which is a separable space.

Now notice that if we define the function $\phi_w \in L^2(\mathbb{S}^{d-1})$ by $\phi_w(\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell \left(\langle \sigma, w\rangle\right)$ for $\sigma, w \in \mathbb{S}^{d-1}$, then for any $v \in \mathbb{R}^s$:

$$\boldsymbol{P} \cdot v \equiv \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot \sum_{j=1}^{s} v_j \cdot \phi_{w_j}$$

Furthermore, given functions $f, g \in L^2(\mathbb{S}^{d-1})$ we define the operator $(f \otimes g) : L^2(\mathbb{S}^{d-1}) \to L^2(\mathbb{S}^{d-1})$ by

$$(f \otimes g)h := \langle g, h\rangle_{\mathbb{S}^{d-1}} \cdot f \qquad \text{for any } h \in L^2(\mathbb{S}^{d-1}).$$

Therefore, using this notation, if we let

$$\mathcal{R}_j := |\mathbb{S}^{d-1}| \cdot (\phi_{w_j} \otimes \phi_{w_j}),$$

then we understand that,

$$\boldsymbol{P}\boldsymbol{P}^* \equiv \frac{1}{s} \cdot \sum_{j=1}^{s} \mathcal{R}_j.$$

Note that $\mathcal{R}_j$ is a rank-one self adjoint operator, thus it is also Hilbert-Schmidt. Since samples $w_1, w_2, \ldots, w_s$ are drawn independently at random, $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_s$ are i.i.d. random operators with expectation,

$$\mathbb{E}[\mathcal{R}_j] = \left|\mathbb{S}^{d-1}\right| \cdot \mathbb{E}_{w_j \sim \mathcal{U}(\mathbb{S}^{d-1})}[\phi_{w_j} \otimes \phi_{w_j}] = \mathcal{K}_d^{(q)}.$$

The reason for the second equality above is that for any function $f \in L^2(\mathbb{S}^{d-1})$ and any $\sigma \in \mathbb{S}^{d-1}$:

$$\begin{aligned}
\left|\mathbb{S}^{d-1}\right| \left[\mathbb{E}[\phi_{w_j} \otimes \phi_{w_j}] f\right](\sigma) &= \left|\mathbb{S}^{d-1}\right| \cdot \mathbb{E}_{w_j}\left[[(\phi_{w_j} \otimes \phi_{w_j}) \cdot f](\sigma)\right] \\
&= \left|\mathbb{S}^{d-1}\right| \cdot \mathbb{E}_{w_j}\left[\langle \phi_{w_j}, f\rangle_{\mathbb{S}^{d-1}} \cdot \phi_{w_j}(\sigma)\right] \\
&= \sum_{\ell=0}^{q} \sum_{\ell'=0}^{q} \alpha_{\ell,d}\alpha_{\ell',d} \cdot \mathbb{E}_w\left[\mathbb{E}_\tau\left[P_d^\ell(\langle \sigma, w\rangle) P_d^{\ell'}(\langle \tau, w\rangle) f(\tau)\right]\right] \\
&= \sum_{\ell=0}^{q} \alpha_{\ell,d} \cdot \mathbb{E}_{\tau\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell(\langle \sigma, \tau\rangle) f(\tau)\right] \\
&= \left[\mathcal{K}_d^{(q)} f\right](\sigma),
\end{aligned}$$

where the fourth line above follows from Lemma 4. Next, we bound the operator norm of $\mathcal{R}_j$. This random operator only takes values that are both positive semi-definite and rank one, so the operator norm of $\mathcal{R}_j$ is equal to the following

$$\begin{aligned}
\|\mathcal{R}_j\|_{op} &= \left\|\left|\mathbb{S}^{d-1}\right| \cdot (\phi_{w_j} \otimes \phi_{w_j})\right\|_{op} \\
&= \left|\mathbb{S}^{d-1}\right| \cdot \left\|\phi_{w_j}\right\|_{\mathbb{S}^{d-1}}^2 \\
&= \beta_{q,d},
\end{aligned}$$

where the last line follows from Lemma 4 and definition of ϕ_{w_j} as well as the fact that $\beta_{q,d} = \sum_{\ell=0}^{q} \alpha_{\ell,d}$. The final ingredient for applying Lemma 17 is to bound $\mathcal{R}_j^2$. We have,

$$\begin{aligned}
\mathcal{R}_j^2 &= \left|\mathbb{S}^{d-1}\right|^2 \cdot (\phi_{w_j} \otimes \phi_{w_j}) \cdot (\phi_{w_j} \otimes \phi_{w_j}) \\
&= \left|\mathbb{S}^{d-1}\right|^2 \cdot \left\|\phi_{w_j}\right\|_{\mathbb{S}^{d-1}}^2 \cdot (\phi_{w_j} \otimes \phi_{w_j}) \\
&= \beta_{q,d} \cdot \left|\mathbb{S}^{d-1}\right| \cdot (\phi_{w_j} \otimes \phi_{w_j}) \\
&= \beta_{q,d} \cdot \mathcal{R}_j.
\end{aligned}$$

Therefore,

$$\mathbb{E}[\mathcal{R}_j^2] = \beta_{q,d} \cdot \mathcal{K}_d^{(q)} =: \mathcal{M}.$$

Now note that by Eq. (6), we have $\operatorname{trace}(\mathcal{M}) = \beta_{q,d} \cdot \operatorname{trace}\left(\mathcal{K}_d^{(q)}\right) = \beta_{q,d}^2$. Also, by Claim 1, $\mathcal{K}_d^{(q)}$ is an orthonormal projection operator, thus $\|\mathcal{M}\|_{op} = \beta_{q,d} \cdot \left\|\mathcal{K}_d^{(q)}\right\|_{op} = \beta_{q,d}$. Therefore, by Lemma 17 we have,

$$\begin{aligned}
\Pr\left[\left\|\boldsymbol{P}\boldsymbol{P}^* - \mathcal{K}_d^{(q)}\right\|_{op} > \epsilon\right] &\le \frac{8 \cdot \operatorname{trace}(\mathcal{M})}{\|\mathcal{M}\|_{op}} \cdot \exp\left(\frac{-s\epsilon^2/2}{\|\mathcal{M}\|_{op} + 2\beta_{q,d}\epsilon/3}\right) \\
&= \frac{8 \cdot \beta_{q,d}^2}{\beta_{q,d}} \cdot \exp\left(\frac{-s\epsilon^2/2}{\beta_{q,d} + 2\beta_{q,d}\epsilon/3}\right) \\
&\le \delta.
\end{aligned}$$

Now recall that $\mathcal{K}_d^{(q)}$ is an orthonormal projection matrix. We claim that the eigenspace of $\boldsymbol{P}\boldsymbol{P}^*$ is a subspace of the eigenspace of $\mathcal{K}_d^{(q)}$. To see why note that we can write,

$$\begin{aligned}
\mathcal{K}_d^{(q)} \cdot \boldsymbol{P}\boldsymbol{P}^* &= \mathcal{K}_d^{(q)} \cdot \left(\frac{\left|\mathbb{S}^{d-1}\right|}{s} \cdot \sum_{j=1}^{s} (\phi_{w_j} \otimes \phi_{w_j}) \right) \\
&= \frac{\left|\mathbb{S}^{d-1}\right|}{s} \cdot \sum_{j=1}^{s} \mathcal{K}_d^{(q)} \cdot (\phi_{w_j} \otimes \phi_{w_j}) \\
&= \frac{\left|\mathbb{S}^{d-1}\right|}{s} \cdot \sum_{j=1}^{s} \left(\left(\mathcal{K}_d^{(q)} \cdot \phi_{w_j} \right) \otimes \phi_{w_j} \right) \\
&= \frac{\left|\mathbb{S}^{d-1}\right|}{s} \cdot \sum_{j=1}^{s} \left(\phi_{w_j} \otimes \phi_{w_j} \right) \\
&= \boldsymbol{P}\boldsymbol{P}^*, \qquad (13)
\end{aligned}$$

where the fourth line above follows because for any $\sigma, w_j \in \mathbb{S}^{d-1}$,

$$\begin{aligned}
\left[\mathcal{K}_d^{(q)} \cdot \phi_{w_j} \right](\sigma) &= \sum_{\ell=0}^{q} \alpha_{\ell,d} \cdot \mathbb{E}_{v \sim \mathcal{U}(\mathbb{S}^{d-1})} \left[P_d^{\ell}\left(\langle \sigma, v \rangle\right) \cdot \phi_{w_j}(v) \right] \\
&= \sum_{\ell=0}^{q} \sum_{\ell'=0}^{q} \frac{\alpha_{\ell,d}\alpha_{\ell',d}}{|\mathbb{S}^{d-1}|} \cdot \mathbb{E}_{v \sim \mathcal{U}(\mathbb{S}^{d-1})} \left[P_d^{\ell}\left(\langle \sigma, v \rangle\right) \cdot P_d^{\ell'}\left(\langle w_j, v \rangle\right) \right] \\
&= \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^{\ell}\left(\langle \sigma, w_j \rangle\right) \\
&= \phi_{w_j}(\sigma),
\end{aligned}$$

where the third line above follows from Lemma 4. Therefore, now we have shown that $\mathcal{K}_d^{(q)} \cdot \boldsymbol{P}\boldsymbol{P}^* = \boldsymbol{P}\boldsymbol{P}^*$ and $\left\| \boldsymbol{P}\boldsymbol{P}^* - \mathcal{K}_d^{(q)} \right\|_{op} \leq \epsilon$. Given the fact that $\mathcal{K}_d^{(q)}$ is a symmetric self-adjoint orthonormal projection and $\boldsymbol{P}\boldsymbol{P}^*$ is also symmetric and self-adjoint, this implies that,

$$\Pr\left[(1-\epsilon) \cdot \mathcal{K}_d^{(q)} \preceq \boldsymbol{P}\boldsymbol{P}^* \preceq (1+\epsilon) \cdot \mathcal{K}_d^{(q)} \right] \geq 1 - \delta$$

which completes the proof.

□

Now we are ready to prove Theorem 13. We prove this theorem by showing that for all $g \in L^2(\mathbb{S}^{d-1})$, leverage function sampling lets us approximate the value of the regression objective function in Eq. (7) when evaluated at g. We do this by showing that our sampling provides the so-called *affine embedding guarantee*.

Theorem 13 (Approximate Regression via Leverage Function Sampling). *For any $\epsilon, \delta > 0$, let $s = c \cdot \frac{\beta_{q,d}}{\epsilon^2}\left(\log \beta_{q,d} + \delta^{-1}\right)$, for sufficiently large fixed constant c, and let $w_1, w_2, \ldots, w_s$ be i.i.d. uniform* *samples on $\mathbb{S}^{d-1}$. Define the quasi-matrix $\boldsymbol{P}: \mathbb{R}^s \to L^2(\mathbb{S}^{d-1})$ as follows, for every $v \in \mathbb{R}^d$:*

$$[\boldsymbol{P} \cdot v](\sigma) := \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s \cdot |\mathbb{S}^{d-1}|}} \cdot \sum_{j=1}^{s} v_j \cdot P_d^{\ell}\left(\langle w_j, \sigma\rangle\right) \quad \textit{for } \sigma \in \mathbb{S}^{d-1}.$$

Also let $\boldsymbol{f} \in \mathbb{R}^s$ be a vector with $\boldsymbol{f}_j := \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot f(w_j)$ for $j = 1, 2, \ldots, s$ and let $\boldsymbol{P}^*$ be the adjoint of $\boldsymbol{P}$. If $\tilde{g}$ is an optimal solution to the following least-squares problem

$$\tilde{g} \in \arg\min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2,$$

then with probability at least $1-\delta$ the following holds,

$$\left\|\mathcal{K}_d^{(q)} \tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 \le (1+\epsilon) \cdot \min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2.$$

Proof. Throughout the proof we use $f^{(q)} := \mathcal{K}_d^{(q)} f$ and $B^* := \left\|f - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2$. The proof is by reduction to affine embedding. Specifically, we prove that, with probability at least $1-\delta$, simultaneously for all $g \in L^2(\mathbb{S}^{d-1})$,

$$(1-\epsilon/3) \cdot \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 \le \|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 + C \le (1+\epsilon/3) \cdot \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2, \tag{14}$$

where C is some fixed value independent of g that only depends on $\mathcal{K}_d^{(q)}$, $\boldsymbol{P}$, $\boldsymbol{f}$, and f. First we show that if we can prove Eq. (14), then the theorem immediately follows. To see why, note that for any $\tilde{g} \in \arg\min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2$ we can write,

$$\begin{aligned}
\left\|\mathcal{K}_d^{(q)} \tilde{g} - f\right\|_{\mathbb{S}^{d-1}}^2 &\le (1-\epsilon/3)^{-1} \left(\|\boldsymbol{P}^* \tilde{g} - \boldsymbol{f}\|_2^2 + C\right) && \text{(By Eq. (14))}\\
&= (1-\epsilon/3)^{-1} \left(\min_{g \in L^2\left(\mathbb{S}^{d-1}\right)} \|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 + C\right)\\
&\le (1-\epsilon/3)^{-1} \left(\|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 + C\right)\\
&\le \frac{1+\epsilon/3}{1-\epsilon/3} \cdot \left\|\mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 && \text{(By Eq. (14))}\\
&\le (1+\epsilon) \cdot \min_{g \in L^2(\mathbb{S}^{d-1})} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2,
\end{aligned}$$

where the last inequality follows because $f \in \arg\min_{g \in L^2(\mathbb{S}^{d-1})} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2$.

Thus, in order to prove the theorem it suffices to prove that the affine embedding property in Eq. (14) holds with probability at least $1-\delta$.

Expression for Least-Squares Excess Cost. We first show that the least-squares objective function in Eq. (7) can be written as a function of the deviation from the optimum $g - f$. More specifically, for any

$g \in L^2(\mathbb{S}^{d-1})$ we have,

$$\begin{aligned}\left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 &= \left\|\mathcal{K}_d^{(q)} g - \mathcal{K}_d^{(q)} f + \mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 \\ &= \left\|\mathcal{K}_d^{(q)} (g-f) + \mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 \\ &= \left\|\mathcal{K}_d^{(q)} (g-f)\right\|_{\mathbb{S}^{d-1}}^2 + \left\|\mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 \\ &= \left\|\mathcal{K}_d^{(q)} (g-f)\right\|_{\mathbb{S}^{d-1}}^2 + B^*, \end{aligned} \tag{15}$$

where the third line above follows from the Pythagorean theorem because $\mathcal{K}_d^{(q)}(g-f) \in \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$ while $\left(\mathcal{K}_d^{(q)} f - f\right) \perp \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$.

Bounding The Sampling Error. We now show that Eq. (15) holds approximately, even after sampling. This almost immediately yields the affine embedding bound of Eq. (14). We can write the discretized objective function value for any $g \in L^2(\mathbb{S}^{d-1})$ as,

$$\begin{aligned}\|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 &= \|\boldsymbol{P}^* g - \boldsymbol{P}^* f + \boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \\ &= \|\boldsymbol{P}^* (g-f) + \boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \\ &= \|\boldsymbol{P}^* (g-f)\|_2^2 + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 + 2\langle \boldsymbol{P}^*(g-f), \boldsymbol{P}^* f - \boldsymbol{f}\rangle. \end{aligned} \tag{16}$$

Let us focus on the last term above. First we show that $\mathcal{K}_d^{(q)} \cdot \boldsymbol{P} = \boldsymbol{P}$. For any $v \in \mathbb{R}^s$:

$$\begin{aligned}\left[\mathcal{K}_d^{(q)} \cdot \boldsymbol{P} \cdot v\right](\sigma) &= \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s \cdot |\mathbb{S}^{d-1}|}} \cdot \sum_{j=1}^{s} v_j \cdot \left[\mathcal{K}_d^{(q)} P_d^\ell\left(\langle w_j, \cdot\rangle\right)\right](\sigma) \\ &= \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s \cdot |\mathbb{S}^{d-1}|}} \cdot \sum_{j=1}^{s} v_j \cdot P_d^\ell\left(\langle w_j, \sigma\rangle\right) \\ &= [\boldsymbol{P} \cdot v](\sigma), \end{aligned}$$

where the second line follows from the definition of $\mathcal{K}_d^{(q)}$ in Definition 8 along with Lemma 4. Now using the fact that $\mathcal{K}_d^{(q)} \cdot \boldsymbol{P} = \boldsymbol{P}$, we can rewrite the last term as,

$$\begin{aligned}\langle \boldsymbol{P}^*(g-f), \boldsymbol{P}^* f - \boldsymbol{f}\rangle &= \langle g-f, \boldsymbol{P}(\boldsymbol{P}^* f - \boldsymbol{f})\rangle_{\mathbb{S}^{d-1}} && (\boldsymbol{P} \text{ is the adjoint of } \boldsymbol{P}^*) \\ &= \left\langle g-f, \mathcal{K}_d^{(q)} \cdot \boldsymbol{P}(\boldsymbol{P}^* f - \boldsymbol{f})\right\rangle_{\mathbb{S}^{d-1}} \\ &= \left\langle \mathcal{K}_d^{(q)}(g-f), \boldsymbol{P}(\boldsymbol{P}^* f - \boldsymbol{f})\right\rangle_{\mathbb{S}^{d-1}} && (\mathcal{K}_d^{(q)} \text{ is self-adjoint}) \end{aligned}$$

By plugging the above into Eq. (16) and applying Cauchy-Schwarz inequality we find that,

$$\|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 \in \|\boldsymbol{P}^*(g-f)\|_2^2 + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \pm 2\left\|\mathcal{K}_d^{(q)}(g-f)\right\|_{\mathbb{S}^{d-1}} \cdot \|\boldsymbol{P}(\boldsymbol{P}^* f - \boldsymbol{f})\|_{\mathbb{S}^{d-1}}. \tag{17}$$

Now we bound $\|\boldsymbol{P}(\boldsymbol{P}^* f - \boldsymbol{f})\|_{\mathbb{S}^{d-1}}$. We show that this quantity is small with probability at least $1 - \delta/2$, in the following claim,

Claim 5 (Approximate Operator Application). *With probability at least* $1-\delta/2$*:*

$$\|\boldsymbol{P}(\boldsymbol{P}^* f - \boldsymbol{f})\|_{\mathbb{S}^{d-1}} \le \frac{\epsilon}{18} \cdot \sqrt{B^*}.$$

We prove this claim later. Now by plugging the bound in Claim 5 into Eq. (17) we find that,

$$\begin{aligned}\|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 &\in \|\boldsymbol{P}^*(g-f)\|_2^2 + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \pm \frac{\epsilon}{9} \cdot \left\|\mathcal{K}_d^{(q)}(g-f)\right\|_{\mathbb{S}^{d-1}} \cdot \sqrt{B^*} \\ &\in \|\boldsymbol{P}^*(g-f)\|_2^2 + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \pm \frac{\epsilon}{18} \cdot \left\|\mathcal{K}_d^{(q)}(g-f)\right\|_{\mathbb{S}^{d-1}}^2 \pm \frac{\epsilon}{18} \cdot B^*,\end{aligned}$$

where the second line comes from the AM-GM inequality. Applying the operator approximation bound of Lemma 18 with error parameter $\epsilon/12$ and failure probability $\delta/2$ gives that the following holds simultaneously for all g, with probability at least $1-\delta$,

$$\|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 \in (1 \pm 5\epsilon/36) \cdot \left\|\mathcal{K}_d^{(q)}(g-f)\right\|_{\mathbb{S}^{d-1}}^2 + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \pm \frac{\epsilon}{18} \cdot B^*$$

Therefore, by plugging Eq. (15) into the above inequality we find that,

$$\begin{aligned}\|\boldsymbol{P}^* g - \boldsymbol{f}\|_2^2 &\in (1 \pm 5\epsilon/36) \cdot \left(\left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 - B^*\right) + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \pm \frac{\epsilon}{18} \cdot B^* \\ &\in (1 \pm 5\epsilon/36) \cdot \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 - (1 \pm 7\epsilon/36) \cdot B^* + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2 \\ &\in (1 \pm \epsilon/3) \cdot \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 - B^* + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2,\end{aligned}$$

where the last line above follows because $B^* = \left\|\mathcal{K}_d^{(q)} f - f\right\|_{\mathbb{S}^{d-1}}^2 \le \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2$ for any g. This shows that the affine embedding guarantee of Eq. (14) holds if we let $C := -B^* + \|\boldsymbol{P}^* f - \boldsymbol{f}\|_2^2$ which is a quantity that only depends on f, $\boldsymbol{f}$, $\boldsymbol{P}^*$, and $\mathcal{K}_d^{(q)}$ and is independent of g.

□

Now we prove Claim 5.

Proof of Claim 5: For conciseness we use $f^{(q)} := \mathcal{K}_d^{(q)} f$ and also define the function $\phi_w \in L^2(\mathbb{S}^{d-1})$ by $\phi_w(\sigma) := \sum_{\ell=0}^q \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^\ell(\langle \sigma, w\rangle)$ for $\sigma, w \in \mathbb{S}^{d-1}$. With this definition for any $v \in \mathbb{R}^s$:

$$\boldsymbol{P} \cdot v \equiv \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot \sum_{j=1}^s v_j \cdot \phi_{w_j}.$$

Furthermore, for any $f \in L^2(\mathbb{S}^{d-1})$ and any $j \in [s]$,

$$\begin{aligned}[\boldsymbol{P}^* f](j) &\equiv \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot \langle \phi_{w_j}, f\rangle_{\mathbb{S}^{d-1}} \\ &= \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \sum_{\ell=0}^q \alpha_{\ell,d} \cdot \mathbb{E}_{\sigma \sim \mathcal{U}(\mathbb{S}^{d-1})} \left[P_d^\ell(\langle \sigma, w_j\rangle) \cdot f(\sigma)\right] \\ &= \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot \left[\mathcal{K}_d^{(q)} f\right](w_j) \\ &= \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot f^{(q)}(w_j).\end{aligned}$$

Therefore, if we let $\boldsymbol{y} \in \mathbb{R}^s$ be the vector $\boldsymbol{y} := \boldsymbol{P}^* f - \boldsymbol{f}$, we have for any $j \in [s]$

$$\boldsymbol{y}(j) = \sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot \left(f^{(q)} - f\right)(w_j).$$

Additionally, for ease of notation let $y := f^{(q)} - f$. Thus we now focus on bounding $\|\boldsymbol{P} \cdot \boldsymbol{y}\|_{\mathbb{S}^{d-1}}^2$. We start by computing the expectation of this quantity with respect to $w_1, w_2, \ldots, w_s$,

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{P} \cdot \boldsymbol{y}\|_{\mathbb{S}^{d-1}}^2\right] &= \mathbb{E}\left[\left\|\sqrt{\frac{|\mathbb{S}^{d-1}|}{s}} \cdot \sum_{j=1}^{s} \phi_{w_j} \cdot \boldsymbol{y}(j)\right\|_{\mathbb{S}^{d-1}}^2\right] \\
&= \frac{|\mathbb{S}^{d-1}|^2}{s^2} \cdot \mathbb{E}\left[\left\|\sum_{j=1}^{s} \phi_{w_j} \cdot y(w_j)\right\|_{\mathbb{S}^{d-1}}^2\right] \\
&= \frac{|\mathbb{S}^{d-1}|^2}{s^2} \sum_{i,j \in [s]} \mathbb{E}_{w_i, w_j}\left[\langle \phi_{w_i}, \phi_{w_j} \rangle_{\mathbb{S}^{d-1}} \cdot y(w_i) y(w_j)\right] \\
&= \frac{|\mathbb{S}^{d-1}|^2}{s^2} \sum_{i \in [s]} \mathbb{E}_{w_i}\left[\|\phi_{w_i}\|_{\mathbb{S}^{d-1}}^2 \cdot y(w_i)^2\right] \qquad (18) \\
&\quad + \frac{|\mathbb{S}^{d-1}|^2}{s^2} \sum_{i \neq j \in [s]} \left\langle \mathbb{E}_{w_i}[y(w_i) \cdot \phi_{w_i}], \mathbb{E}_{w_j}[y(w_j) \cdot \phi_{w_j}] \right\rangle_{\mathbb{S}^{d-1}}, \qquad (19)
\end{aligned}$$

First we consider the term in Eq. (18). By Lemma 4 we can write,

$$\begin{aligned}
\frac{|\mathbb{S}^{d-1}|^2}{s^2} \sum_{i \in [s]} \mathbb{E}_{w_i}\left[\|\phi_{w_i}\|_{\mathbb{S}^{d-1}}^2 \cdot y(w_i)^2\right] &= \frac{|\mathbb{S}^{d-1}|^2}{s^2} \sum_{i \in [s]} \mathbb{E}_{w_i}\left[\frac{\beta_{q,d}}{|\mathbb{S}^{d-1}|} \cdot y(w_i)^2\right] \\
&= \frac{\beta_{q,d}}{s} \cdot \|y\|_{\mathbb{S}^{d-1}}^2.
\end{aligned}$$

Next we consider the term in Eq. (19). Using the definition of $y = f^{(q)} - f$, We show that for any $\sigma \in \mathbb{S}^{d-1}$ and any $i \in [s]$,

$$\begin{aligned}
\mathbb{E}_{w_i}[y(w_i)\phi_{w_i}(\sigma)] &= \mathbb{E}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})}\left[\sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{|\mathbb{S}^{d-1}|} P_d^{\ell}\left(\langle \sigma, w \rangle\right) y(w)\right] \\
&= \frac{1}{|\mathbb{S}^{d-1}|} \cdot \left[\mathcal{K}_d^{(q)} y\right](\sigma) \\
&= \left[\mathcal{K}_d^{(q)} (f^{(q)} - f)\right](\sigma) \\
&= \left[\mathcal{K}_d^{(q)} \left(\mathcal{K}_d^{(q)} f - f\right)\right](\sigma) \\
&= 0,
\end{aligned}$$

where the last line above follows from Claim 1. Thus,

$$\frac{|\mathbb{S}^{d-1}|^2}{s^2} \sum_{i \neq j \in [s]} \left\langle \mathbb{E}_{w_i}[y(w_i) \cdot \phi_{w_i}], \mathbb{E}_{w_j}[y(w_j) \cdot \phi_{w_j}] \right\rangle_{\mathbb{S}^{d-1}} = 0$$

By plugging these equalities into Eq. (18) and Eq. (19) we find that,

$$\mathbb{E}\left[\|\boldsymbol{P}\cdot\boldsymbol{y}\|_{\mathbb{S}^{d-1}}^2\right] = \frac{\beta_{q,d}}{s}\cdot\|y\|_{\mathbb{S}^{d-1}}^2 = \frac{\beta_{q,d}}{s}\cdot\|f^{(q)}-f\|_{\mathbb{S}^{d-1}}^2 = \frac{\beta_{q,d}}{s}\cdot B^*$$

Thus, by Markov's inequality and using the fact that $s=\Omega\left(\frac{\beta_{q,d}}{\epsilon^2\cdot\delta}\right)$, the claim follows. ■

D Efficient Algorithm for Spherical Harmonic Interpolation

In this section we prove our main theorem about our spherical harmonic interpolation algorithm.

Theorem 14 (Efficient Spherical Harmonic Interpolation). *Algorithm 1 returns a function $y\in\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ such that, with probability at least $1-\delta$:*

$$\left\|y-f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \le \epsilon\cdot\left\|f^{(q)}-f\right\|_{\mathbb{S}^{d-1}}^2,$$

where $f^{(q)} := \mathcal{K}_d^{(q)} f$. Suppose we can compute the Gegenbauer polynomial $P_d^\ell(t)$ at every point $t\in[-1,1]$ in constant time. Algorithm 1 queries the function f at $s=\mathcal{O}\left(\frac{\beta_{q,d}}{\epsilon^2}\left(\log\beta_{q,d}+\delta^{-1}\right)\right)$ points on the sphere $\mathbb{S}^{d-1}$ and runs in $\mathcal{O}(s^2\cdot d+s^\omega)$ time. This algorithm evaluates $y(\sigma)$ in $\mathcal{O}(d\cdot s)$ time for any $\sigma\in\mathbb{S}^{d-1}$.

Proof. First note that the random points $w_1,w_2,\ldots,w_s$ in line 3 of Algorithm 1 are i.i.d. sample with uniform distribution on the surface of $\mathbb{S}^{d-1}$. Therefore, we can invoke Theorem 13. More specifically, if we let $\boldsymbol{P}$ be the quasi-matrix defined in Theorem 13 corresponding to the random points $w_1,w_2,\ldots,w_s$ sampled in line 3 and if we let $\boldsymbol{f}$ be the vector of function samples defined in line 5 of the algorithm, then with probability at least $1-\delta$, any optimal solution to the following least-squares problem

$$\tilde{g}\in\arg\min_{g\in L^2(\mathbb{S}^{d-1})}\|\boldsymbol{P}^*g-\boldsymbol{f}\|_2^2, \tag{20}$$

satisfies the following,

$$\left\|\mathcal{K}_d^{(q)}\tilde{g}-f\right\|_{\mathbb{S}^{d-1}}^2 \le (1+\epsilon)\cdot\min_{g\in L^2(\mathbb{S}^{d-1})}\left\|\mathcal{K}_d^{(q)}g-f\right\|_{\mathbb{S}^{d-1}}^2. \tag{21}$$

Now note that the least-squares problem in Eq. (20) has at least one optimal solution $\tilde{g}$ which is in the eigenspace of the operator $\boldsymbol{P}\boldsymbol{P}^*$. More specifically, there exists a vector $\boldsymbol{z}\in\mathbb{R}^s$ such that $\tilde{g}=\boldsymbol{P}\cdot\boldsymbol{z}$ is an optimal solution for Eq. (20). Therefore, we can focus on finding this optimal solution by solving the following least-squares problem

$$\boldsymbol{z}\in\arg\min_{\boldsymbol{x}\in\mathbb{R}^s}\|\boldsymbol{P}^*\boldsymbol{P}\boldsymbol{x}-\boldsymbol{f}\|_2^2,$$

and then letting $\tilde{g}=\boldsymbol{P}\cdot\boldsymbol{z}$. This $\tilde{g}$ is guaranteed to be an optimal solution for Eq. (20), thus it satisfies Eq. (21). We solve the above least-squares problem using the kernel trick. In fact we show that $\boldsymbol{P}^*\boldsymbol{P}$ is equal to the kernel matrix $\boldsymbol{K}$ computed in line 4 of Algorithm 1. To see why, note that for any $i,j\in[s]$ we have,

$$
\begin{aligned}
[\boldsymbol{P}^*\boldsymbol{P}]_{i,j} &= \left\langle \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s\cdot|\mathbb{S}^{d-1}|}} \cdot P_d^\ell\left(\langle w_i, \cdot\rangle\right), \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{\sqrt{s\cdot|\mathbb{S}^{d-1}|}} \cdot P_d^\ell\left(\langle w_j, \cdot\rangle\right) \right\rangle_{\mathbb{S}^{d-1}} \\
&= \sum_{\ell=0}^{q}\sum_{\ell'=0}^{q} \frac{\alpha_{\ell,d}\alpha_{\ell',d}}{s\cdot|\mathbb{S}^{d-1}|} \cdot \left\langle P_d^\ell\left(\langle w_i, \cdot\rangle\right), P_d^{\ell'}\left(\langle w_j, \cdot\rangle\right) \right\rangle_{\mathbb{S}^{d-1}} \\
&= \sum_{\ell=0}^{q}\sum_{\ell'=0}^{q} \frac{\alpha_{\ell,d}\alpha_{\ell',d}}{s} \cdot \mathbb{E}_{v\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[P_d^\ell\left(\langle w_i, v\rangle\right) \cdot P_d^{\ell'}\left(\langle w_j, v\rangle\right)\right] \\
&= \sum_{\ell=0}^{q} \frac{\alpha_{\ell,d}}{s} \cdot P_d^\ell\left(\langle w_i, w_j\rangle\right) = \boldsymbol{K}_{i,j},
\end{aligned}
$$

where the fourth line above follows from Lemma 4. Therefore, we are interested in the optimal solution of the following least-squares problem

$$
\boldsymbol{z} \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^s} \|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\|_2^2 .
$$

The least-squares solution to the above problem is $\boldsymbol{z} = \boldsymbol{K}^\dagger \boldsymbol{f}$ which is exactly what is computed in line 6 of the algorithm. Now note that, the function $\tilde{g} = \boldsymbol{P}\cdot\boldsymbol{z}$ satisfies Eq. (21). Because $\tilde{g} = \boldsymbol{P}\cdot\boldsymbol{z} \in \mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ and because $\mathcal{K}_d^{(q)}$ is an orthonormal projection operator into $\mathcal{H}^{(q)}(\mathbb{S}^{d-1})$, we have $\mathcal{K}_d^{(q)}\cdot\tilde{g} = \tilde{g} = \boldsymbol{P}\cdot\boldsymbol{z}$. This together with Eq. (21) imply that,

$$
\|\boldsymbol{P}\cdot\boldsymbol{z} - f\|_{\mathbb{S}^{d-1}}^2 \le (1+\epsilon)\cdot \min_{g\in L^2\left(\mathbb{S}^{d-1}\right)} \left\|\mathcal{K}_d^{(q)} g - f\right\|_{\mathbb{S}^{d-1}}^2 .
$$

Now if we invoke Claim 2 with $C = 1+\epsilon$ on the above inequality we find that,

$$
\left\|\boldsymbol{P}\cdot\boldsymbol{z} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \le \epsilon\cdot\left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2 .
$$

Finally, one can easily see that the function $y \in \mathcal{H}^{(q)}(\mathbb{S}^{d-1})$ that Algorithm 1 outputs in line 7 is exactly equal to $y = \boldsymbol{P}\cdot\boldsymbol{z}$. This completes the accuracy bound of the theorem.

Runtime and Sample Complexity. these bounds follow from observing that:

- $s\cdot d$ time is needed to generate $w_1, w_2, \ldots, w_s$ in line 3 of the algorithm. To do this, we first generate random Gaussian points in $\mathbb{R}^d$ and then project then onto $\mathbb{S}^{d-1}$ by normalizing them.
- $s^2\cdot d$ operations are needed to form the kernel matrix $\boldsymbol{K}$ in line 4 of the algorithm.
- s queries to function f are needed to form the samples vector $\boldsymbol{f}$ in line 5 of the algorithm.
- s^ω time is needed to compute the least-squares solution $\boldsymbol{z} = \boldsymbol{K}^\dagger\boldsymbol{f}$ in line 6 of the algorithm.
- $s\cdot d$ operations are needed to evaluate the output function $y(\sigma)$ in line 7 of the algorithm.

This completes the proof of Theorem 14. □

E Lower Bound: Claims and Lemmas

In this section we prove the Claims and Lemmas used in our lower bound analysis for proving Theorem 15.

Claim 3. *Given the random input $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)}$ generated as described above, to solve Problem 2, an algorithm must return a function $\tilde{f}^{(q)} \in \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$ such that $\|\tilde{f}^{(q)} - f\|_{\mathbb{S}^{d-1}}^2 = 0$.*

Proof. Note that Problem 2 requires recovering a function $\tilde{f}^{(q)} \in \mathcal{H}^{(q)}\left(\mathbb{S}^{d-1}\right)$ such that:

$$\left\|\tilde{f}^{(q)} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \leq \epsilon \cdot \left\|f^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2, \tag{22}$$

where $f^{(q)} = \mathcal{K}_d^{(q)} f$. Using the definition of the input function $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)}$, we can write,

$$\begin{aligned} f^{(q)} = \mathcal{K}_d^{(q)} f &= \sum_{\ell=0}^{q} \mathcal{K}_d^{(q)} \cdot \boldsymbol{Y}_\ell \cdot v^{(\ell)} \\ &= \sum_{\ell=0}^{q} \left(\sum_{\ell'=0}^{q} \boldsymbol{Y}_{\ell'} \boldsymbol{Y}_{\ell'}^* \right) \cdot \boldsymbol{Y}_\ell \cdot v^{(\ell)} \\ &= \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)} = f, \end{aligned}$$

where the equality in the second line above follows from Eq. (11) and the addition theorem in Theorem 5, and the third line follows because the operator $\boldsymbol{Y}_\ell$ has orthonormal columns and thus $\boldsymbol{Y}_{\ell'}^* \boldsymbol{Y}_\ell = I_{\alpha_{\ell,d}} \cdot \mathbb{1}_{\{\ell=\ell'\}}$. Therefore, plugging this into Eq. (22) gives,

$$\left\|\tilde{f}^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2 = \left\|\tilde{f}^{(q)} - f^{(q)}\right\|_{\mathbb{S}^{d-1}}^2 \leq \epsilon \cdot \|f^{(q)} - f\|_{\mathbb{S}^{d-1}}^2 = \epsilon \cdot \|f - f\|_{\mathbb{S}^{d-1}}^2 = 0.$$

□

Lemma 16. *If a deterministic algorithm solves Problem 2 with probability at least $1/10$ over our random input distribution $f = \sum_{\ell=0}^{q} \boldsymbol{Y}_\ell \cdot v^{(\ell)}$, then with probability at least $1/10$, the output of the algorithm $\tilde{f}^{(q)}$ satisfies $\boldsymbol{Y}_\ell^* \tilde{f}^{(q)} = v^{(\ell)}$ for all integers $\ell \leq q$.*

Proof. By Claim 3, the output of the algorithm that solves Problem 2, satisfies $\left\|\tilde{f}^{(q)} - f\right\|_{\mathbb{S}^{d-1}}^2 = 0$. Therefore, by orthonormality of the columns of the operator $\boldsymbol{Y}_\ell$, we can write,

$$\boldsymbol{Y}_\ell^* \tilde{f}^{(q)} = \boldsymbol{Y}_\ell^* f + \boldsymbol{Y}_\ell^* (\tilde{f}^{(q)} - f) = \sum_{\ell'=0}^{q} \boldsymbol{Y}_\ell^* \boldsymbol{Y}_{\ell'} \cdot v^{(\ell')} = v^{(\ell)}.$$

□