SIAM J. NUMER. ANAL.
Vol. 0, No. 0, pp. 000–000

DISTRIBUTED FILTERED HYPERINTERPOLATION FOR NOISY DATA ON THE SPHERE*

SHAO-BO LIN†, YU GUANG WANG‡, AND DING-XUAN ZHOU§

Abstract. Problems in astrophysics, space weather research, and geophysics usually need to analyze big noisy data on the sphere. This paper develops distributed filtered hyperinterpolation for noisy data on the sphere, which assigns the data fitting task to multiple servers to find a good approximation of the mapping of input and output data. For each server, the approximation is a filtered hyperinterpolation on the sphere by a small proportion of quadrature nodes. The distributed strategy allows parallel computing for data processing and model selection. It reduces computational cost for each server while preserving the approximation capability compared to the filtered hyperinterpolation. We prove a quantitative relation between the approximation capability of distributed filtered hyperinterpolation and the numbers of input data and servers. Numerical examples show the efficiency and accuracy of the proposed method.

Key words. distributed learning, filtered hyperinterpolation, noisy data, big data, sphere

AMS subject classifications. 68Q32, 65D05, 41A50, 33C55, 65T60

DOI. 10.1137/19M1281095

1. Introduction.

In cosmic microwave background analysis, global ionospheric prediction of geomagnetic storms, climate change modeling, environmental governance and meteorology, and remote sensing, data are collected on the sphere and usually big and noisy [1, 13, 14, 32, 36, 41]. One of the critical tasks of big data analysis on the sphere is to find an effective data fitting strategy to approximate the mapping between input and output data. There have been many useful methods for fitting spherical data, for example, approximations by spherical harmonics [37], spherical basis functions [25, 26, 20, 39], spherical wavelets [13], spherical needlets [2, 38, 12, 22, 31, 47], spherical kernel methods [11, 28], and spherical filtered hyperinterpolation [44]. When noise is sufficiently small and decreases with the size of data, least squares regularization can be used to reduce noise in learning representation; see, e.g., [25, 20]. This method is, however, not suitable when the size of noisy data is big, as then the regularization condition implies that noise must be close to zero. In this paper, we propose a new strategy based on distributed learning—*distributed filtered hyperintepolation*, which assigns the data fitting task to multiple servers, then synthesizes them as a global prediction model.

*Received by the editors August 14, 2019; accepted for publication (in revised form) November 2, 2020; published electronically DATE.
https://doi.org/10.1137/19M1281095
Funding: The work of the first author was supported by National Natural Science Foundation of China grant 61876133. The work of the second author was supported by the Australian Research Council under Discovery Project DP180100506. The work of the third author was supported by Research Grant Council of Hong Kong project CityU 11338616. This material is based on work supported by National Science Foundation grant DMS-1439786 while the second author was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, Rhode Island, during the Point Configurations in Geometry, Physics, and Computer Science program.

†School of Management, Xi'an Jiaotong University, Xi'an, China (sblin1983@gmail.com).
‡School of Mathematics and Statistics, University of New South Wales, Sydney, NSW, Australia and Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany (yuguang.wang@unsw.edu.au).
§School of Data Science and Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong (mazhou@cityu.edu.hk).

Spherical filtered hyperinterpolation, developed by Sloan [43] and Sloan and Womersley [44], is a constructive approach: Given degree L (which indicates the level of precision), use N data on the sphere to find a filtered expansion of spherical harmonics up to degree $2L$. The filtering strategy uses an appropriate filter and the data at nodes of a quadrature rule. When the quadrature rule that is a set of pairs of points on the sphere and real weights is exact for numerical integration of polynomials up to degree $(c+1)L$ (see the definition of (2.4)), the filtered hyperinterpolation reaches the best polynomial approximation [44, 45]. The computational cost is thus determined by the number N of data, which has at least order $\mathcal{O}(L^2)$ [19]. The cost becomes heavy as degree L or the number of data increases. One way to reduce the computational time is to distribute the approximation task to multiple servers, each of which works on a fraction of the total computation using a small proportion of all data, and then synthesize the computed fitting models of all servers to produce a global predictor. To achieve that, we apply a *distributed learning* strategy [29, 30] to spherical filtered hyperinterpolation, which leads to distributed filtered hyperinterpolation (DFH).

The proposed DFH can fit N noisy data $y_i = f^*(\mathbf{x}_i) + \epsilon_i$, $i = 1, \ldots, N$ for continuous function f^* on the sphere and independent bounded noises ϵ_i. We show that the approximation error of DFH for such noisy data depends on the number of data and the smoothness of the target function f^*. We also consider the DFH with independent random interpolation points, which we call *distributed filtered hyperinterpolation with random sampling*, and prove that for the distribution of random points satisfying appropriate conditions, DFH has the same approximation capability as that with a "deterministic" quadrature rule.

The rest of the paper is organized as follows. In section 2, we introduce the filtered kernel, the quadrature rule, and "ordinary" spherical filtered hyperinterpolation. In section 3, we define distributed filtered hyperinterpolation with the "deterministic" quadrature rule and give the relation of its approximation error and numbers of data and servers. Section 4 defines the distributed filtered hyperinterpolation with random sampling and gives its error estimate. Section 5 gives numerical examples of the distributed filtered hyperinterpolation, where we study the impact of the numbers of data, servers, and noise on approximation error. Section 6 gives the proofs for the main results.

2. Filtered kernel, quadrature rule, and filtered hyperinterpolation. In this section, we introduce the spherical filtered hyperinterpolation defined by filtered kernels and the quadrature rule.

2.1. Filtered kernels. For $d \geq 2$, let $\mathbf{x} \cdot \mathbf{y}$ be the inner product of two points $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^{d+1}$ and the Euclidean norm $|\mathbf{x}| := \sqrt{\mathbf{x} \cdot \mathbf{x}}$. Let $\mathbb{S}^d := \{\mathbf{x} \in \mathbb{R}^{d+1} : |\mathbf{x}| = 1\}$ be the unit sphere of $\mathbb{R}^{d+1}$. The $\mathbb{S}^d$ is a compact metric space with geodesic distance

$$\mathrm{dist}(\mathbf{x}, \mathbf{y}) := \arccos(\mathbf{x} \cdot \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}^d,$$

as the metric. For $1 \leq p < \infty$, let $L_p(\mathbb{S}^d)$ be the real-valued L_p space on $\mathbb{S}^d$ with Lebesgue measure $\omega := \omega_d$ and L_p norm $\|f\|_{L_p(\mathbb{S}^d)} := (\int_{\mathbb{S}^d} |f(\mathbf{x})|^p \mathrm{d}\omega(\mathbf{x}))^{1/p}$ for $f \in L_p(\mathbb{S}^d)$. In particular, $L_2(\mathbb{S}^d)$ is a Hilbert space with inner product

$$\langle f, g \rangle_{L_2(\mathbb{S}^d)} := \int_{\mathbb{S}^d} f(\mathbf{x}) g(\mathbf{x}) \mathrm{d}\omega(\mathbf{x}), \quad f, g \in L_p(\mathbb{S}^d).$$

Denote by $L_\infty(\mathbb{S}^d)$ the space of all real-valued continuous functions on $\mathbb{S}^d$ with uniform norm $\|f\|_{L_\infty} := \max_{\mathbf{x} \in \mathbb{S}^d} |f(\mathbf{x})|$. The volume of $\mathbb{S}^d$ is $|\mathbb{S}^d| := \omega_d(\mathbb{S}^d) := \frac{2\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$.

For $\ell \in \mathbb{N}_0 := \{0, 1, \dots\}$, the restriction to $\mathbb{S}^d$ of a homogeneous harmonic polynomial of degree ℓ is called a spherical harmonic of degree ℓ. Let $\mathbb{H}_\ell^d$ be the space of all spherical harmonics of degree ℓ and, for $L \in \mathbb{N}_0$, Π_L^d the space of all spherical polynomials of degree up to L. Then $\Pi_L^d = \bigoplus_{\ell=0}^{L} \mathbb{H}_\ell^d$. The dimension of $\mathbb{H}_\ell^d$ is

$$Z_{d,\ell} := \dim \mathbb{H}_\ell^d = \begin{cases} \dfrac{2\ell + d - 1}{\ell + d - 1}\dbinom{\ell + d - 1}{\ell}, & \ell \geq 1, \\ 1, & \ell = 0, \end{cases}$$

and then the dimension of Π_L^d is $\sum_{\ell=0}^{L} Z_{d,\ell} = Z_{d+1,L} \asymp L^d$. Here for two sequences a_ℓ and b_ℓ, $\ell \in \mathbb{N}_0$, $a_\ell \asymp b_\ell$ means that there exists constants c, c' such that $c' b_\ell \leq a_\ell \leq c b_\ell$ for all $\ell \in \mathbb{N}_0$. The Laplace–Beltrami operator Δ^* on $\mathbb{S}^d$ has the eigenfunctions $Y_{\ell,m}$ and eigenvalues $\lambda_\ell := \ell(\ell + d - 1)$, $\ell \in \mathbb{N}_0$, $m = 1, \dots, Z_{d,\ell}$: $\Delta^* Y_{\ell,m} = \lambda_\ell Y_{\ell,m}$.

Let $P_\ell^{(d+1)}(t)$, $t \in [-1, 1]$, be the normalized Gegenbauer polynomial which satisfies $P_\ell^{(d+1)}(1) = 1$ and the orthogonality relation

$$\int_{-1}^{1} P_\ell^{(d+1)}(t) P_{\ell'}^{(d+1)}(t)(1 - t^2)^{\frac{d-2}{2}} \mathrm{d}t = \frac{|\mathbb{S}^d|}{|\mathbb{S}^{d-1}| Z_{d,\ell}} \delta_{\ell,\ell'},$$

where $\delta_{\ell,\ell'}$ is the Kronecker symbol. Let $\eta : [0, \infty) \to \mathbb{R}$ be a filter with specified smoothness $\kappa \geq 1$ satisfying

$$\eta \in C^\kappa(\mathbb{R}_+); \quad \operatorname{supp} \eta \subseteq [1/2, 2]; \quad \eta(t)^2 + \eta(2t)^2 = 1 \text{ for } t \in [1/2, 1]. \tag{2.1}$$

The filtered kernel $K_n(\mathbf{x} \cdot \mathbf{x}')$, $n \geq 1$, is then given by

$$K_n(\mathbf{x} \cdot \mathbf{x}') = \sum_{\ell=0}^{\infty} \eta\left(\frac{\ell}{n}\right) \frac{Z_{d,\ell}}{|\mathbb{S}^d|} P_\ell^{(d+1)}(\mathbf{x} \cdot \mathbf{x}'); \tag{2.2}$$

see [38]. The approximation property of the filtered kernel depends on the smoothness of the filter η. We refer the reader to, e.g., [44, p. 101] and [47] for examples of filtered kernels satisfying (2.1). Since the support of η is in $[1/2, 2]$, $K_n(\mathbf{x} \cdot \mathbf{y})$ is a polynomial of either $\mathbf{x}$ or $\mathbf{y}$ of degree up to $2n - 1$. Here, we give an example of filter η, which is in $C^5(\mathbb{R}_+)$ and defined by a piecewise polynomial. For $0 \leq t \leq 1$, $\eta(t) = 1$; for $t \geq 2$, $\eta(t) = 0$; and for $1 < t < 2$,

$$\begin{aligned} \eta(t) = 1 + (t-1)^6 \big[&-462 + 1980(t-1) - 3465(t-1)^2 + 3080(t-1)^3 \\ &- 1386(t-1)^4 + 252(t-1)^5\big]. \end{aligned}$$

We will use this example in the experiment in section 5.

2.2. Spherical quadrature rules. The geometric properties of a finite set $X_N := \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, $N \geq 2$, of points on $\mathbb{S}^d$ can be described by mesh norm, separation radius, and mesh ratio, as we introduce now. The *mesh norm* (or covering radius) of X_N is

$$h(X_N) := \max_{\mathbf{x} \in \mathbb{S}^d} \min_{\mathbf{x}_i \in X_N} \operatorname{dist}(\mathbf{x}, \mathbf{x}_i).$$

The mesh norm is the minimal radius with which the caps with centers at points of X_N cover $\mathbb{S}^d$. The separation radius of X_N is

$$\delta(X_N) := \frac{1}{2} \min_{j \neq k} \operatorname{dist}(\mathbf{x}_j, \mathbf{x}_k).$$

This is half the smallest geodesic distance between any pair of points in X_N. The mesh ratio of X_N is the minimum of distances between points of X_N,

$$\rho(X_N) := \frac{h(X_N)}{\delta(X_N)} \geq 1,$$

which measures how uniformly the points of X_N are distributed on $\mathbb{S}^d$. We say a sequence of point sets $\{X_N\}_{N=2}^{\infty}$ *τ-quasi uniform* if there is a constant $\tau \geq 2$ such that $\rho(X_N) \leq \tau$ for all $N \geq 2$. The existence of a τ-quasi uniform sequence of point sets is proved in [39]. Assume the sequence of point sets $\{X_N\}_{N=2}^{\infty}$ is τ-quasi uniform. Then

$$h(X_N) \leq \tau\delta(X_N) \leq \frac{\tau}{2N^{1/d}}. \tag{2.3}$$

It then follows from (2.3) and [27, Lemma 2] that

$$\mathbb{S}^d \subseteq \bigcup_{\mathbf{x}_i \in X_N} \mathcal{C}(\mathbf{x}_i, \tau/(2N^{1/d})) \quad \text{and} \quad \max_{\mathbf{x}_i \in X_N} \left| X_N \bigcap \mathcal{C}(\mathbf{x}_i, \tau/(2N^{1/d})) \right| \leq 2\pi^{d-1}\tau^d,$$

where $|A|$ is the cardinality of the finite set A and $\mathcal{C}(\mathbf{x}, r) := \{\mathbf{y} \in \mathbf{S}^d : \mathrm{dist}(\mathbf{x}, \mathbf{y}) \leq r\}$ the spherical cap with center $\mathbf{x}$ and radius r, $r > 0$.

We say a set $\mathcal{Q}_N := \{(w_i, \mathbf{x}_i) : w_i \in \mathbb{R} \text{ and } \mathbf{x}_i \in \mathbb{S}^d, i = 1, \dots, N\}$, $N \geq 2$, a *quadrature rule* on $\mathbb{S}^d$, where w_i are called weights of $\mathcal{Q}_N$. For $n \geq 0$, a quadrature rule $\mathcal{Q}_N$ is said to be exact for polynomials of degree up to n if

$$\int_{\mathbb{S}^d} P(\mathbf{x}) \mathrm{d}\omega(\mathbf{x}) = \sum_{i=1}^{N} w_i P(\mathbf{x}_i) \quad \forall P \in \Pi_n^d. \tag{2.4}$$

The following lemma gives a sequence of polynomial-exact quadrature rules whose point sets are τ-quasi uniform; see [7, Theorem 3.1] and [35, 38, 18, 23, 33].

Lemma 2.1 ([7, 35, 38]). *If $\{X_N\}_{N=2}^{\infty}$ is τ-quasi uniform, then for $N \geq 2$, there exist positive weights w_i, $i = 1, \dots, N$, such that $0 \leq w_i \leq c_2 N^{-1}$ and*

$$\int_{\mathbb{S}^d} P(\mathbf{x}) \mathrm{d}\omega(\mathbf{x}) = \sum_{\mathbf{x}_i \in X_N} w_i P(\mathbf{x}_i) \quad \forall P \in \Pi^d_{c_3 N^{1/d}},$$

where c_2 and c_3 are constants depending only on τ and d.

For $1 \leq p < \infty$, let $L_{p,\mu} := L_p(\mathbb{S}^d, \mu)$ be the L_p space on $\mathbb{S}^d$ with respect to a probability measure μ, endowed with norm $\|f\|_{p,\mu} := (\int_{\mathbb{S}^d} |f(\mathbf{x})|^p \mathrm{d}\mu(\mathbf{x}))^{1/p}$. The following theorem shows that if $X_N = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{S}^d$ is a set of i.i.d. random points with distribution μ, then with high probability, the quadrature rule is exact for polynomials with a specific degree. We postpone the proof of Theorem 2.2 until section 6.

Theorem 2.2. *Let $X_N = \{\mathbf{x}_i\}_{i=1}^N$ be i.i.d. random points on $\mathbb{S}^d$ with distribution μ, which satisfies*

$$\|f\|_{L_1(\mathbb{S}^d)} \leq c_4 \|f\|_{1,\mu} \quad \forall f \in L_1(\mathbb{S}^d) \cap L_{1,\mu} \tag{2.5}$$

for a positive absolute constant c_4. Then, for integer N satisfying $N/n^{2d} > c$ for a sufficiently large constant c, there exists a quadrature rule $\mathcal{Q}_N := \{(w_{i,n}, \mathbf{x}_i)\}_{i=1}^N$ such that

$$\int_{\mathbb{S}^d} P_n(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) = \sum_{i=1}^{N} w_{i,n} P_n(\mathbf{x}_i) \quad \forall P_n \in \Pi_n^d$$

holds, and $\sum_{i=1}^{N}|w_{i,n}|^2 \leq \frac{2}{N}$ *and* $w_{i,n} \geq 0$ *for all* $i = 1,\ldots,N$*, with confidence at least* $1 - 4\exp\left\{-CN/n^d\right\}$*, where* C *is a constant depending only on* c_4 *and* d*.*

The confidence level in Theorem 2.2 is an exponential of n and N, which is different from the polynomial confidence level given by [24, Theorem 4.1]. This exponential relation plays a crucial role in the error estimation for distributed filtered hyperinterpolation with spherical noisy data. It is also different from [28] in that the established quadrature rule with positive weights satisfies (2.5). The condition in (2.5) describes the distortion of distribution μ from the uniform distribution on $\mathbb{S}^d$ (which is also the spherical Lebesgue measure). The probabilistic quadrature rule in Theorem 2.2 is critical to our construction of distributed filtered hyperinterpolation with random sampling. We will study the details in section 4.2.

2.3. Spherical filtered hyperinterpolation approximation. Using the filtered kernel and the quadrature rule, we define the spherical filtered hyperinterpolation as follows.

Definition 2.3 ([44]). *Let* $d \geq 2$ *and* $1 \leq p \leq \infty$*. For* $f \in L_p(\mathbb{S}^d)$*, the filtered hyperinterpolation (approximation) with a quadrature rule* $\mathcal{Q}_N := \{(w_i, \mathbf{x}_i)\}_{i=1}^N$ *is*

$$V_{n,N}(f;\mathbf{x}) := \sum_{i=1}^{N} w_i f(\mathbf{x}_i) K_n(\mathbf{x}_i \cdot \mathbf{x}), \quad n \geq 1. \tag{2.6}$$

The approximation property of the filtered hyperinterpolation depends on the smoothness of the function space. For $r \in \mathbb{R}_+$, let $b_\ell^r := (1+\lambda_\ell)^{r/2} \asymp (1+\ell)^r$. Let $\{Y_{\ell,m} : \ell = 0,1,\ldots,\ m = 1,\ldots,Z_{d,\ell}\}$ be an orthonormal basis for the space $L_2(\mathbb{S}^d)$ and

$$\widehat{f}_{\ell m} := \langle f, Y_{\ell,m}\rangle_{L_2(\mathbb{S}^d)} := \int_{\mathbb{S}^d} f(\mathbf{x}) Y_{\ell m}(\mathbf{x}) \mathrm{d}\omega(\mathbf{x})$$

the Fourier coefficients of $f \in L_2(\mathbb{S}^d)$. For $d \geq 2$, $1 \leq p \leq \infty$, and $r \in \mathbb{R}_+$, the Sobolev space $\mathbb{W}_p^r(\mathbb{S}^d)$ is the space of functions f satisfying $\sum_{\ell=0}^{\infty} b_\ell^r \widehat{f}_{\ell m} Y_{\ell,m} \in L_p(\mathbb{S}^d)$, endowed with the norm $\|f\|_{\mathbb{W}_p^r(\mathbb{S}^d)} := \|\sum_{\ell=0}^{\infty} b_\ell^r \widehat{f}_{\ell m} Y_{\ell,m}\|_{L_p(\mathbb{S}^d)}$.

The following lemma, which is proved by Wang and Sloan [45], shows the relation of the approximation error of the filtered hyperinterpolation approximation $V_{n,N}$ and the smoothness of the function space under the condition that the associated quadrature rule is exact for polynomials of degree up to $3n-1$.

Lemma 2.4 ([45]). *Let* $d \geq 2$*,* $1 \leq p \leq \infty$*, and* $r > d/p$*. Let* $V_{n,N}$ *be the filtered hyperinterpolation in* (2.6) *with quadrature rule* $\mathcal{Q}_N$ *exact for polynomials of degree up to* $3n-1$ *and with the filter* η *in* $C^\kappa(\mathbb{R}_+)$*,* $\kappa \geq \lfloor \frac{d+3}{2} \rfloor$*. Then, for* $f \in \mathbb{W}_p^r(\mathbb{S}^d)$*,*

$$\|f - V_{n,N}(f)\|_{L_p(\mathbb{S}^d)} \leq c_5\, n^{-r} \|f\|_{\mathbb{W}_p^r(\mathbb{S}^d)}, \tag{2.7}$$

where c_5 *depends only on* $d, p, r,$ *and* η*. The order* n^{-r} *in* (2.7) *is optimal.*

3. Distributed filtered hyperinterpolation with deterministic sampling. A *data set* $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{|D|}$ on $\mathbb{S}^d$ is a set of pairs of points $\Lambda_D := \{\mathbf{x}_i\}_{i=1}^{|D|}$ on the sphere and real numbers y_i. Elements of D are called *data*. The points $\mathbf{x}_i$ of Λ_D are called the *sampling points* of D. To distinguish the quadrature rule with random points, which we will investigate later, we say a data set D has *deterministic sampling* for (fixed) sampling points. In this section, we introduce a new filtered hyperinterpolation for which distributed learning can be used. The data y_i are the values of a function f^* on $\mathbb{S}^d$ plus noise, that is,

$$y_i = f^*(\mathbf{x}_i) + \epsilon_i, \quad \mathbf{E}[\epsilon_i] = 0, \quad |\epsilon_i| \leq M \quad \forall i = 1, \ldots, |D|. \tag{3.1}$$

The D satisfying (3.1) is then called *noisy data set associated with* f^*.

3.1. Filtered hyperinterpolation for noisy data: Deterministic sampling. We first study the performance of the filtered hyperinterpolation for a noisy data set D whose data are stored on a "big enough" machine.

DEFINITION 3.1. *For $s \in \mathbb{N}_0$ and $\{\mathbf{x}_i\}_{i=1}^{|D|}$, let $\mathcal{Q}_{|D|}$ be the quadrature rule given by Lemma* 2.1, *which is exact for polynomials of degree up to s and has positive weights $\{w_{i,s,D}\}_{i=1}^{|D|}$ satisfying $0 \leq w_{i,s,D} \leq c_2|D|^{-1}$. The filtered hyperinterpolation for noisy data associated with a function f^* on $\mathbb{S}^d$ is*

$$f^{\diamond}_{D,n}(x) := \sum_{i=1}^{|D|} w_{i,s,D}\, y_i K_n(\mathbf{x}_i \cdot \mathbf{x}), \tag{3.2}$$

where K_n is a filtered kernel in (2.2) *for $\eta \in C^{\kappa}(\mathbb{R}_+)$ with $\kappa \geq \lfloor \frac{d+3}{2} \rfloor$ and $n \leq s$.*

The kernel K_n provides a smoothing method for the function f^* using data D. As we shall see below, the approximation error of this filtered hyperinterpolation has the convergence rate depending on the smoothness of function f^*.

If Λ_D is τ-quasi uniform, it then follows from Lemma 2.1 that $s = c_3 N^{1/d}$. We do not assume the magnitude of the noise to be extremely small; the filter η of the filtered kernel K_n shall then be chosen properly to minimize the impact of noise on interpolation. It is a problem similar to "model selection" in statistical and machine learning [9]. To say it precisely, if the support n is too large, then the filtered hyperinterpolation $f^{\diamond}_{D,n}$ will have precise approximation at the data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^{D}$, but $f^{\diamond}_{D,n}$ may not be a good approximation of f^* due to the noise. If the support is too small, the performance of the filtered hyperinterpolation $f^{\diamond}_{D,n}(\mathbf{x}_i)$ is not good, even at the interpolation points. It is then preferable to set n as a parameter in the training process. For Λ_D, the quadrature rule in Definition 3.1 (which is from Lemma 2.1) is valid for n sufficiently large. The parameter selection thus needs only a few steps of computation, while the filtered hyperinterpolation in Definition 3.1 allows us to handle massive noisy spherical data. This property of filtered hyperinterpolation is different from other methods, such as regularized least squares [20]. The latter needs to compute the inverse of the kernel matrix for each regularization parameter.

The following theorem shows that the filtered hyperinterpolation $f^{\diamond}_{D,n}$ can approximate f^* well provided that the support of the filtered kernel is appropriately tuned and the sampling point set Λ_D is τ-quasi uniform for $\tau \geq 2$.

THEOREM 3.2. *Let $d \geq 2$ and $r > d/2$. Assume that the sampling point set Λ_D of the data set D is τ-quasi uniform for $\tau \geq 2$ and that $\frac{c_3}{6}|D|^{1/(2r+d)} \leq n \leq \frac{c_3}{3}|D|^{1/(2r+d)}$ for constant c_3 in Lemma* 2.1. *Then the filtered hyperinterpolation $f^{\diamond}_{D,n}$ for noisy data set D with target function $f^* \in \mathbb{W}_2^r(\mathbb{S}^d)$ satisfies*

$$\mathbf{E}\left\{\|f^{\diamond}_{D,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\right\} \leq C_1 |D|^{-2r/(2r+d)}, \tag{3.3}$$

where C_1 is a constant independent of $|D|$ and n.

Remark 3.3. Here the condition $r > d/2$ is the embedding condition such that any function in $\mathbb{W}_2^r(\mathbb{S}^d)$ has a representation of a continuous function on $\mathbb{S}^d$. The numerical computation of the filtered hyperinterpolation then makes sense.

We give the proof of Theorem 3.2 in section 6. As mentioned above, since Λ_D is τ-quasi uniform, the quadrature rule for the filtered hyperinterpolation in Definition 3.1 is exact for Π_s^d with $s = c|D|^{1/d}$. As we choose $n \leq c|D|^{1/(2r+d)} \leq s$ in Theorem 3.2, $f_{D,n}^{\diamond}$ reproduces polynomials in Π_s^d. Theorem 3.2 illustrates that if the scattered data Λ_D have good geometric property, for example, τ-quasi uniformity and the support of the filter η is appropriately chosen, then the spherical filtered hyperinterpolation for the noisy data set D can approximate a sufficiently smooth target function on the sphere with high precision in a probabilistic sense. By [17], the rate $|D|^{-2r/(2r+d)}$ in (3.3) cannot be essentially improved in the scenario of (3.1). Thus, Theorem 3.2 provides a feasibility analysis of the spherical filtered hyperinterpolation for spherical data with random noise.

3.2. Distributed filtered hyperinterpolation: Deterministic sampling. We say a large data set D is *distributively stored* on m local machines if for $j = 1, \dots, m$, $m \geq 2$, the jth machine contains a subset D_j of D and there is no common data between any pair of machines, that is, $D_j \cap D_{j'} = \emptyset$ for $j \neq j'$ and $D = \cup_{j=1}^m D_j$. The data sets $D_1, \dots, D_m$ are called *distributed data sets* of D. In this case, the filtered hyperinterpolation $f_{D,n}^{\diamond}$ which needs access to the entire data set D, is infeasible. Instead, in this section, we construct a *distributed filtered hyperinterpolation for the distributed data sets* $\{D_j\}_{j=1}^m$ *of* D by the divide-and-conquer strategy [29].

DEFINITION 3.4. *The distributed filtered hyperinterpolation* $\overline{f_{D,n}^{\diamond}}$ *for distributed data sets* $\{D_j\}_{j=1}^m$ *of a noisy data set* D *associated with function* f^* *on* $\mathbb{S}^d$ *is a synthesized estimator of local estimators* $f_{D_j,n}^{\diamond}$, $j = 1, 2, \dots, m$, *each of which is the spherical filtered hyperinterpolation* (3.2) *for noisy data set* D_j*:*

$$\overline{f^{\diamond}}(\{D_j\}_{j=1}^m, n; \mathbf{x}) := \overline{f_{D,n}^{\diamond}}(\mathbf{x}) := \sum_{j=1}^{m} \frac{|D_j|}{|D|} f_{D_j,n}^{\diamond}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^d. \tag{3.4}$$

The synthesis here is a process when the local estimators communicate to a central processor to produce the global estimator $\overline{f_{D,n}^{\diamond}}$.

The following theorem shows that the distributed filtered hyperinterpolation $\overline{f_{D,n}^{\diamond}}$ has similar approximation performance as $f_{D,n}^{\diamond}$ if the number of local machines is not large.

THEOREM 3.5. *Let* $d \geq 2$, $r > d/2$, $m \geq 2$, *and* D *be a noisy data set satisfying* (3.1)*. Let* $\{D_j\}_{j=1}^m$ *be* m *distributed data sets of* D*. For* $j = 1, \dots, m$, *the sampling point set* Λ_{D_j} *of* D *is* τ*-quasi uniform for* $\tau \geq 2$*. If the distributed filtered hyperinterpolation* $\overline{f_{D,n}^{\diamond}}$ *for* $\{D_j\}_{j=1}^m$ *satisfies that the target function* f^* *is in* $\mathbb{W}_2^r(\mathbb{S}^d)$, $\frac{c_3}{6}|D|^{1/(2r+d)} \leq n \leq \frac{c_3}{3}|D|^{1/(2r+d)}$ *for constant* c_3 *in Lemma* 2.1 *and* $\min_{j=1,\dots,m} |D_j| \geq |D|^{\frac{d}{2r+d}}$, *then,*

$$\mathbf{E}\left\{\|\overline{f_{D,n}^{\diamond}} - f^*\|_{L_2(\mathbb{S}^d)}^2\right\} \leq C_2 |D|^{-2r/(2r+d)}, \tag{3.5}$$

where C_2 *is a constant independent of* $|D|$, $|D_1|, \dots, |D_m|$ *and* n.

Remark 3.6. The condition $\min_{j=1,\dots,m} |D_j| \geq |D|^{\frac{d}{2r+d}}$ has a close connection to the number m of local machines. In particular, if $|D_1| = \cdots = |D_m|$, since each D_j is τ-quasi uniform, $\min_{j=1,\dots,m} |D_j| \geq |D|^{\frac{d}{2r+d}}$ is equivalent to $m \leq |D|^{\frac{2r}{2r+d}}$.

The proof of Theorem 3.5 is postponed until section 6. Theorem 3.5 illustrates that if $\min_{j=1,\dots,m}|D_j| \geq |D|^{\frac{d}{2r+d}}$, then with the same assumption as Theorem 3.2, the distributed filtered hyperinterpolation will have the same approximation performance as the filtered hyperinterpolation that treats all the distributed data sets as a whole "big enough" machine.

4. Distributed filtered hyperinterpolation with random sampling. We say a data set D has *random sampling* if its sampling points are i.i.d. random points on $\mathbb{S}^d$. In this section, we construct a filtered hyperinterpolation for noisy data satisfying (3.1) with random sampling points.

4.1. Filtered hyperinterpolation for noisy data: Random sampling. The filtered hyperinterpolation for noisy data with random sampling can be constructed as follows. Let $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{|D|}$ and $n \in \mathbb{N}$. Let the quadrature rule $\mathcal{Q}_{|D|} := \{(w^*_{i,n,D}, \mathbf{x}_i)\}_{i=1}^{|D|}$ as given by Theorem 2.2, which is exact for polynomials of degree n. For $m \geq 2$, let

$$w_{i,n,D} = \begin{cases} w^*_{i,n,D}, & \text{if } \sum_{i=1}^{|D|} |w^*_{i,n,D}|^2 \leq 2/m, \\ 0, & \text{otherwise,} \end{cases} \qquad \forall i = 1, \dots, |D|. \tag{4.1}$$

Definition 4.1. *The filtered hyperinterpolation for noisy data $D := \{(\mathbf{x}_i, y_i)\}_{i=1}^{|D|}$ with random sampling points $\{\mathbf{x}_i\}_{i=1}^{|D|}$ is*

$$f_{D,n}(\mathbf{x}) := \sum_{i=1}^{|D|} w_{i,n,D}\, y_i K_n(\mathbf{x}_i \cdot \mathbf{x}). \tag{4.2}$$

The following theorem gives the approximation error of the filtered hyperinterpolation in Definition 4.1 for sufficiently smooth functions.

Theorem 4.2. *Let $d \geq 2$ and $r > d/2$. Let the noisy data set D with i.i.d. random sampling points on $\mathbb{S}^d$ and distribution μ satisfying* (2.5)*. For integer n satisfying $\frac{c_3}{6}|D|^{1/(2r+d)} \leq n \leq \frac{c_3}{3}|D|^{1/(2r+d)}$ with constant c_3 in Lemma* 2.1*, the filtered hyperinterpolation $f_{D,n}$ for noisy data set D with target function $f^* \in \mathbb{W}_2^r(\mathbb{S}^d)$ has the approximation error*

$$\mathbf{E}\left\{\|f_{D,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\right\} \leq C_3 |D|^{-2r/(2r+d)}, \tag{4.3}$$

where C_3 is a constant independent of $|D|$.

We give the proof of Theorem 4.2 in section 6. Theorems 3.2 and 4.2 show that the filtered hyperinterpolation approximations with random sampling and deterministic sampling achieve the same optimal convergence rate.

4.2. Distributed filtered hyperinterpolation: Random sampling. The distributed filtered hyperinterpolation with random sampling is a weighted average of filtered hyperinterpolation approximations for data on local machines. Here the weight for a local machine is the proportion of the data used by the machine to all data. Let $f_{D_j,n}$ be the filtered hyperinterpolation for data D_j. Similar to (3.4), the global estimator $\overline{f}_{D,n}$ is defined as follows.

TABLE 1
Computational steps and time for distributed and nondistributed learning by filtered hyperinterpolation. N is the total data number, m is the number of servers.

	Nondistributed	Distributed
Total steps	$\mathcal{O}(N)$	$\mathcal{O}(N)$
Real time	$\mathcal{O}(N)$	$\mathcal{O}(N/m)$

DEFINITION 4.3. *Let $d \geq 2$ and D be a noisy data set satisfying* (3.1). *The sampling points of D are i.i.d. random points on $\mathbb{S}^d$. For $m \geq 2$, let $\{D_j\}_{j=1}^m$ be m distributed data sets of D, and for $j = 1, \ldots, m$, let $f_{D_j,n}$ be the filtered hyperinterpolation for D_j given by Definition* 4.1. *The distributed filtered hyperinterpolation for distributed data sets $\{D_j\}_{j=1}^m$ of D is*

$$\overline{f}_{D,n} := \sum_{j=1}^{m} \frac{|D_j|}{|D|} f_{D_j,n}. \tag{4.4}$$

The f^* in (3.1) *is called a target function to $\overline{f}_{D,n}$.*

The following theorem gives an upper bound of approximation error for distributed filtered hyperinterpolation with random sampling.

THEOREM 4.4. *Let $d \geq 2$, $r > d/2$, $m \geq 2$, and D be a noisy data set satisfying* (3.1). *The sampling points are i.i.d. random points on $\mathbb{S}^d$ with distribution μ in* (2.5). *If the target function $f^* \in \mathbb{W}_2^r(\mathbb{S}^d)$, $\frac{c_3}{6}|D|^{1/(2r+d)} \leq n \leq \frac{c_3}{3}|D|^{1/(2r+d)}$ with constant c_3 in Lemma* 2.1 *and $\min_{j=1,\ldots,m} |D_j| \geq |D|^{\frac{d+\nu}{2r+d}}$ for some ν in $(0, 2r)$, then*

$$\mathbf{E}\left\{\|\overline{f}_{D,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\right\} \leq C_4 |D|^{-2r/(2r+d)}, \tag{4.5}$$

where C_4 is a constant independent of $|D|$, $|D_1|, \ldots, |D_m|$ and n.

The proof of Theorem 4.4 will be given in section 6. From Theorems 4.4 and 3.5, we see that the distributed filtered hyperinterpolation approximations with random sampling and deterministic sampling can both achieve the convergence rate of order $|D|^{-2r/(2r+d)}$. To achieve this approximation order, the condition on the number of local machines of the random sampling is stronger than the deterministic case since the former requires $\min_{j=1,\ldots,m} |D_j| \geq |D|^{(d+\nu)/(2r+d)}$ for $\nu \in (0, 2r)$, while the latter only needs $\min_{j=1,\ldots,m} |D_j| \geq |D|^{d/(2r+d)}$.

Here, we only consider error estimates for Lebesgue measure. It would be interesting to consider error estimates for distributed learning with respect to other measures as done in [54, 53, 55].

Computational complexity. We show in Table 1 a computational cost comparison of distributed and nondistributed filtered hyperinterpolation. The total computational steps for both are the same. Nevertheless, in the distributed case, we can, in parallel, compute the individual estimator. For the jth server, it takes $\mathcal{O}(|D_j|)$ steps. So, suppose we equally distribute the data to m servers: The distributed estimator will reduce the computational time to $1/m$ of the time for the nondistributed.

5. Numerical examples. In this section, we test distributed filtered hyperinterpolation on noisy data on $\mathbb{S}^2$. We use Womersley's *symmetric spherical t-designs*[1]

[1]https://web.maths.unsw.edu.au/%7Ersw/Sphere/EffSphDes/.

[50, 10] as the quadrature rule for distributed filtered hyperinterpolation. The symmetric spherical t-design is an equal-weighted quadrature with $\mathcal{O}(t^2)$ nodes satisfying (2.4) for degree $n \le t$, where the order $\mathcal{O}(t^2)$ is optimal as a consequence of [4].

Let $(u)_+ := \max\{u, 0\}$ for $u \in \mathbb{R}$. The normalized Wendland function is given by

$$\phi(u) := \widetilde{\phi}\left(\frac{8u}{15\sqrt{\pi}}\right), \quad u \in \mathbb{R},$$

where $\widetilde{\phi}(u)$ is the original Wendland function

$$\widetilde{\phi}(u) := (1-u)_+^8(32u^3 + 25u^2 + 8u + 1).$$

See [8, 49, 52]. The $\phi(\mathbf{x}\cdot\mathbf{z}) \in H^{4.5}(\mathbb{S}^2)$ is a radial basis function on the sphere $\mathbb{S}^2$ with center at $\mathbf{z} \in \mathbb{S}^2$ [26, 40]. We then define

$$f(\mathbf{x}) := \sum_{i=1}^{6} \phi(\mathbf{x}\cdot\mathbf{z}_i), \tag{5.1}$$

which is the linear combination of radial basis functions $\phi(\mathbf{x}\cdot\mathbf{z}_i)$ with centers at $\mathbf{z}_1 = (1,0,0)$, $\mathbf{z}_2 = (-1,0,0)$, $\mathbf{z}_3 = (0,1,0)$, $\mathbf{z}_4 = (0,-1,0)$, $\mathbf{z}_5 = (0,0,1)$, $\mathbf{z}_6 = (0,0,-1)$. The smoothness of f is 4.5, i.e., $f \in H^{4.5}(\mathbb{S}^2)$. We use the function f plus Gaussian white noise as the noisy data. That is, at each node $\mathbf{x}_i \in \mathbb{S}^2$,

$$y_i = f(\mathbf{x}_i) + \epsilon, \quad \epsilon \sim N(0, \sigma^2), \tag{5.2}$$

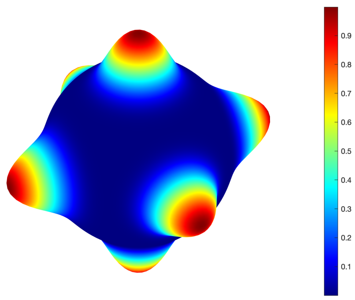
where $\mathbf{x}_i$ are the nodes of a symmetric spherical t-design and $\sigma \ge 0$.

Figures 1(a) and (b) show the pictures of f and a realization of $f+\epsilon$ with noise level $\sigma = 0.1$. Figure 1(c) shows the distributed filtered hyperinterpolation approximation $\overline{f^{\diamond}_{D,n}}$ of degree $n = 25$ for noisy data y_i. The distributed filtered hyperinterpolation uses $m = 100$ machines and C^5-filter η [46, 47, 48], which satisfies the condition of Theorem 3.5. On the jth machine, $j = 1, \dots, 100$, the filtered hyperinterpolation uses a $3n = 75$-design with 2,852 nodes, where the design is rotated from Womersley's symmetric spherical 75-design [50] by the rotation matrix

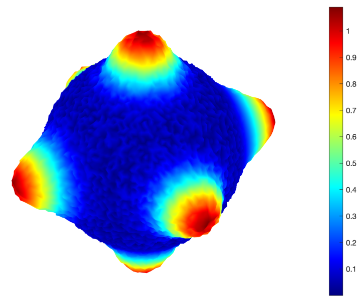
$$\rho_j := \begin{pmatrix} \cos(\theta_j) & -\sin(\theta_j) & 0 \\ \sin(\theta_j) & \cos(\theta_j) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with $\theta_j = j\pi/m$. The rotated spherical design satisfies the same polynomial exactness property for numerical integration as the unrotated symmetric spherical design since the rotation of a spherical design is still a spherical design with the same *separation* and *filling radii* [5, 6, 15]. The distributed filtered hyperinterpolation then satisfies the condition of Theorem 3.5, which uses 285,200 points in total. The function values are evaluated at 10,000 generalized spiral points on $\mathbb{S}^2$, which are equally distributed points [3, 42]. Figure 1(d) shows the approximation error of $\overline{f^{\diamond}_{D,n}}$ to f, which illustrates that errors are small compared to the magnitude of the function f.

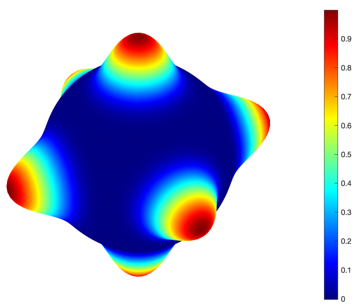
Figure 2 shows the convergence rate (with respect to n) of the approximation error of the distributed filtered hyperinterpolation for y_i with noise standard variance $\sigma = 0, 0.0001, 0.001, 0.01, 0.1$. It illustrates that the approximation rate increases as the noise level becomes higher (i.e., when σ is larger). When $\sigma = 0$ and the data are not contaminated by noise, the approximation error reaches the highest convergence



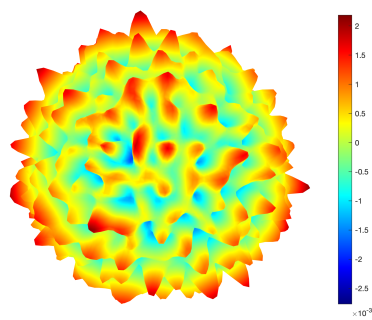
(a) f



(b) f plus noise



(c) $\overline{f_{D,n}^{\diamond}}$



(d) Error

FIG. 1. (a) *Function f in* (5.1)*, which is in Sobolev space* $H^{4.5}(\mathbb{S}^2)$. (b) *Noisy data $f+\epsilon$ with noise level $\sigma = 0.1$.* (c) *Distributed filtered hyperinterpolation $\overline{f_{D,n}^{\diamond}}$ with $n = 25$, $m = 100$, $\sigma = 0.1$ and the noisy data set D given on the nodes of symmetric spherical* 75*-design.* (d) *Error $\overline{f_{D,n}^{\diamond}} - f$.*

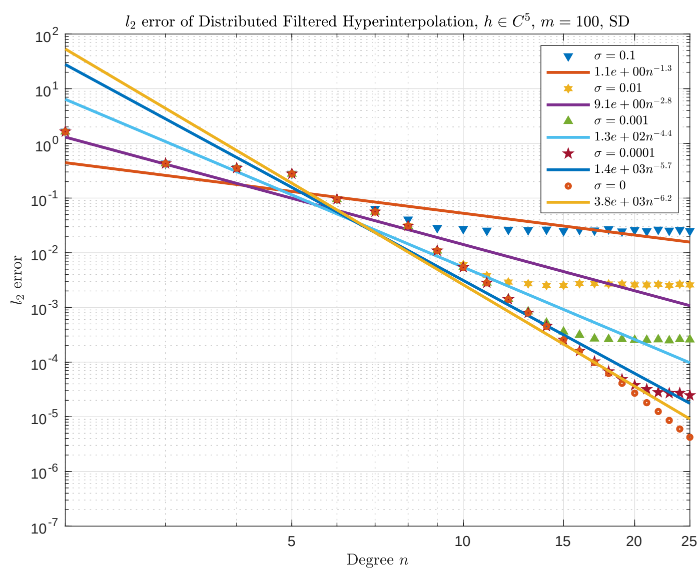


FIG. 2. *Errors vs. degree n for distributed filtered hyperinterpolation, $n \leq 25$, $m = 100$,* $\sigma = 0, 0.0001, 0.001, 0.01, 0.1$.

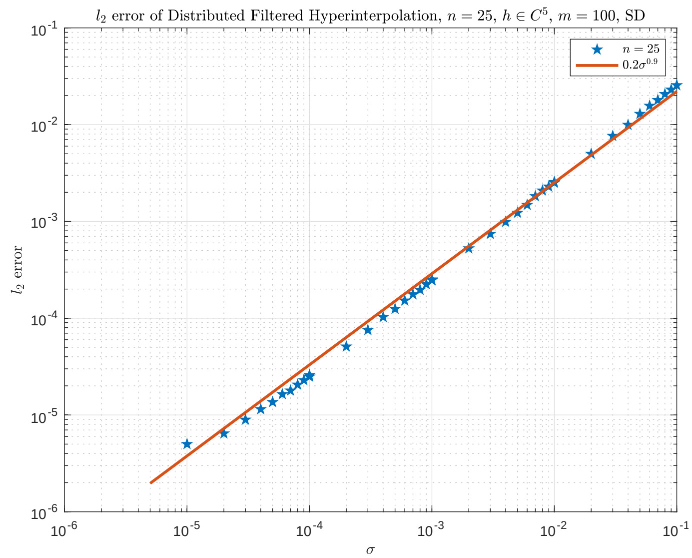


FIG. 3. *Errors vs. standard variance σ for distributed filtered hyperinterpolation, $n = 25$, $m = 100$.*

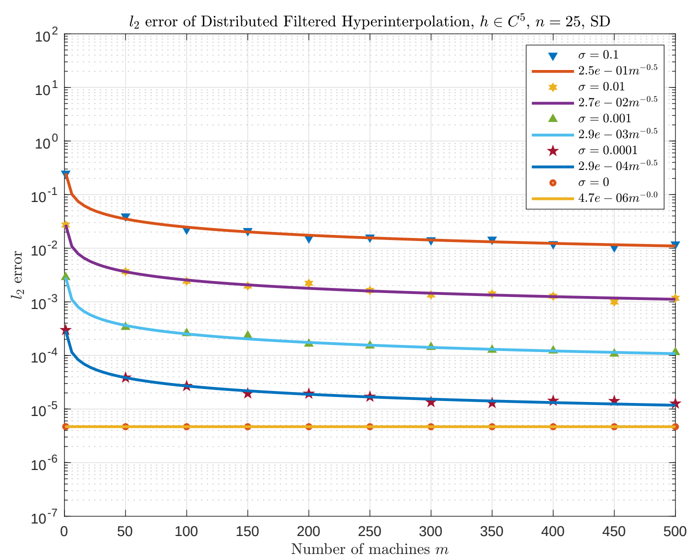


FIG. 4. *Errors vs. number of machines m for distributed filtered hyperinterpolation, $n = 25$, $m \leq 500, \sigma = 0, 0.0001, 0.001, 0.01, 0.1$.*

rate at order $n^{-6.7}$. The convergence rate of the distributed filtered hyperinterpolation is higher than the upper bound of Theorem 3.5 since the function f is sufficiently smooth.

Figure 3 shows that the approximation error of the distributed filtered hyperinterpolation of degree $n = 25$ with $m = 100$ machines converges at the rate $0.2\sigma^{0.9}$ with a decrease of noise level σ. The noise level of data impacts the approximation precision of distributed filtered hyperinterpolation.

Figure 4 illustrates the impact of the number m of machines on approximation capability. As we can see for a noise level $\sigma > 0$, the approximation error converges at a rate of around $m^{-0.5}$. This means that the noise level changes the absolute magnitude of approximation error but has little impact on the trend of approximation rate with the increase of the number of machines.

6. Proofs.

6.1. Proof for Theorem 2.2.

Theorem 2.2 uses the norming set method [35] to derive the probabilistic quadrature rule. To prove Theorem 2.2, we need the following four lemmas. The first one is the Nikolskiî inequality on the sphere, which was proved in [21, 34].

LEMMA 6.1. *Let $1 \le p < q \le \infty$, $n \ge 1$ be an integer, and let $P \in \Pi_n^d$. Then*

$$\|P\|_{L^q(\mathbb{S}^d)} \le \tilde{C}_1 n^{\frac{d}{p}-\frac{d}{q}} \|P\|_{L^p(\mathbb{S}^d)},$$

where the constant $\tilde{C}_1 > 0$ depends only on d, p, q.

The second one is the following concentration inequality, which was established in [51].

LEMMA 6.2. *Let $\mathcal{G}$ be a set of functions on compact metric space Z. For every $g \in \mathcal{G}$, $\mathbf{E}g \ge 0$, $|g - \mathbf{E}g| \le B$ almost everywhere, and $\mathbf{E}(g^2) \le \tilde{c}(\mathbf{E}g)^\alpha$ for some $B \ge 0$, $0 \le \alpha \le 1$ and $\tilde{c} \ge 0$. Then, for any $\varepsilon > 0$,*

$$\mathbf{P}\left\{\sup_{g\in\mathcal{G}} \frac{\left|\mathbf{E}g - \frac{1}{m}\sum_{i=1}^m g(z_i)\right|}{\sqrt{(\mathbf{E}g)^\alpha + \varepsilon^\alpha}} > \varepsilon^{1-\frac{\alpha}{2}}\right\} \le 2\mathcal{N}(\mathcal{G},\varepsilon)\exp\left\{-\frac{m\varepsilon^{2-\alpha}}{2(\tilde{c} + \frac{1}{3}B\varepsilon^{1-\alpha})}\right\},$$

where $\mathcal{N}(\mathcal{G}, \varepsilon)$ denotes the covering number [51] of $\mathcal{G}$ with radius ε.

The third one is a covering number estimate for Banach spaces, as given in [56].

LEMMA 6.3. *Let $\mathbb{B}$ be a finite-dimensional Banach space. Let B_R be the closed ball of radius R centered at the origin given by $B_R := \{f \in \mathbb{B} : \|f\|_{\mathbb{B}} \le R\}$. Then*

$$\log \mathcal{N}(B_R, \varepsilon) \le \dim(\mathbb{B}) \log\left(\frac{4R}{\varepsilon}\right).$$

To state the last lemma, we need following definition.

DEFINITION 6.4. *Let $\mathcal{X}$ be a finite dimensional vector space with norm $\|\cdot\|_{\mathcal{X}}$, and let $\mathcal{Z} \subset \mathcal{X}^*$ be a finite set. We say that $\mathcal{Z}$ is a norm generating set for $\mathcal{X}$ if the mapping $T_{\mathcal{Z}} : \mathcal{X} \to \mathbb{R}^{|\mathcal{Z}|}$ defined by $T_{\mathcal{Z}}(x) = (z(x))_{z\in\mathcal{Z}}$ is injective and $T_{\mathcal{Z}}$ is called sampling operator.*

Let $W := T_{\mathcal{Z}}(\mathcal{X})$ be the range of $T_{\mathcal{Z}}$. Then the injectivity of $T_{\mathcal{Z}}$ implies that $T_{\mathcal{Z}}^{-1} : W \to \mathcal{X}$ exists. Let $\|\cdot\|_{\mathbb{R}^{|\mathcal{Z}|}}$ be the norm of $\mathbb{R}^{|\mathcal{Z}|}$ norm, and let $\|\cdot\|_{\mathbb{R}^{|\mathcal{Z}|^*}}$ be the dual norm on $\mathbf{R}^{|\mathcal{Z}|^*}$ for $\|\cdot\|_{\mathbb{R}^{|\mathcal{Z}|}}$. Equip W with the induced norm, and let $\|T_{\mathcal{Z}}^{-1}\| := \|T_{\mathcal{Z}}^{-1}\|_{W\to\mathcal{X}}$. In addition, let $\mathcal{K}_+$ be the positive cone of $\mathbb{R}^{|\mathcal{Z}|}$, which is the set of all $(r_z)_{z\in\mathcal{Z}} \in \mathbb{R}^{|\mathcal{Z}|}$ such that $r_z \ge 0$. Then the following lemma [35] holds.

LEMMA 6.5. *Let $\mathcal{Z}$ be a norm generating set for $\mathcal{X}$, with $T_{\mathcal{Z}}$ the corresponding sampling operator. If $g \in \mathcal{X}^*$ with $\|g\|_{\mathcal{X}^*} \le \mathcal{A}$, then there exist positive numbers $\{a_z\}_{z\in\mathcal{Z}}$ depending only on g such that for every $x \in \mathcal{X}$,*

$$g(x) = \sum_{z\in\mathcal{Z}} a_z z(x), \quad \|(a_z)\|_{\mathbb{R}^{|\mathcal{Z}|^*}} \le \mathcal{A}\|T_{\mathcal{Z}}^{-1}\|.$$

Also, if W contains an interior point $v_0 \in \mathcal{K}_+$ and if $g(T_{\mathcal{Z}}^{-1}v) \ge 0$ when $v \in W \cap \mathcal{K}_+$, then we may choose $a_z \ge 0$.

We are ready to prove Theorem 2.2.

Proof of Theorem 2.2. For $p = 1, 2$, without loss of generality, we prove Theorem 2.2 for $P_n \in \Pi_n^d$ satisfying $\|P_n\|_{p,\mu} = A$ for some constant $A > 0$. For an arbitrary $P_n \in \Pi_n^d$ with $\|P_n\|_{p,\mu} = A$, it follows from (2.5) and Lemma 6.1 that

$$\begin{aligned}
&\|P_n\|_{L_\infty(\mathbb{S}^d)} \leq \tilde{C}_1 n^{\frac{d}{p}} \|P_n\|_{L_p(\mathbb{S}^d)} \leq c_4^{1/p} \tilde{C}_1 n^{\frac{d}{p}} \|P_n\|_{p,\mu} \quad \text{and} \\
&\mathbf{E}\left\{|P_n|^{2p}\right\} = \int_{\mathbb{S}^d} |P_n(\mathbf{x})|^{2p} \mathrm{d}\mu(\mathbf{x}) \leq \|P_n\|_{L_\infty(\mathbb{S}^d)}^p \int_{\mathbb{S}^d} |P_n(\mathbf{x})|^p \mathrm{d}\mu(\mathbf{x}) \\
&\qquad \leq c_4 (\tilde{C}_1)^p n^d \|P_n\|_{p,\mu}^p \mathbf{E}\left[|P_n|^p\right].
\end{aligned}$$

Let $Z = \mathbb{S}^d$, $g(z_i) = |P_n(\mathbf{x}_i)|^p$, $B = 2c_4(\tilde{C}_1)^p n^d A^p$, $\tilde{c} = c_4(\tilde{C}_1)^p n^d A^p$, $m = N$, $\alpha = 1$, and $\mathcal{G}_p = \{|P_n|^p : P_n \in \Pi_n^d, \|P_n\|_{p,\mu} = A\}$ in Lemma 6.2. Then, for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbf{P}\left\{ \sup_{P_n \in \Pi_n^d, \|P_n\|_{p,\mu} = A} \frac{\left| \|P_n\|_{p,\mu}^p - \frac{1}{N}\sum_{i=1}^N |P_n(\mathbf{x}_i)|^p \right|}{\sqrt{\|P_n\|_{p,\mu}^p + \varepsilon}} > \sqrt{\varepsilon} \right\} \\
&\quad \leq 2\mathcal{N}(\mathcal{G}_p, \varepsilon) \exp\left\{ -\frac{N\varepsilon}{\tilde{C}_2 n^d A^p} \right\},
\end{aligned}$$

where $\tilde{C}_2 = 10c_4(\tilde{C}_1)^p/3$. For $p = 1$, we have $|P_n| - |P_n^*| \leq |P_n - P_n^*|$ for any $P_n, P_n^* \in \Pi_n^d$. Then it follows from the definition of the covering number that $\mathcal{N}(\mathcal{G}_1, \varepsilon) \leq \mathcal{N}(\mathcal{G}_1', \varepsilon)$, where $\mathcal{G}_1' := \{P_n \in \Pi_n^d : \|P_n\|_{p,\mu} = A\}$. For $p = 2$, $|P_n|^2 \in \Pi_{2n}^d$. Let $\mathcal{G}_2' := \{P_n \in \Pi_{2n}^d : \|P_n\|_{p,\mu} = A\}$. Then $\mathcal{N}(\mathcal{G}_2, \varepsilon) = \mathcal{N}(\mathcal{G}_2', \varepsilon)$. It then follows from Lemma 6.3 that for $p = 1, 2$,

$$\begin{aligned}
&\mathbf{P}\left\{ \sup_{P_n \in \Pi_n^d, \|P_n\|_{p,\mu} = A} \frac{\left| \|P_n\|_{p,\mu}^p - \frac{1}{N}\sum_{i=1}^N |P_n(\mathbf{x}_i)|^p \right|}{\sqrt{\|P_n\|_{p,\mu}^p + \varepsilon}} > \sqrt{\varepsilon} \right\} \\
&\quad \leq 2\exp\left\{ (2n)^d \log\frac{4A^p}{\varepsilon} - \frac{N\varepsilon}{\tilde{C}_2 n^d A^p} \right\},
\end{aligned}$$

where we use $\dim \mathcal{G}_p \leq (pn)^d$ for $p = 1, 2$. Let $\varepsilon = A^p/4$. As $N/n^d > c$ for a sufficiently large $c > 0$, $(2n^d) \log \frac{4A^p}{\varepsilon} < \frac{N\varepsilon}{\tilde{C}_2 n^d A^p}$. Then, with confidence

$$1 - 2\exp\left\{ -\tilde{C}_3 N/n^d \right\}, \tag{6.1}$$

there holds

$$\left| \|P_n\|_{p,\mu}^p - \frac{1}{N}\sum_{i=1}^N |P_n(\mathbf{x}_i)|^p \right| \leq \sqrt{\varepsilon(\|P_n\|_{p,\mu}^p + \varepsilon)} = \frac{\sqrt{5}}{4}\|P_n\|_{p,\mu}^p.$$

Then, with the same confidence as (6.1),

$$\frac{1}{3}\|P_n\|_{p,\mu}^p \leq \frac{1}{N}\sum_{i=1}^N |P_n(\mathbf{x}_i)|^p \leq \frac{5}{3}\|P\|_{p,\mu}^p \quad \forall P_n \in \Pi_n^d,\ p = 1, 2. \tag{6.2}$$

Now we use (6.2) with $p = 2$ and Lemma 6.5 to prove Theorem 2.2. In Lemma 6.5, we take $\mathcal{X} = \Pi_n^d$, $\|P_n\|_{\mathcal{X}} = \|P_n\|_{2,\mu}$, and $\mathcal{Z}$ to be the set of point evaluation functionals $\{\delta_{\mathbf{x}_i}\}_{i=1}^N$. The operator $T_{\mathcal{Z}}$ is then the restriction map $P_n \mapsto P_n|_{X_N}$ and $\|f\|_{X_N,2} := (\frac{1}{N}\sum_{i=1}^N |f(\mathbf{x}_i)|^2)^{\frac{1}{2}}$. It follows from (6.2) that with confidence at least $1 - 2\exp\{-\tilde{C}_3 N/n^d\}$, there holds $\|T_{\mathcal{Z}}^{-1}\| \leq \sqrt{\frac{5}{3}}$. We take g to be the functional $g : P_n \mapsto \int_{\mathbb{S}^d} P_n(x)\mathrm{d}\mu(x)$. By the Hölder inequality, $\|g\|_{\mathcal{X}^*} \leq 1$. Lemma 6.5 then shows that there exists a set of real numbers $\{w_{i,n}\}_{i=1}^N$ such that

$$\int_{\mathbb{S}^d} P_n(x)\mathrm{d}\mu(x) = \sum_{i=1}^N w_{i,n} P_n(\mathbf{x}_i) \quad \forall P_n \in \Pi_n^d.$$

holds and $\frac{1}{N}\sum_{i=1}^N (\frac{w_{i,n}}{1/N})^2 \leq 2$ with confidence at least $1 - 2\exp\{-\tilde{C}_3 N/n^d\}$.

Finally, we use the second assertion of Lemma 6.5 and (6.2) with $p = 1$ to prove the positivity of $w_{i,n}$. Since $1 \in \Pi_n^d$, we have that $v_0 := 1|_{X_N} = (1, 1, \ldots, 1)$ is an interior point of $\mathcal{K}_+$. For $P_n \in \Pi_n^d$, $T_{\mathcal{Z}} P_n = P_n|_{X_N}$ is in $W \cap \mathcal{K}_+$ if and only if $P_n(\mathbf{x}_i) \geq 0$ for all $\mathbf{x}_i \in X_N$. For an arbitrary P_n satisfying $P_n(\mathbf{x}_i) \geq 0$ with $\mathbf{x}_i \in X_N$, define $\xi_i(P_n) = P_n(\mathbf{x}_i)$. From Lemma 6.1 and (2.5), we obtain the following estimates: For $i = 1, \ldots, N$,

$$\begin{aligned}
|\xi_i| &\leq \|P_n\|_{L_\infty(\mathbb{S}^d)} \leq \tilde{C}_1 n^d \|P_n\|_{L_1(\mathbb{S}^d)} \leq \tilde{C}_1 c_4 n^d \|P_n\|_{1,\mu}, \\
|\xi_i - \mathbf{E}\xi_i| &\leq 2\|P_n\|_{L_\infty(\mathbb{S}^d)} \leq 2\tilde{C}_1 c_4 n^d \|P_n\|_{1,\mu}, \\
\mathbf{E}\xi_i^2 &\leq \|P_n\|_{L_\infty(\mathbb{S}^d)} \|P_n\|_{1,\mu} \leq \tilde{C}_1 c_4 n^d \|P_n\|_{1,\mu}^2.
\end{aligned}$$

Applying Lemma 6.2 with $B = 2\tilde{C}_1 c_4 n^d A$, $\tilde{c} = \tilde{C}_1 c_4 n^d A^2$, and $\alpha = 0$ to the set $\{P_n : P_n \in \Pi_n^d, \|P_n\|_{1,\mu} = A\}$, by Lemma 6.3, we obtain for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{P_n \in \Pi_n^d, P_n|_{X_N} \geq 0, \|P_n\|_{1,\mu} = A} \left| g(P_n) - \frac{1}{N}\sum_{i=1}^N P_n(\mathbf{x}_i)\right| > \varepsilon\right\} \\
&\quad \leq 2\exp\left\{n^d \log\frac{4A}{\varepsilon} - \frac{N\varepsilon^2}{2\tilde{C}_1 c_4 n^d A(A + 2\varepsilon/3)}\right\}.
\end{aligned}$$

Let $\varepsilon = A/4$. We then obtain that with confidence $1 - 2\exp\{-\tilde{C}_4 N/n^d\}$, there holds

$$\left| g(P_n) - \frac{1}{N}\sum_{i=1}^N P_n(\mathbf{x}_i)\right| \leq \frac{1}{4}\|P_n\|_{1,\mu} \quad \forall P_n \in \Pi_n^d.$$

This and (6.2) imply that, with confidence $1 - 4\exp\{-CN/n^d\}$ (where C depends only on $\tilde{C}_3$ and $\tilde{C}_4$), for any P_n satisfying $P_n(\mathbf{x}_i) \geq 0$ $\forall \mathbf{x}_i \in X_N$, the inequality

$$\left| g(P_n) - \frac{1}{N}\sum_{i=1}^N P_n(\mathbf{x}_i)\right| \leq \frac{3}{4}\frac{1}{N}\sum_{i=1}^N P_n(\mathbf{x}_i)$$

holds and then

$$g(P_n) \geq \frac{1}{4}\frac{1}{N}\sum_{i=1}^N P_n(\mathbf{x}_i) \geq 0.$$

It hence follows from Lemma 6.5 that $w_{i,n} \geq 0$, thus completing the proof of Theorem 2.2. $\square$

6.2. Proofs for the theorems in section 3. The following lemma shows that the filtered kernel has the following localization property, as proved by [38] and also [46, 47, 48].

LEMMA 6.6 ([38]). *Let $d \geq 2$ and η be a filter in $C^{\kappa}(\mathbb{R}_+)$ with $1 \leq \kappa < \infty$ such that η is constant on $[0,a]$ for some $0 < a < 2$. Then*

$$|K_n(\cos\theta)| \leq \frac{cn^d}{(1+n\theta)^{\kappa}}, \quad n \geq 1,$$

where c is a constant depending only on d, η, and κ.

Lemma 6.6 gives the following upper bound of the L_p norm of the filtered kernel.

LEMMA 6.7. *Let $d \geq 2$, $1 \leq p \leq \infty$, and η be a filter in $C^{\kappa}(\mathbb{R}_+)$ with $\kappa \geq \lfloor \frac{d+3}{2} \rfloor$ such that η is constant on $[0,a]$ for some $0 < a < 2$. Then*

$$\|K_n(\mathbf{x}\cdot\cdot)\|_{L_p(\mathbb{S}^d)} \leq c_1 n^{d(1-1/p)} \quad \forall \mathbf{x} \in \mathbb{S}^d,\ n \geq 1,$$

where c_1 is a constant depending only on d, η, κ, and p.

The above lemma for $p = 1$ was proven in [47] (see also [38] for $\kappa \geq d+1$). The case $p > 1$ can be obtained from the case $p = 1$ with the fact that $K_n \in \Pi_{2n}^d$ and the Nikolskiî inequality for spherical polynomials [34].

Proof of Theorem 3.2. Define

$$f^{\diamond,*}_{D,n}(x) := \sum_{i=1}^{|D|} w_{i,s,D} f^*(\mathbf{x}_i) K_n(\mathbf{x}_i \cdot \mathbf{x}). \tag{6.3}$$

As $\mathbf{E}\{\epsilon_i\} = 0$ for any $i = 1, \ldots, |D|$,

$$\begin{aligned}\mathbf{E}\left\{f^{\diamond}_{D,n}(x)\right\} &= \mathbf{E}\left\{\sum_{i=1}^{m} w_{i,s,D} y_i K_n(\mathbf{x}_i\cdot\mathbf{x})\right\} = \mathbf{E}\left\{\sum_{i=1}^{m} w_{i,s,D}(f^*(\mathbf{x}_i)+\epsilon_i) K_n(\mathbf{x}_i\cdot\mathbf{x})\right\}\\ &= \sum_{i=1}^{m} w_{i,s,D} f^*(\mathbf{x}_i) K_n(\mathbf{x}_i\cdot\mathbf{x}) + \sum_{i=1}^{m} w_{i,s,D}\mathbf{E}\{\epsilon_i\} K_n(\mathbf{x}_i\cdot\mathbf{x}) = f^{\diamond,*}_{D,n}(x).\end{aligned}$$

Then

$$\mathbf{E}\left\{f^{\diamond,*}_{D,n}(x) - f^{\diamond}_{D,n}(x)\right\} = 0. \tag{6.4}$$

This implies

$$\begin{aligned}&\mathbf{E}\left\{\|f^{\diamond}_{D,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\ &= \int_{\mathbb{S}^d} \mathbf{E}\{(f^*(x) - f^{\diamond}_{D,n}(x))^2\} \mathrm{d}\omega(\mathbf{x})\\ &= \int_{\mathbb{S}^d} \mathbf{E}\{(f^*(x) - f^{\diamond,*}_{D,n}(x) + f^{\diamond,*}_{D,n}(x) - f^{\diamond}_{D,n}(x))^2\} \mathrm{d}\omega(\mathbf{x})\\ &= \int_{\mathbb{S}^d} (f^{\diamond,*}_{D,n}(x) - f^*(x))^2 \mathrm{d}\omega(\mathbf{x}) + \int_{\mathbb{S}^d} \mathbf{E}\{(f^{\diamond,*}_{D,n}(x) - f^{\diamond}_{D,n}(x))^2\} \mathrm{d}\omega(\mathbf{x})\\ &:= \mathcal{A}^{\diamond}_{D,n} + \mathcal{S}^{\diamond}_{D,n}.\end{aligned} \tag{6.5}$$

Lemma 2.4 gives

$$\mathcal{A}^{\diamond}_{D,n} \leq c_5^2\, n^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}. \tag{6.6}$$

To bound $\mathcal{S}^{\diamond}_{D,n}$, we observe from (3.1) that

$$\begin{aligned}
\mathbf{E}\left\{(f^{\diamond,*}_{D,n}(\mathbf{x}) - f^{\diamond}_{D,n}(\mathbf{x}))^2\right\} &= \mathbf{E}\left\{\left(\sum_{i=1}^{|D|}(y_i - f^*(\mathbf{x}_i))w_{i,s,D}K_n(\mathbf{x}_i\cdot\mathbf{x})\right)^2\right\}\\
&= \mathbf{E}\left\{\left(\sum_{i=1}^{|D|}\epsilon_i w_{i,s,D}K_n(\mathbf{x}_i\cdot\mathbf{x})\right)^2\right\}\\
&\leq M^2\sum_{i=1}^{|D|}w_{i,s,D}^2|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2,
\end{aligned}$$

where the last inequality uses the independence of $\epsilon_1,\ldots,\epsilon_{|D|}$. This together with Lemmas 6.7 and 2.1 shows

$$\begin{aligned}
\mathcal{S}^{\diamond}_{D,n} &\leq M^2\int_{\mathbb{S}^d}\sum_{i=1}^{|D|}w_{i,s,D}^2|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2\mathrm{d}\omega(\mathbf{x})\\
&= M^2\sum_{i=1}^{|D|}w_{i,s,D}^2\int_{\mathbb{S}^d}|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2\mathrm{d}\omega(\mathbf{x}) \leq c_1M^2n^d\sum_{i=1}^{|D|}w_{i,s,D}^2 \leq \frac{c_1c_2^2M^2n^d}{|D|}. \qquad (6.7)
\end{aligned}$$

Putting (6.7) and (6.6) to (6.5), we obtain

$$\mathbf{E}\left\{\|f^{\diamond}_{D,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\right\} \leq c_5^2n^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)} + \frac{c_1c_2^2M^2n^d}{|D|}, \tag{6.8}$$

with $\frac{c_3}{6}|D|^{\frac{1}{2r+d}} \leq n \leq \frac{c_3}{3}|D|^{\frac{1}{2r+d}}$. Then $\mathbf{E}\{\|f^{\diamond}_{D,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\} \leq C_1|D|^{-\frac{2r}{2r+d}}$ with $C_1 := 36^rc_5^2c_3^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)} + 3^{-d}c_1c_2^2c_3^dM^2$, thus completing the proof. ∎

To prove Theorem 3.5, we need the following lemma, which is a modified version of [16, Proposition 4].

LEMMA 6.8. *For* $\overline{f^{\diamond}_{D,n}}$ *in Definition* 3.4, *there holds*

$$\begin{aligned}
&\mathbf{E}\left\{\left\|\overline{f^{\diamond}_{D,n}} - f^*\right\|^2_{L_2(\mathbb{S}^d)}\right\}\\
&\leq \sum_{j=1}^m\frac{|D_j|^2}{|D|^2}\mathbf{E}\left\{\|f^{\diamond}_{D_j,n} - f^*\|^2_{L_2(\mathbb{S}^d)}\right\} + \sum_{j=1}^m\frac{|D_j|}{|D|}\left\|f^{\diamond,*}_{D_j,n} - f^*\right\|^2_{L_2(\mathbb{S}^d)}, \qquad (6.9)
\end{aligned}$$

where $f^{\diamond,*}_{D_j,n}$ *is given by* (6.3).

Proof. Due to (3.4) and $\sum_{j=1}^m \frac{|D_j|}{|D|}=1$, we have

$$\begin{aligned}\|\overline{f^\diamond_{D,n}}-f^*\|^2_{L_2(\mathbb{S}^d)} &= \left\|\sum_{j=1}^m \frac{|D_j|}{|D|}(f^\diamond_{D_j,n}-f^*)\right\|^2_{L_2(\mathbb{S}^d)}\\ &=\sum_{j=1}^m \frac{|D_j|^2}{|D|^2}\|f^\diamond_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\\ &\quad+\sum_{j=1}^m \frac{|D_j|}{|D|}\left\langle f^\diamond_{D_j,n}-f^*, \sum_{k\neq j}\frac{|D_k|}{|D|}(f^\diamond_{D_k,n}-f^*)\right\rangle_{L_2(\mathbb{S}^d)}.\end{aligned}$$

Taking expectations gives

$$\begin{aligned}&\mathbf{E}\left\{\|\overline{f^\diamond_{D,n}}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\ &\quad=\sum_{j=1}^m \frac{|D_j|^2}{|D|^2}\mathbf{E}\left\{\|f^\diamond_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\ &\qquad+\sum_{j=1}^m \frac{|D_j|}{|D|}\left\langle \mathbf{E}_{D_j}\{f^\diamond_{D_j,n}\}-f^*, \mathbf{E}\left\{\overline{f^\diamond_{D,n}}\right\}-f^*-\frac{|D_j|}{|D|}\left(\mathbf{E}_{D_j}\{f^\diamond_{D_j,n}\}-f^*\right)\right\rangle_{L_2(\mathbb{S}^d)},\end{aligned}$$

where

$$\begin{aligned}&\sum_{j=1}^m \frac{|D_j|}{|D|}\left\langle \mathbf{E}_{D_j}\{f^\diamond_{D_j,n}\}-f^*, \mathbf{E}\left\{\overline{f^\diamond_{D,n}}\right\}-f^*\right\rangle_{L_2(\mathbb{S}^d)}\\ &\quad=\mathbf{E}\left\{\left\langle \overline{f^\diamond_{D,n}}-f^*, \mathbf{E}\left\{\overline{f^\diamond_{D,n}}\right\}-f^*\right\rangle_{L_2(\mathbb{S}^d)}\right\}=\left\|\mathbf{E}\left\{\overline{f^\diamond_{D,n}}\right\}-f^*\right\|^2_{L_2(\mathbb{S}^d)}.\end{aligned}$$

Then

$$\begin{aligned}&\mathbf{E}\left\{\|\overline{f^\diamond_{D,n}}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\ &\quad=\sum_{j=1}^m \frac{|D_j|^2}{|D|^2}\mathbf{E}\left\{\|f^\diamond_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\ &\qquad-\sum_{j=1}^m \frac{|D_j|^2}{|D|^2}\left\|\mathbf{E}\{f^\diamond_{D_j,n}\}-f^*\right\|^2_{L_2(\mathbb{S}^d)}+\left\|\mathbf{E}\left\{\overline{f^\diamond_{D,n}}\right\}-f^*\right\|^2_{L_2(\mathbb{S}^d)}.\end{aligned}$$

By (6.4), $\mathbf{E}\{\overline{f^\diamond_{D,n}}\}=\sum_{j=1}^m \frac{|D_j|}{|D|}f^{\diamond,*}_{D_j,n}$. This plus $\sum_{j=1}^m \frac{|D_j|}{|D|}=1$ gives

$$\begin{aligned}\left\|\mathbf{E}\left\{\overline{f^\diamond_{D,n}}\right\}-f^*\right\|^2_{L_2(\mathbb{S}^d)} &= \left\|\sum_{j=1}^m \frac{|D_j|}{|D|}\left(f^{\diamond,*}_{D_j,n}-f^*\right)\right\|^2_{L_2(\mathbb{S}^d)}\\ &\leq \sum_{j=1}^m \frac{|D_j|}{|D|}\|f^{\diamond,*}_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)},\end{aligned}$$

thus proving the bound in (6.9). ∎

Proof of Theorem 3.5. By Lemma 6.8, we only need to estimate the bounds of $\mathbf{E}\{\|f^{\diamond}_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\}$ and $\|f^{\diamond,*}_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}$. Since $\min_{j=1,\dots,m}|D_j|\geq|D|^{d/(2r+d)}$ and D_j is τ-quasi uniform, it follows from Lemma 2.1 that there exists a quadrature rule for each local machine which is exact for polynomials of degree $3n-1$ for $n\leq\frac{c_3}{3}|D|^{1/(2r+d)}$. From (6.8) with $D=D_j$, for $j=1,\dots,m$,

$$\mathbf{E}\left\{\|f^{\diamond}_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\leq c_5^2n^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+\frac{c_1c_2^2M^2n^d}{|D_j|}.$$

This together with $\sum_{i=1}^m\frac{|D_j|}{|D|}=1$ gives

$$\begin{aligned}\sum_{j=1}^m\frac{|D_j|^2}{|D|^2}\mathbf{E}&\left\{\|f^{\diamond}_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\&\leq 36^rc_5^2c_3^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}|D|^{-\frac{2r}{2r+d}}\\&\quad+3^{-d}c_1c_2^2c_3^dM^2\sum_{j=1}^m\frac{|D_j|^2}{|D|^2}\frac{|D|^{\frac{d}{2r+d}}}{|D_j|}=C_1|D|^{-\frac{2r}{2r+d}},\end{aligned}\tag{6.10}$$

where $C_1:=36^rc_5^2c_3^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+3^{-d}c_1c_2^2c_3^dM^2$.

For each $j=1,\dots,m$, D_j is τ-quasi uniform. Lemma 2.1 implies that there exists a quadrature rule with nodes of D_j and $|D_j|$ positive weights such that $f^*_{D_j,n}$ is a filtered hyperinterpolation for the noise-free data set $\{\mathbf{x}_i,f^*(\mathbf{x}_i)\}_{\mathbf{x}_i\in D_j}$. Lemma 2.4 then gives

$$\left\|f^{\diamond,*}_{D_j,n}-f^*\right\|^2_{L_2(\mathbb{S}^d)}\leq c_5^2n^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}\quad\forall j=1,2,\dots,m.$$

This together with $\sum_{j=1}^m\frac{|D_j|}{|D|}=1$ and $\frac{c_3}{6}|D|^{\frac{1}{2r+d}}\leq n\leq\frac{c_3}{3}|D|^{\frac{1}{2r+d}}$ gives

$$\sum_{j=1}^m\frac{|D_j|}{|D|}\left\|f^{\diamond,*}_{D_j,n}-f^*\right\|^2_{L_2(\mathbb{S}^d)}\leq 36^rc_5^2c_3^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}|D|^{-\frac{2r}{2r+d}}.\tag{6.11}$$

Using (6.10) and (6.11) in Lemma 6.8,

$$\mathbf{E}\left\{\|\overline{f^{\diamond}_{D,n}}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\leq C_2|D|^{-\frac{2r}{2r+d}}.$$

This then proves (3.5) with $C_2=2^{2r+1}\cdot9^rc_5^2c_3^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+3^{-d}c_1c_2^2c_3^dM^2$. ◻

6.3. Proofs for the theorems in section 4.

Proof of Theorem 4.2. Let $\{w_{i,n,D}\}_{i=1}^{|D|}$ be the real numbers computed in (4.1). Since $\{\mathbf{x}_i\}_{i=1}^{|D|}$ is a set of random points on $\mathbb{S}^d$, we define four events as follows. Let Ω_D be the event such that $\sum_{i=1}^{|D|}|w_{i,n,D}|^2\leq\frac{2}{|D|}$ and Ω_D^c be the complement of Ω_D, i.e., Ω_D^c be the event $\sum_{i=1}^{|D|}|w_{i,n,D}|^2>\frac{2}{|D|}$. Let Ξ_D be the event that $\{(w_{i,n,D},\mathbf{x}_i)\}_{i=1}^{|D|}$ is a quadrature rule exact for polynomials in Π_n^d and Ξ_D^c be the complement event of Ξ_D. Then, by Theorem 2.2,

$$\mathbf{P}\{\Omega_D^c\}\leq\mathbf{P}\{\Xi_D^c\}\leq4\exp\left\{-C|D|/n^d\right\}.\tag{6.12}$$

We write

$$\mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\} = \mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega_D\right\}\mathbf{P}\{\Omega_D\} + \mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega_D^c\right\}\mathbf{P}\{\Omega_D^c\}. \tag{6.13}$$

Under the event Ω_D^c, we have from (4.1) and (4.2) that $f_{D,n}(x)=0$. Then, by (6.12),

$$\mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega_D^c\right\}\mathbf{P}\{\Omega_D^c\} \leq 4\|f^*\|^2_{L_\infty(\mathbb{S}^d)}\exp\bigl\{-C|D|/n^d\bigr\}. \tag{6.14}$$

Now we estimate the first term of the right-hand side of (6.13) when the event Ω_D takes place. Under this circumstance, we let

$$f^*_{D,n}(\mathbf{x}) := \sum_{i=1}^{|D|} w_{i,n,D} f^*(\mathbf{x}_i) K_n(\mathbf{x}_i\cdot\mathbf{x}). \tag{6.15}$$

Let $\Lambda_D := \{\mathbf{x}_i\}_{i=1}^{|D|}$. By the independence between $\{\epsilon_i\}_{i=1}^{|D|}$ and Λ_D and $\mathbf{E}\{\epsilon_i\}=0$, $i=1,\dots,|D|$, we obtain

$$\begin{aligned}
\mathbf{E}\left\{f_{D,n}(\mathbf{x})\big|\Lambda_D\right\} &= \mathbf{E}\left\{\sum_{i=1}^m w_{i,n,D} y_i K_n(\mathbf{x}_i\cdot\mathbf{x})\big|\Lambda_D\right\}\\
&= \mathbf{E}\left\{\sum_{i=1}^m w_{i,n,D}(f^*(\mathbf{x}_i)+\epsilon_i) K_n(\mathbf{x}_i\cdot\mathbf{x})\big|\Lambda_D\right\}\\
&= \sum_{i=1}^m w_{i,n,D} f^*(\mathbf{x}_i) K_n(\mathbf{x}_i\cdot\mathbf{x}) + \sum_{i=1}^m w_{i,n,D}\mathbf{E}\{\epsilon_i\} K_n(\mathbf{x}_i\cdot\mathbf{x})\\
&= f^*_{D,n}(\mathbf{x}).
\end{aligned}$$

Hence,

$$\mathbf{E}\left\{\left(f^*_{D,n}(\mathbf{x})-f_{D,n}(\mathbf{x})\right)\big|\Lambda_D\right\} = 0. \tag{6.16}$$

This allows us to write

$$\begin{aligned}
&\mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\big|\Omega_D\right\}\\
&= \mathbf{E}\left\{\int_{\mathbb{S}^d}\mathbf{E}\{(f^*(\mathbf{x})-f_{D,n}(\mathbf{x}))^2\big|\Lambda_D\}\mathrm{d}\omega(\mathbf{x})\big|\Omega_D\right\}\\
&= \mathbf{E}\left\{\int_{\mathbb{S}^d}\mathbf{E}\{(f^*(\mathbf{x})-f^*_{D,n}(\mathbf{x})+f^*_{D,n}(\mathbf{x})-f_{D,n}(\mathbf{x}))^2\big|\Lambda_D\}\mathrm{d}\omega(\mathbf{x})\big|\Omega_D\right\}\\
&= \mathbf{E}\left\{\int_{\mathbb{S}^d}\mathbf{E}\{(f^*_{D,n}(\mathbf{x})-f_{D,n}(\mathbf{x}))^2\big|\Lambda_D\}\mathrm{d}\omega(\mathbf{x})\big|\Omega_D\right\}\\
&\quad + \mathbf{E}\left\{\int_{\mathbb{S}^d}\mathbf{E}\{(f^*_{D,n}(\mathbf{x})-f^*(\mathbf{x}))^2\big|\Lambda_D\}\mathrm{d}\omega(\mathbf{x})\big|\Omega_D\right\}\\
&:= \mathcal{S}_{D,n}+\mathcal{A}_{D,n}.
\end{aligned} \tag{6.17}$$

Given Λ_D, if the event Ω_D occurs, by $|\epsilon_i| \leq M$,

$$\begin{aligned}\mathbf{E}\left\{(f^*_{D,n}(\mathbf{x}) - f_{D,n}(\mathbf{x}))^2\middle|\Lambda_D\right\} &= \mathbf{E}\left\{\left(\sum_{i=1}^{|D|}\epsilon_i w_{i,n,D} K_n(\mathbf{x}_i\cdot\mathbf{x})\right)^2\middle|\Lambda_D\right\}\\ &\leq M^2\sum_{i=1}^{|D|} w_{i,n,D}^2|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2,\end{aligned}$$

where we used the independence of $\epsilon_1,\dots,\epsilon_{|D|}$. This with Lemma 6.7 shows

$$\begin{aligned}\mathcal{S}_{D,n} &\leq M^2\mathbf{E}\left\{\int_{\mathbb{S}^d}\sum_{i=1}^{|D|} w_{i,n,D}^2|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2\mathrm{d}\omega(\mathbf{x})\middle|\Omega_D\right\}\\ &= M^2\mathbf{E}\left\{\sum_{i=1}^{|D|} w_{i,n,D}^2\int_{\mathbb{S}^d}|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2\mathrm{d}\omega(\mathbf{x})\middle|\Omega_D\right\}\\ &\leq c_1^2M^2n^d\mathbf{E}\left\{\sum_{i=1}^{|D|} w_{i,n,D}^2\right\} \leq \frac{2c_1^2M^2n^d}{|D|}. \qquad (6.18)\end{aligned}$$

We now turn to bound $\mathcal{A}_{D,n}$. We split $\mathcal{A}_{D,n}$ as

$$\begin{aligned}\mathcal{A}_{D,n} &= \mathbf{E}\left\{\int_{\mathbb{S}^d}\mathbf{E}\left\{(f^*(\mathbf{x}) - f^*_{D,n}(\mathbf{x}))^2\middle|\Lambda_D\right\}\mathrm{d}\omega(\mathbf{x})\middle|\Xi_D,\Omega_D\right\}\mathbf{P}\{\Xi_D\}\\ &\quad+\mathbf{E}\left\{\int_{\mathbb{S}^d}\mathbf{E}\left\{(f^*(\mathbf{x}) - f^*_{D,n}(\mathbf{x}))^2\middle|\Lambda_D\right\}\mathrm{d}\omega(\mathbf{x})\middle|\Xi_D^c,\Omega_D\right\}\mathbf{P}\{\Xi_D^c\}\\ &:= \mathcal{A}_{D,n,1}+\mathcal{A}_{D,n,2}. \qquad (6.19)\end{aligned}$$

To estimate $\mathcal{A}_{D,n,2}$ given the event $\Omega_D\cap\Xi_D^c$, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\left(f^*(\mathbf{x}) - f^*_{D,n}(\mathbf{x})\right)^2 &\leq 2\|f^*\|^2_{L_\infty(\mathbb{S}^d)} + 2\left|\sum_{i=1}^{|D|} w_{i,n,D}f^*(\mathbf{x}_i)K_n(\mathbf{x}_i\cdot\mathbf{x})\right|^2\\ &\leq 2\|f^*\|^2_{L_\infty(\mathbb{S}^d)} + 2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}\sum_{i=1}^{|D|}a_{i,n,D}^2\sum_{i=1}^{|D|}|K_n(\mathbf{x}_i\cdot\mathbf{x})|^2,\end{aligned}$$

which, with (6.12) and Lemma 6.7, gives

$$\mathcal{A}_{D,n,2} \leq 2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}(|\mathbb{S}^d| + 2c_1^2n^d)\exp\left\{-C|D|/n^d\right\}. \qquad (6.20)$$

To bound $\mathcal{A}_{D,n,1}$, we observe that when the event $\Omega_D\cap\Xi_D$ takes place, $\{w_{i,n,D}\}_{i=1}^{|D|}$ is a set of positive weights for quadrature rule $\mathcal{Q}_{|D|,n}$. We then obtain from Lemma 2.4 and $f^*\in\mathbb{W}_2^r(\mathbb{S}^d)$ with $r > d/2$ that

$$\mathcal{A}_{D,n,1} \leq c_5^2n^{-2r}\|f\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}. \qquad (6.21)$$

By (6.21), (6.20), and (6.19), we obtain

$$\mathcal{A}_{D,n} \leq c_5^2n^{-2r}\|f^*\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)} + 2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}(|\mathbb{S}^d| + 2c_1^2n^d)\exp\{-C|D|/n^d\}. \qquad (6.22)$$

This and (6.18) and (6.17) give

$$\begin{aligned}
&\mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\big|\Omega_D\right\}\\
&\leq c_5^2 n^{-2r}\|f\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}(|\mathbb{S}^d|+2c_1^2n^d)\exp\{-C|D|/n^d\}+\frac{2c_1^2M^2n^d}{|D|}.
\end{aligned}$$

Putting the above estimate and (6.14) into (6.13), we obtain

$$\begin{aligned}
\mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\} &\leq c_5^2 n^{-2r}\|f\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+\frac{2c_1^2M^2n^d}{|D|}\\
&\quad+2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}(|\mathbb{S}^d|+2c_1^2n^d+2)\exp\{-C|D|/n^d\}.
\end{aligned}\tag{6.23}$$

Taking account of $\frac{c_3}{6}|D|^{\frac{1}{2r+d}}\leq n\leq\frac{c_3}{3}|D|^{\frac{1}{2r+d}}$ and $r>d/2$, we then have

$$n^d\exp\left\{-C|D|/n^d\right\}\leq\left(\frac{c_3}{3}\right)^d|D|^{\frac{d}{2r+d}}\exp\left\{-C|D|^{\frac{2r}{2r+d}}\right\}\leq\tilde{C}_5|D|^{-\frac{2r}{2r+d}},$$

where $\tilde{C}_5$ is a constant independent of $|D|$. Thus,

$$\mathbf{E}\left\{\|f_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\leq C_3|D|^{-\frac{2r}{2r+d}}$$

with C_3 a constant independent of $|D|$, thus completing the proof. □

To prove Theorem 4.4, we need the following lemma, which can be obtained by a similar proof as Lemma 6.8.

LEMMA 6.9. *For $\overline{f}_{D,n}$ in Definition 4.3, there holds*

$$\begin{aligned}
\mathbf{E}\left\{\|\overline{f}_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}&\leq\sum_{j=1}^m\frac{|D_j|^2}{|D|^2}\mathbf{E}\left\{\|f_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\
&\quad+\sum_{j=1}^m\frac{|D_j|}{|D|}\left\|\mathbf{E}\{f_{D_j,n}\}-f^*\right\|^2_{L_2(\mathbb{S}^d)}.
\end{aligned}$$

Proof of Theorem 4.4. By Lemma 6.9, we only need to estimate the bounds of $\mathbf{E}\{\|f_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\}$ and $\|\mathbf{E}\{f_{D_j,n}\}-f^*\|^2_{L_2(\mathbb{S}^d)}$. To estimate the first, we obtain from (6.23) with $D=D_j$ that for $j=1,\ldots,m$,

$$\begin{aligned}
\mathbf{E}\left\{\|f_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}&\leq c_5^2n^{-2r}\|f\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+\frac{2c_1^2M^2n^d}{|D_j|}\\
&\quad+2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}\left(|\mathbb{S}^d|+2c_1^2n^d+2\right)\exp\{-C|D_j|/n^d\}.
\end{aligned}$$

Since $\min_{1\leq j\leq m}|D_j|\geq|D|^{\frac{d+\nu}{2r+d}}$, $\frac{c_3}{6}|D|^{\frac{1}{2r+d}}\leq n\leq\frac{c_3}{3}|D|^{\frac{1}{2r+d}}$, $2r>d$, and $0<\nu<2r$,

$$2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}\left(|\mathbb{S}^d|+2c_1^2n^d+2\right)\exp\{-C|D_j|/n^d\}\leq\tilde{C}_7|D|^{-\frac{2r}{2r+d}},$$

where $\tilde{C}_7$ is a constant depending only on $r,c_1,C,d,$ and f^*. Thus, there exists a constant $\tilde{C}_8$ independent of $m,n,|D_1|,\ldots,|D_m|$, and $|D|$ such that

$$\begin{aligned}
&\sum_{j=1}^m\frac{|D_j|^2}{|D|^2}\mathbf{E}\left\{\|f_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\
&\leq\tilde{C}_8\left(|D|^{-\frac{2r}{2r+d}}+\sum_{j=1}^m\frac{|D_j|^2}{|D|^2}\frac{|D|^{\frac{d}{2r+d}}}{|D_j|}\right)=(\tilde{C}_8+1)|D|^{-\frac{2r}{2r+d}},
\end{aligned}\tag{6.24}$$

where we used $\sum_{i=1}^m \frac{|D_j|}{|D|}=1$. To bound $\|\mathbf{E}\{f_{D_j,n}\}-f^*\|^2_{L_2(\mathbb{S}^d)}$, let Λ_{D_j} be the set of points of the data set D_j. Then we use (6.16) and Jensen's inequality to obtain

$$\begin{aligned}\left\|\mathbf{E}\{f_{D_j,n}\}-f^*\right\|^2_{L_2(\mathbb{S}^d)} &= \left\|\mathbf{E}\{\mathbf{E}\{f_{D_j,n}|\Lambda_{D_j}\}-f^*\}\right\|^2_{L_2(\mathbb{S}^d)}\\ &= \left\|\mathbf{E}\{f^*_{D_j,n}-f^*\}\right\|^2_{L_2(\mathbb{S}^d)} \leq \mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}.\end{aligned}\tag{6.25}$$

We now use a similar proof as Theorem 4.2 to prove the error bound of distributed filtered hyperinterpolation $\overline{f}_{D,n}$. For each $j=1,\ldots,m$, we let Ω_{D_j} be the event such that the sum of the quadrature weights $\sum_{i=1} w^2_{i,n,D_j}\leq 2/|D_j|$ and $\Omega^c_{D_j}$ be the complement of Ω_{D_j}. Write

$$\begin{aligned}\mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\} &= \mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega_{D_j}\right\}\mathbf{P}\{\Omega_{D_j}\}\\ &\quad+\mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega^c_{D_j}\right\}\mathbf{P}\{\Omega^c_{D_j}\},\end{aligned}\tag{6.26}$$

where $\mathbf{E}\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega^c_{D_j}\}\mathbf{P}\{\Omega^c_{D_j}\}\leq 4\|f^*\|^2_{L_\infty(\mathbb{S}^d)}\exp\{-C|D_j|/n^d\}$.

By (6.22) with $D=D_j$, the second term of the right-hand side in (6.26) becomes

$$\begin{aligned}&\mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}|\Omega_{D_j}\right\}\mathbf{P}\{\Omega_{D_j}\}\\ &\quad\leq c_5^2 n^{-2r}\|f\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}(|\mathbb{S}^d|+2c_1^2n^d)\exp\{-C|D_j|/n^d\}.\end{aligned}$$

These two estimates give

$$\begin{aligned}&\mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\\ &\quad\leq c_5^2 n^{-2r}\|f\|^2_{\mathbb{W}_2^r(\mathbb{S}^d)}+2\|f^*\|^2_{L_\infty(\mathbb{S}^d)}(|\mathbb{S}^d|+2c_1^2n^d+2)\exp\{-C|D_j|/n^d\}.\end{aligned}$$

By $\min_{1\leq j\leq m}|D_j|\geq |D|^{\frac{d+\nu}{2r+d}}$, $\frac{c_3}{6}|D|^{\frac{1}{2r+d}}\leq n\leq \frac{c_3}{3}|D|^{\frac{1}{2r+d}}$, and $2r>d$, $0<\nu<2r$,

$$\mathbf{E}\left\{\|f^*_{D_j,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\leq \tilde{C}_9|D|^{-\frac{2r}{2r+d}},$$

which with (6.25) and $\sum_{j=1}^m\frac{|D_j|}{|D|}=1$ gives

$$\sum_{j=1}^m\frac{|D_j|}{|D|}\left\|\mathbf{E}\{f_{D_j,n}\}-f^*\right\|^2_{L_2(\mathbb{S}^d)}\leq \tilde{C}_9|D|^{-\frac{2r}{2r+d}}.\tag{6.27}$$

Using (6.24) and (6.27) in Lemma 6.9 then gives

$$\mathbf{E}\left\{\|\overline{f}_{D,n}-f^*\|^2_{L_2(\mathbb{S}^d)}\right\}\leq C_4|D|^{-\frac{2r}{2r+d}},$$

thus completing the proof. ∎

Acknowledgments. The authors thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak, et al., *Planck* 2018 *Results.* I. *Overview and the Cosmological Legacy of Planck*, preprint, arXiv:1807.06205, 2018.

[2] P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard, *Adaptive density estimation for directional data using needlets*, Ann. Statist., 37 (2009), pp. 3362–3395, https://doi.org/10.1214/09-AOS682.

[3] R. Bauer, *Distribution of points on a sphere with application to star catalogs*, J. Guid. Control Dyn., 23 (2000), pp. 130–137.

[4] A. Bondarenko, D. Radchenko, and M. Viazovska, *Optimal asymptotic bounds for spherical designs*, Ann. Math. (2), 178 (2013), pp. 443–452, https://doi.org/10.4007/annals.2013.178.2.2.

[5] J. Brauchart, J. Dick, E. Saff, I. Sloan, Y. Wang, and R. Womersley, *Covering of spheres by spherical caps and worst-case error for equal weight cubature in Sobolev spaces*, J. Math. Anal. Appl., 431 (2015), pp. 782–811.

[6] J. S. Brauchart, A. B. Reznikov, E. B. Saff, I. H. Sloan, Y. G. Wang, and R. S. Womersley, *Random point sets on the sphere—Hole radii, covering, and separation*, Exp. Math., 27 (2018), pp. 62–81, https://doi.org/10.1080/10586458.2016.1226209.

[7] G. Brown and F. Dai, *Approximation of smooth functions on compact two-point homogeneous spaces*, J. Funct. Anal., 220 (2005), pp. 401–423, https://doi.org/10.1016/j.jfa.2004.10.005.

[8] A. Chernih, I. H. Sloan, and R. S. Womersley, *Wendland functions with increasing smoothness converge to a Gaussian*, Adv. Comput. Math., 40 (2014), pp. 185–200, https://doi.org/10.1007/s10444-013-9304-5.

[9] F. Cucker and D.-X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge Monogr. Appl. Comput. Math. 24, Cambridge University Press, Cambridge, 2007, https://doi.org/10.1017/CBO9780511618796.

[10] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Spherical codes and designs*, Geom. Dedicata, 6 (1977), pp. 363–388, https://doi.org/10.1007/bf03187604.

[11] M. Fan, D. Paul, T. C. M. Lee, and T. Matsuo, *Modeling tangential vector fields on a sphere*, J. Amer. Statist. Assoc., 113 (2018), pp. 1625–1636, https://doi.org/10.1080/01621459.2017.1356322.

[12] M. Fan, D. Paul, T. C. M. Lee, and T. Matsuo, *A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials*, Ann. Appl. Stat., 12 (2018), pp. 459–489, https://doi.org/10.1214/17-AOAS1104.

[13] W. Freeden, T. Gervens, and M. Schreiner, *Constructive Approximation on the Sphere*, Clarendon Press, Oxford University Press, New York, 1998.

[14] Q. T. L. Gia, I. H. Sloan, R. S. Womersley, and Y. G. Wang, *Isotropic sparse regularization for spherical harmonic representations of random fields on the sphere*, Appl. Comput. Harmon. Anal., (2019), https://doi.org/10.1016/j.acha.2019.01.005.

[15] P. Grabner and I. H. Sloan, *Lower bounds and separation for spherical designs*, in Uniform Distribution Theory and Applications, Mathematisches Forschungsinstitut Oberwolfach Report 49/2013, 2013, pp. 2887–2890.

[16] Z.-C. Guo, S.-B. Lin, and D.-X. Zhou, *Learning theory of distributed spectral algorithms*, Inverse Probl., 33 (2017), 074009, http://stacks.iop.org/0266-5611/33/i=7/a=074009.

[17] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer Series in Statistics, Springer-Verlag, Berlin, 2002, https://doi.org/10.1007/b97848.

[18] K. Hesse, H. N. Mhaskar, and I. H. Sloan, *Quadrature in Besov spaces on the Euclidean sphere*, J. Complexity, 23 (2007), pp. 528–552, https://doi.org/10.1016/j.jco.2006.10.004.

[19] K. Hesse, I. H. Sloan, and R. S. Womersley, *Numerical integration on the sphere*, Handb. Geomath., (2010), pp. 1185–1219.

[20] K. Hesse, I. H. Sloan, and R. S. Womersley, *Radial basis function approximation of noisy scattered data on the sphere*, Numer. Math., 137 (2017), pp. 579–605, https://doi.org/10.1007/s00211-017-0886-6.

[21] A. I. Kamzolov, *Approximation of functions on the sphere* S^n, Serdica, 10 (1984), pp. 3–10.

[22] G. Kerkyacharian, T. M. Pham Ngoc, and D. Picard, *Localized spherical deconvolution*, Ann. Statist., 39 (2011), pp. 1042–1068, https://doi.org/10.1214/10-AOS858.

[23] Q. T. Le Gia and H. N. Mhaskar, *Localized linear polynomial operators and quadrature formulas on the sphere*, SIAM J. Numer. Anal., 47 (2008/2009), pp. 440–466, https://doi.org/10.1137/060678555.

[24] Q. T. Le Gia and H. N. Mhaskar, *Localized linear polynomial operators and quadrature formulas on the sphere*, SIAM J. Numer. Anal., 47 (2008/2009), pp. 440–466, https://doi.org/10.1137/060678555.
[25] Q. T. Le Gia, F. J. Narcowich, J. D. Ward, and H. Wendland, *Continuous and discrete least-squares approximation by radial basis functions on spheres*, J. Approx. Theory, 143 (2006), pp. 124–133, https://doi.org/10.1016/j.jat.2006.03.007.
[26] Q. T. Le Gia, I. H. Sloan, and H. Wendland, *Multiscale analysis in Sobolev spaces on the sphere*, SIAM J. Numer. Anal., 48 (2010), pp. 2065–2090, https://doi.org/10.1137/090774550.
[27] S. Lin, F. Cao, X. Chang, and Z. Xu, *A general radial quasi-interpolation operator on the sphere*, J. Approx. Theory, 164 (2012), pp. 1402–1414, https://doi.org/10.1016/j.jat.2012.07.001.
[28] S.-B. Lin, *Nonparametric regression using needlet kernels for spherical data*, J. Complexity, 50 (2019), pp. 66–83, https://doi.org/10.1016/j.jco.2018.09.003.
[29] S.-B. Lin, X. Guo, and D.-X. Zhou, *Distributed learning with regularized least squares*, J. Mach. Learn. Res., 18 (2017), pp. Paper No. 92, 31.
[30] S.-B. Lin and D.-X. Zhou, *Distributed kernel-based gradient descent algorithms*, Constr. Approx., 47 (2018), pp. 249–276, https://doi.org/10.1007/s00365-017-9379-1.
[31] D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, and N. Vittorio, *Spherical needlets for cosmic microwave background data analysis*, Monthly Not. Roy. Astron. Soc., 383 (2008), pp. 539–545, https://doi.org/10.1111/j.1365-2966.2007.12550.x.
[32] M. Mendillo, *Storms in the ionosphere: Patterns and processes for total electron content*, Rev. Geophys., 44 (2006), https://doi.org/10.1029/2005RG000193.
[33] H. N. Mhaskar, *Weighted quadrature formulas and approximation by zonal function networks on the sphere*, J. Complexity, 22 (2006), pp. 348–370, https://doi.org/10.1016/j.jco.2005.10.003.
[34] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, *Approximation properties of zonal function networks using scattered data on the sphere*, Adv. Comput. Math., 11 (1999), pp. 121–137, https://doi.org/10.1023/A:1018967708053.
[35] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, *Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature*, Math. Comp., 70 (2001), pp. 1113–1130, https://doi.org/10.1090/S0025-5718-00-01240-0.
[36] P. Mukhtarov, D. Pancheva, B. Andonov, and L. Pashova, *Global TEC maps based on GNSS data: 1. Empirical background TEC model*, JGR Space Phys., 118 (2013), pp. 4594–4608, https://doi.org/10.1002/jgra.50413.
[37] C. Müller, *Spherical Harmonics*, Springer-Verlag, Berlin, 1966.
[38] F. J. Narcowich, P. Petrushev, and J. D. Ward, *Localized tight frames on spheres*, SIAM J. Math. Anal., 38 (2006), pp. 574–594, https://doi.org/10.1137/040614359.
[39] F. J. Narcowich, X. Sun, J. D. Ward, and H. Wendland, *Direct and inverse Sobolev error estimates for scattered data interpolation via spherical basis functions*, Found. Comput. Math., 7 (2007), pp. 369–390, https://doi.org/10.1007/s10208-005-0197-7.
[40] F. J. Narcowich and J. D. Ward, *Scattered data interpolation on spheres: Error estimates and locally supported basis functions*, SIAM J. Math. Anal., 33 (2002), pp. 1393–1410, https://doi.org/10.1137/S0036141001395054.
[41] Planck Collaboration and R. Adam et al., *Planck* 2015 *results*—I. *Overview of products and scientific results*, Astron. Astrophys., 594 (2016), p. A1, https://doi.org/10.1051/0004-6361/201527101.
[42] E. A. Rakhmanov, E. Saff, and Y. Zhou, *Minimal discrete energy on the sphere*, Math. Res. Lett., 1 (1994), pp. 647–662.
[43] I. H. Sloan, *Polynomial interpolation and hyperinterpolation over general regions*, J. Approx. Theory, 83 (1995), pp. 238–254, https://doi.org/10.1006/jath.1995.1119.
[44] I. H. Sloan and R. S. Womersley, *Filtered hyperinterpolation: A constructive polynomial approximation on the sphere*, GEM Int. J. Geomath., 3 (2012), pp. 95–117, https://doi.org/10.1007/s13137-011-0029-7.
[45] H. Wang and I. H. Sloan, *On filtered polynomial approximation on the sphere*, J. Fourier Anal. Appl., 23 (2017), pp. 863–876, https://doi.org/10.1007/s00041-016-9493-7.
[46] Y. Wang, *Filtered polynomial approximation on the sphere*, Bull. Aust. Math. Soc., 93 (2016), pp. 162–163, https://doi.org/10.1017/S000497271500132X.
[47] Y. G. Wang, Q. T. Le Gia, I. H. Sloan, and R. S. Womersley, *Fully discrete needlet approximation on the sphere*, Appl. Comput. Harmon. Anal., 43 (2017), pp. 292–316, https://doi.org/10.1016/j.acha.2016.01.003.

[48] Y. G. Wang, I. H. Sloan, and R. S. Womersley, *Riemann localisation on the sphere*, J. Fourier Anal. Appl., 24 (2018), pp. 141–183, https://doi.org/10.1007/s00041-016-9496-4.
[49] H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Adv. Comput. Math., 4 (1995), pp. 389–396, https://doi.org/10.1007/BF02123482.
[50] R. S. Womersley, *Efficient spherical designs with good geometric properties*, in Contemporary Computational Mathematics—A Celebration of the 80th Birthday of Ian Sloan. Vols. 1 and 2, Springer-Verlag, Berlin, 2018, pp. 1243–1285.
[51] Q. Wu and D.-X. Zhou, *SVM soft margin classifiers: Linear programming versus quadratic programming*, Neural Comput., 17 (2005), pp. 1160–1187, https://doi.org/10.1162/0899766053491896.
[52] Z. M. Wu, *Compactly supported positive definite radial functions*, Adv. Comput. Math., 4 (1995), pp. 283–292, https://doi.org/10.1007/BF03177517.
[53] D.-X. Zhou, *Deep distributed convolutional neural networks: Universality*, Anal. Appl. (Singap.), 16 (2018), pp. 895–919, https://doi.org/10.1142/S0219530518500124.
[54] D.-X. Zhou, *Distributed Approximation with Deep Convolutional Neural Networks*, manuscript, 2018.
[55] D.-X. Zhou, *Universality of deep convolutional neural networks*, Appl. Comput. Harmon. Anal., (2019), https://doi.org/10.1016/j.acha.2019.06.004.
[56] D.-X. Zhou and K. Jetter, *Approximation with polynomial kernels and SVM classifiers*, Adv. Comput. Math., 25 (2006), pp. 323–344, https://doi.org/10.1007/s10444-004-7206-2.