Numerische Mathematik

CrossMark

Radial basis function approximation of noisy scattered data on the sphere

Kerstin Hesse[1] [iD] · **Ian H. Sloan[2]** [iD] ·
Robert S. Womersley[2] [iD]

Abstract In this paper we consider the approximation of noisy scattered data on the sphere by radial basis functions generated by a strictly positive definite kernel. The approximation is the minimizer in the native space for that kernel of a quadratic functional in which the smoothing term is a multiple of the square of the native space norm. The balance between data fitting and smoothness is controlled by a smoothing parameter, the choice of which should depend on the nature and magnitude of the noise. The main results concern the choice of that smoothing parameter, under the assumption that the noise is deterministic rather than random. Four strategies for choosing the smoothing parameter are considered: Morozov's discrepancy principle, and three a priori strategies. For each of these strategies we derive an L_2 error bound. The error bounds are similar, with the discrepancy principle giving marginally the best bound. A numerical example supports the theoretical results.

✉ Kerstin Hesse
  kerstin.hesse@math.uni-paderborn.de

  Ian H. Sloan
  I.Sloan@unsw.edu.au

  Robert S. Womersley
  R.Womersley@unsw.edu.au

[1] Department of Mathematics, Paderborn University, Warburger Straße 100, 33098 Paderborn, Germany

[2] School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia

🌀 Springer

1 Introduction

This paper is a contribution to the approximation of noisy scattered data on the d-dimensional unit sphere $\mathbb{S}^d$. The most important case in practice is the case $d = 2$, for which there are numerous geophysical problems in which the need is to fit noisy data available only at scattered points. Some examples are gravitational potential, magnetic field intensity, topography, and ground or sea level temperature or pressure (see for example [8]). Often a series of independent local approximations on patches will suffice, but here we seek a single global approximation.

Our ultimate approximation consists of radial basis functions centered at the given data points, thus allowing a finer approximation wherever the data points are concentrated, perhaps because of greater variability of the physical quantity in that region. The radial basis functions are generated by a single strictly positive definite kernel.

The approximation is defined (see Sect. 3) to be the minimizer in the native space for that kernel of a certain quadratic functional. This functional is the sum of the squared ℓ_2 norm of the discrepancy between the values of the approximation and the noisy data values at the data points, and a smoothing term (or regularization term) consisting of a smoothing parameter (or regularization parameter) λ multiplied by the squared native space norm of the approximation.

The main results (see Theorem 4.1) concern the choice of that smoothing parameter. The choice should depend on the nature and magnitude of the noise. The statistical and simulation literatures contain a large number of strategies for choosing the smoothing parameter for ill-posed problems and variable selection, see for example [4,32]. In this paper the data errors are considered to be deterministic rather than random. By this we mean that we assume that the data errors at the N data sites $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbb{S}^d$ are given by an error vector (or noise vector) $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T$, with the assumption that only $\|\boldsymbol{\varepsilon}\|_2$ is known. (Here for $k = 1, 2, \ldots, N$, ε_k is the data error at $\mathbf{x}_k$, and $\|\boldsymbol{\varepsilon}\|_2$ is the Euclidean norm of the noise vector). We do not assume that this noise has a known or unknown stochastic distribution, and our error estimates are deterministic and not in terms of expectation values.

For a given level of data error, we consider four strategies for choosing the smoothing parameter λ, namely Morozov's discrepancy principle (which is an *a posteriori* strategy) and three a priori strategies. In all four cases we obtain, under the requirement that the native space is equivalent to a Sobolev space H^s with $s > d/2$, L_2 bounds on the approximation error, see Theorem 4.1, that are a sum of two terms. One term is of the order h_X^s and the other one is of the order $h_X^{d/2}\|\boldsymbol{\varepsilon}\|_2$, where h_X is the mesh norm of the set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of the data sites.

The theoretical results, which make essential use of an L_2 sampling inequality for the sphere, slightly favor the discrepancy principle.

The theory developed in this paper is especially relevant to the compactly supported radial basis functions of Wendland (see [28–30]), for which the native space is in every case equivalent to a Sobolev space. The compactness of the support allows the use of

sparse matrix technology in solving the resulting linear systems, greatly increasing the efficiency of the approximation scheme.

The structure of the paper is as follows: In Sect. 2 we introduce the required notation and background material on the sphere, function spaces on the sphere, and radial basis functions. In Sect. 3 the approximation is introduced, and discussed in the context of penalized least squares approximation (see, for example, [1,25] for a general discussion of penalized least squares approximation, [2] for a particular polynomial variant, and [6,24] for this topic in the context of learning theory). In Sect. 4 we present the main result, Theorem 4.1, on L_2 error estimates for the smoothing approximation from noisy scattered data on the sphere using radial basis functions. In Sect. 5 we state the L_2 sampling inequality for the sphere and sketch its proof. Using the method from [15] we can lift the L_2 sampling inequality from [3, Theorem 4.1] by the use of charts to the sphere, and so obtain an analogous L_2 sampling inequality for the sphere. This is an essential part of the proof of our main result. In Sect. 6 we finally give the proof of the main result, Theorem 4.1. In Sect. 7 we give a numerical illustration of the different strategies and discuss their practical implementation.

2 Notation and background

In this section, the general notation and background material are introduced.

Throughout the paper generic positive constants are denoted by $c, \tilde{c}, \ldots$, and may have different values at different places. They may depend on the sphere dimension d, the Sobolev space index s, and the choice of the radial basis function in our approximation scheme.

For two sequences $\{a_n\}$ and $\{b_n\}$, the notation $a_n \asymp b_n$ means that there exist positive constants c and $\tilde{c}$ such that $ca_n \le b_n \le \tilde{c}a_n$ for all n.

2.1 The sphere, spherical geometry, and the mesh norm

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}$, let $\mathbf{x} \cdot \mathbf{y}$ denote the Euclidean inner product and let $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ denote the Euclidean norm. The *unit sphere* $\mathbb{S}^d$ in $\mathbb{R}^{d+1}$ is given by

$$\mathbb{S}^d := \left\{ \mathbf{x} \in \mathbb{R}^{d+1} \ : \ \|\mathbf{x}\|_2 = 1 \right\}.$$

The unit sphere $\mathbb{S}^d$ has the (Lebesgue) surface area

$$|\mathbb{S}^d| = \frac{2\pi^{(d+1)/2}}{\Gamma\big((d+1)/2\big)}.$$

The separation between any two points $\mathbf{x}$ and $\mathbf{y}$ on $\mathbb{S}^d$ is measured by the *geodesic distance* $\mathrm{dist}(\mathbf{x}, \mathbf{y}) \in [0, \pi]$, defined by

$$\mathrm{dist}(\mathbf{x}, \mathbf{y}) := \arccos(\mathbf{x} \cdot \mathbf{y}).$$

Let $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^d$ denote a set of N distinct points on $\mathbb{S}^d$. The *mesh norm* of X is defined by

$$h_X := \sup_{\mathbf{y} \in \mathbb{S}^d} \min_{\mathbf{x}_j \in X} \operatorname{dist}(\mathbf{y}, \mathbf{x}_j).$$

2.2 Function spaces on the sphere

The space $L_2 = L_2(\mathbb{S}^d)$ is the usual Hilbert space of square-integrable functions on $\mathbb{S}^d$ endowed with the inner product

$$(f, g)_{L_2} := \int_{\mathbb{S}^d} f(\mathbf{x}) g(\mathbf{x}) \, d\omega_d(\mathbf{x}),$$

and the induced norm is $\|f\|_{L_2} := \sqrt{(f, f)_{L_2}}$. Here $d\omega_d$ denotes the (Lebesgue) surface measure on $\mathbb{S}^d$. The space of continuous functions on $\mathbb{S}^d$ is denoted by $C = C(\mathbb{S}^d)$ and is endowed with the supremum norm

$$\|f\|_C := \sup_{\mathbf{x} \in \mathbb{S}^d} |f(\mathbf{x})|.$$

The space $\mathbb{P}_L = \mathbb{P}_L(\mathbb{S}^d)$ of all *spherical polynomials* on $\mathbb{S}^d$ of degree $\leq L$ consists of the restrictions to $\mathbb{S}^d$ of all polynomials on $\mathbb{R}^{d+1}$ of degree $\leq L$. The dimension of $\mathbb{P}_L$ is given by

$$d_L = \dim\left(\mathbb{P}_L\right) = \frac{(2L + d)\Gamma(L + d)}{\Gamma(d + 1)\Gamma(L + 1)} \asymp (L + 1)^d.$$

A *spherical harmonic* of degree $\ell \in \mathbb{N}_0$ is the restriction to $\mathbb{S}^d$ of a harmonic homogeneous polynomial of exact degree ℓ on $\mathbb{R}^{d+1}$ (see [7, Section 11.2]). The space $\mathbb{H}_\ell = \mathbb{H}_\ell(\mathbb{S}^d)$ of spherical harmonics of degree $\ell \in \mathbb{N}_0$ (together with the zero function) has the dimension $Z(d, \ell) = \dim(\mathbb{H}_\ell)$, given by

$$Z(d, 0) = 1; \quad Z(d, \ell) = \frac{(2\ell + d - 1)\Gamma(\ell + d - 1)}{\Gamma(d)\Gamma(\ell + 1)}, \quad \ell \in \mathbb{N}.$$

Note that $Z(d, \ell) \asymp (\ell + 1)^{d-1}$. In this paper, for any $\ell \in \mathbb{N}_0$, the set

$$\left\{Y_{\ell,k} : k = 1, 2, \ldots, Z(d, \ell)\right\} \tag{2.1}$$

denotes a real L_2-orthonormal basis of $\mathbb{H}_\ell$. Furthermore, $\mathbb{P}_L = \bigoplus_{\ell=0}^{L} \mathbb{H}_\ell$, and, in particular, the union over $\ell = 0, 1, \ldots, L$ of the sets (2.1) forms an L_2-orthonormal basis of $\mathbb{P}_L$.

For fixed $\alpha, \beta > -1$, $P_\ell^{(\alpha,\beta)}$ denotes the *Jacobi polynomial* of degree ℓ with indices α and β (see [23, Chapter IV]). The Jacobi polynomials $\{P_\ell^{(\alpha,\beta)}\}_{\ell \in \mathbb{N}_0}$ are orthogonal

on the interval $[-1, 1]$ with respect to the weighted inner product

$$(f, g)_{L_2^{(\alpha,\beta)}([-1,1])} := \int_{-1}^{1} f(t)g(t)(1-t)^{\alpha}(1+t)^{\beta} \, dt,$$

and the induced norm is $\|f\|_{L_2^{(\alpha,\beta)}([-1,1])} = \sqrt{(f, f)_{L_2^{(\alpha,\beta)}([-1,1])}}$, where the normalization is such that (see [23, (4.1.1)])

$$P_{\ell}^{(\alpha,\beta)}(1) = \frac{\Gamma(\ell + \alpha + 1)}{\Gamma(\alpha + 1)\Gamma(\ell + 1)} \tag{2.2}$$

and (see [23, (4.3.3)])

$$\|P_{\ell}^{(\alpha,\beta)}\|_{L_2^{(\alpha,\beta)}([-1,1])}^2 = \frac{2^{\alpha+\beta+1}}{(2\ell + \alpha + \beta + 1)} \frac{\Gamma(\ell + \alpha + 1)\Gamma(\ell + \beta + 1)}{\Gamma(\ell + 1)\Gamma(\ell + \alpha + \beta + 1)}. \tag{2.3}$$

We denote the space of all measurable functions f on $[-1, 1]$ for which $\|f\|_{L_2^{(\alpha,\beta)}([-1,1])} < \infty$ by $L_2^{(\alpha,\beta)}([-1, 1])$. The Jacobi polynomials $\{P_{\ell}^{(\alpha,\beta)}\}_{\ell \in \mathbb{N}_0}$ form a complete orthogonal system in $L_2^{(\alpha,\beta)}([-1, 1])$.

The spherical harmonics on $\mathbb{S}^d$ of degree ℓ satisfy the *addition theorem* (see [7, Section 11.4] and [20, Section 4.1, Lemma 4.5 and Theorem 4.7]): for any L_2-orthonormal basis (2.1) of $\mathbb{H}_{\ell}$ we have

$$\sum_{k=1}^{Z(d,\ell)} Y_{\ell,k}(\mathbf{x})Y_{\ell,k}(\mathbf{y}) = \frac{Z(d,\ell)}{|\mathbb{S}^d|} \frac{P_{\ell}^{((d-2)/2,(d-2)/2)}(\mathbf{x} \cdot \mathbf{y})}{P_{\ell}^{((d-2)/2,(d-2)/2)}(1)}. \tag{2.4}$$

The union over all $\ell \in \mathbb{N}_0$ of the L_2-orthonormal bases (2.1) of $\mathbb{H}_{\ell}$ forms a complete orthonormal system for L_2. Thus any function $f \in L_2$ can be expanded into a *Fourier series* (or Laplace series) with respect to this orthonormal system: in the L_2 sense

$$f = \sum_{\ell=0}^{\infty} \sum_{k=1}^{Z(d,\ell)} \widehat{f}_{\ell,k} Y_{\ell,k},$$

with the Fourier coefficients given by

$$\widehat{f}_{\ell,k} = (f, Y_{\ell,k})_{L_2} = \int_{\mathbb{S}^d} f(\mathbf{x}) Y_{\ell,k}(\mathbf{x}) \, d\omega_d(\mathbf{x}).$$

The *Sobolev space* $H^s = H^s(\mathbb{S}^d)$, where $s \geq 0$, is defined as the space of those functions in L_2 for which the norm

$$\|f\|_{H^s} := \left(\sum_{\ell=0}^{\infty} (\ell + 1)^{2s} \sum_{k=1}^{Z(d,\ell)} |\widehat{f}_{\ell,k}|^2 \right)^{1/2} \tag{2.5}$$

is finite (see [9, Sections 5.1 and 5.2] for the case of $\mathbb{S}^2$ and [20, Section 6.1, Definition 6.2 and Theorem 6.3]). The space H^s is a Hilbert space with the inner product

$$(f, g)_{H^s} = \sum_{\ell=0}^{\infty} (\ell + 1)^{2s} \sum_{k=1}^{Z(d,\ell)} \widehat{f}_{\ell,k} \widehat{g}_{\ell,k},$$

which induces the norm (2.5).

For $s > d/2$, the space H^s is embedded into C, that is, $H^s \subset C$ and there exists a positive constant c such that $\|f\|_C \leq c\|f\|_{H^s}$ for all $f \in H^s$. For $s > d/2$, the space H^s is a *reproducing kernel Hilbert space*, that is, there exists a kernel $K_s : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$ such that (i) $K_s(\mathbf{x}, \mathbf{y}) = K_s(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{S}^d$, (ii) $K_s(\cdot, \mathbf{x}) \in H^s$ for any fixed $\mathbf{x} \in \mathbb{S}^d$, and (iii) the reproducing property holds

$$\left(f, K_s(\cdot, \mathbf{x})\right)_{H^s} = f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{S}^d \text{ and all } f \in H^s.$$

It is easily verified that the reproducing kernel of H^s is

$$K_s(\mathbf{x}, \mathbf{y}) = \sum_{\ell=0}^{\infty} \sum_{k=1}^{Z(d,\ell)} \frac{Y_{\ell,k}(\mathbf{x}) Y_{\ell,k}(\mathbf{y})}{(\ell + 1)^{2s}}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}^d.$$

2.3 Strictly positive definite functions and the native space

A function $\Phi \in L_2^{((d-2)/2,(d-2)/2)}([-1, 1])$ has the Jacobi series expansion

$$\Phi = \sum_{\ell=0}^{\infty} a_\ell \frac{Z(d, \ell)}{|\mathbb{S}^d|} \frac{P_\ell^{((d-2)/2,(d-2)/2)}}{P_\ell^{((d-2)/2,(d-2)/2)}(1)},$$

with the coefficients (see (2.2) and (2.3) for the normalization)

$$a_\ell := 2\pi^{d/2} \frac{\Gamma(\ell + 1)}{\Gamma(\ell + d/2)} \int_{-1}^{1} \Phi(t) P_\ell^{((d-2)/2,(d-2)/2)}(t)(1 - t^2)^{(d-2)/2} \, dt.$$

Then the kernel $\phi : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$, defined by

$$\phi(\mathbf{x}, \mathbf{y}) := \Phi(\mathbf{x} \cdot \mathbf{y}) = \sum_{\ell=0}^{\infty} a_\ell \frac{Z(d, \ell)}{|\mathbb{S}^d|} \frac{P_\ell^{((d-2)/2,(d-2)/2)}(\mathbf{x} \cdot \mathbf{y})}{P_\ell^{((d-2)/2,(d-2)/2)}(1)}$$

$$= \sum_{\ell=0}^{\infty} a_\ell \sum_{k=1}^{Z(d,\ell)} Y_{\ell,k}(\mathbf{x}) Y_{\ell,k}(\mathbf{y}), \tag{2.6}$$

is a *zonal function*, that is, its values depend only on the inner product of $\mathbf{x}$ and $\mathbf{y}$ (or equivalently, on the geodesic distance dist$(\mathbf{x}, \mathbf{y})$). The last representation in (2.6) follows from the addition theorem (2.4).

If $a_\ell \geq 0$ for all $\ell \in \mathbb{N}_0$, and if $a_\ell > 0$ for infinitely many even indices ℓ and for infinitely many odd indices ℓ, then (see [5,21,31]) the kernel ϕ, defined by (2.6), is *strictly positive definite*, that is, for any $N \in \mathbb{N}$, for any set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^d$ of N distinct points, and for any $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)^T \in \mathbb{R}^N$,

$$\sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_j \alpha_k \phi(\mathbf{x}_j, \mathbf{x}_k) \geq 0,$$

and equality occurs only if $\boldsymbol{\alpha} = \mathbf{0} = (0, 0, \ldots, 0)^T$. In other words, the kernel ϕ is strictly positive definite if for any $N \in \mathbb{N}$ and for any set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^d$ of N distinct points the matrix $\left[\phi(\mathbf{x}_j, \mathbf{x}_k)\right]_{j,k=1,2,\ldots,N}$ is positive definite.

A strictly positive definite zonal function ϕ, given by (2.6), is also called a *radial basis function (RBF)* (for the sphere) or a *spherical (radial) basis function*, since the value $\phi(\mathbf{x}, \mathbf{y})$ depends only on the geodesic distance between $\mathbf{x}$ and $\mathbf{y}$ on $\mathbb{S}^d$.

For a continuous function $f : \mathbb{S}^d \to \mathbb{R}$, and a given set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of N distinct points on $\mathbb{S}^d$, an *RBF approximant* of f is a function of the form

$$\sum_{j=1}^{N} \alpha_j \phi(\cdot, \mathbf{x}_j), \tag{2.7}$$

where the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_N \in \mathbb{R}$ are suitably chosen. In this paper, we approximate a function f given in the form of noisy scattered data on a finite point set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ by an RBF approximant (2.7), that is, the centers $\mathbf{x}_j$ in (2.7) are at the points where the data of f is given.

Let ϕ be a strictly positive definite zonal kernel, given by (2.6), and assume that the coefficients a_ℓ satisfy $a_\ell > 0$ for all $\ell \in \mathbb{N}_0$ and that $\sum_{\ell=0}^{\infty} a_\ell Z(d, \ell) < \infty$. Define the linear space

$$\mathscr{F}_\phi := \left\{ \sum_{j=1}^{N} \alpha_j \phi(\cdot, \mathbf{x}_j) \, : \, \alpha_j \in \mathbb{R}, \ \mathbf{x}_j \in \mathbb{S}^d, \ j = 1, 2, \ldots, N, \ \text{and} \ N \in \mathbb{N} \right\}.$$

The *native space* $\mathscr{N}_\phi$ is defined to be the closure of $\mathscr{F}_\phi$ with respect to the norm $\|f\|_\phi := \sqrt{(f, f)_\phi}$, $f \in \mathscr{F}_\phi$, induced by the inner product

$$\left(\sum_{j=1}^{N} \alpha_j \phi(\cdot, \mathbf{x}_j), \sum_{k=1}^{M} \beta_k \phi(\cdot, \mathbf{y}_k) \right)_\phi := \sum_{j=1}^{N} \sum_{k=1}^{M} \alpha_j \beta_k \phi(\mathbf{x}_j, \mathbf{y}_k).$$

(We note that $(\cdot, \cdot)_\phi$ is indeed an inner product for $\mathscr{F}_\phi$ because of the strict positive definiteness of ϕ.) The space $\mathscr{N}_\phi$ is a Hilbert space with the inner product

$$(f, g)_\phi = \sum_{\ell=0}^{\infty} \frac{1}{a_\ell} \sum_{k=1}^{Z(d,\ell)} \widehat{f}_{\ell,k} \widehat{g}_{\ell,k}$$

and the induced norm

$$\|f\|_\phi = \sqrt{(f,f)_\phi} = \left(\sum_{\ell=0}^{\infty} \frac{1}{a_\ell} \sum_{k=1}^{Z(d,\ell)} |\widehat{f}_{\ell,k}|^2 \right)^{1/2}, \tag{2.8}$$

where a_ℓ, $\ell \in \mathbb{N}_0$, are the coefficients of the kernel ϕ, see (2.6).

The native space $\mathscr{N}_\phi$ is a reproducing kernel Hilbert space, with the reproducing kernel ϕ, since ϕ is symmetric and $\phi(\cdot, \mathbf{x}) \in \mathscr{N}_\phi$ for every fixed $\mathbf{x} \in \mathbb{S}^d$, and

$$\big(f, \phi(\cdot, \mathbf{x})\big)_\phi = f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{S}^d \text{ and all } f \in \mathscr{N}_\phi.$$

For $s > d/2$ we see from (2.8) and (2.5) that $\mathscr{N}_\phi$ can be identified with H^s if $a_\ell \asymp (\ell+1)^{-2s}$, and that the norms $\|\cdot\|_\phi$ and $\|\cdot\|_{H^s}$ are then equivalent.

3 Smoothing approximation

In this section the smoothing approximation is introduced. We assume that (inexact) real number data values $F_1, F_2, \ldots, F_N$ are given at points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbb{S}^d$. Also given is a radial basis function ϕ. The smoothing approximation is the uniquely determined minimizer in the native space $\mathscr{N}_\phi$ of a certain quadratic functional. This functional contains two terms: the first one (the *discrepancy*) measures how well the data values are fitted, and the second one is a smoothing term, given by the square of the native space norm multiplied by a smoothing parameter $\lambda \geq 0$. Strategies for choosing λ are considered in the next section; for the present λ is assumed given. The next theorem shows that the function in the native space $\mathscr{N}_\phi$ that minimizes the quadratic functional is a radial basis function approximant, with the RBFs centered at the data points. The characterization of a smoothing approximation as the minimizer of a quadratic functional is, of course, well known, see [26]. See also [1,25] for a discussion of penalized least squares approximation, which is another name for the smoothing approximation approach.

Theorem 3.1 *Let ϕ be a strictly positive definite zonal kernel, given by (2.6) with $a_\ell > 0$ for all $\ell \in \mathbb{N}_0$ and such that $\sum_{\ell=0}^{\infty} a_\ell Z(d,\ell) < \infty$, and let $\mathscr{N}_\phi$ denote the native space of ϕ with norm $\|\cdot\|_\phi$. Let $\lambda \in \mathbb{R}_0^+$, let $N \in \mathbb{N}$, and let $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^d$ be a set of N distinct points. Assume that approximate data values at X, namely $\mathbf{F} = (F_1, F_2, \ldots, F_N)^T \in \mathbb{R}^N$, are given. Define the quadratic functional*

$$\mu_\lambda(g) := \sum_{i=1}^{N} \big[g(\mathbf{x}_i) - F_i \big]^2 + \lambda \|g\|_\phi^2, \quad g \in \mathscr{N}_\phi. \tag{3.1}$$

If $\lambda > 0$, then μ_λ has a unique minimizer f_λ in $\mathscr{N}_\phi$, given by

$$f_\lambda = \sum_{j=1}^{N} \alpha_j \phi(\cdot, \mathbf{x}_j), \tag{3.2}$$

with coefficients uniquely determined by the conditions $f_\lambda(\mathbf{x}_i) + \lambda\alpha_i = F_i, i = 1, \ldots, N$, that is,

$$\sum_{j=1}^{N} \phi(\mathbf{x}_i, \mathbf{x}_j)\alpha_j + \lambda\alpha_i = F_i, \quad i = 1, \ldots, N. \tag{3.3}$$

If $\lambda = 0$, then there exists a unique minimizer f_0 in $\mathscr{N}_\phi$ of μ_0 that has minimal norm $\|\cdot\|_\phi$. This minimizer is of the form (3.2) and is uniquely determined by the conditions $f_0(\mathbf{x}_i) = F_i$ for $i = 1, 2, \ldots, N$ (that is, $f_0(\mathbf{x}_i) = \sum_{j=1}^{N} \phi(\mathbf{x}_i, \mathbf{x}_j)\alpha_j = F_i, i = 1, \ldots, N$).

The nature of the functional μ_λ is that if λ is close to zero then much weight is given to fitting the data, that is, we expect $f_\lambda(\mathbf{x}_i) \approx F_i, i = 1, 2, \ldots, N$, and little importance is given to keeping the norm $\|f_\lambda\|_\phi$ small. We are then close to the interpolation scenario. As λ increases, data fitting becomes less and less important, and more importance is given to keeping the norm $\|f_\lambda\|_\phi$ small.

From (3.3) the coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)^T$ satisfies the linear system

$$(\mathbf{A} + \lambda\mathbf{I})\boldsymbol{\alpha} = \mathbf{F}, \tag{3.4}$$

where $\mathbf{F} = (F_1, F_2, \ldots, F_N)^T$, $\mathbf{I}$ is the $N \times N$ identity matrix, and

$$\mathbf{A} := \left[\phi(\mathbf{x}_i, \mathbf{x}_j)\right]_{i, j=1,2,\ldots,N}. \tag{3.5}$$

The linear system (3.4) has a unique solution for all right hand sides $\mathbf{F}$, since $\mathbf{A} + \lambda\mathbf{I}$ is positive definite for all $\lambda \geq 0$.

In the next section we consider the case $F_i = f(\mathbf{x}_i) + \varepsilon_i, i = 1, 2, \ldots, N$, for some (unknown) function $f \in \mathscr{N}_\phi$ and errors (noise) ε_i in the data, and explore the question of choosing an appropriate value of λ for a given level of noise.

Given Theorem 3.1, it is clearly sufficient to restrict our attention to functions g of the form $g = \sum_{j=1}^{N} \alpha_j\phi(\cdot, \mathbf{x}_j)$. In this case $g(\mathbf{x}_i) = (\mathbf{A}\boldsymbol{\alpha})_i$ and $\|g\|_\phi^2 = \boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha}$, and for such g the functional (3.1) can be written as

$$\tilde{\mu}_\lambda(\boldsymbol{\alpha}) := \mu_\lambda(g) = \|\mathbf{A}\boldsymbol{\alpha} - \mathbf{F}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_{\mathbf{A}}^2, \tag{3.6}$$

where the $\mathbf{A}$ norm is defined by $\|\boldsymbol{\alpha}\|_{\mathbf{A}} := \sqrt{\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha}}$, and the corresponding inner product is defined by $(\boldsymbol{\alpha}, \boldsymbol{\beta})_{\mathbf{A}} := \boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\beta}$, for $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^N$. From (3.6) it is clear that the functional μ_λ restricted to the finite dimensional approximation space $V_X = \text{span}\{\phi(\cdot, \mathbf{x}_j) : j = 1, 2, \ldots, N\}$ can be interpreted as a Tikhonov functional (see [13, Section 2.2]), and hence strategies for the choice of the regularization parameter in Tikhonov regularization can be applied to determine λ. Indeed, the four choices for λ in Sect. 4 are all motivated by parameter choice strategies for Tikhonov regularization.

The next lemma shows that the first term of $\mu_\lambda(f_\lambda)$ (the *discrepancy*) is a strictly monotonic increasing function of λ. This is needed in the next section. The lemma is inspired by [13, Sections 2.2 and 2.5].

Lemma 3.1 *Let the assumptions and notation be the same as in Theorem 3.1, and assume in addition that $\mathbf{F} \neq \mathbf{0}$. Then the discrepancy function $J : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ defined by*

$$J(\lambda) := \sum_{i=1}^{N} \left[f_\lambda(\mathbf{x}_i) - F_i \right]^2, \quad \lambda \in \mathbb{R}_0^+,$$

is continuous and strictly monotonic increasing, with the range of J given by $\left[0, \|\mathbf{F}\|_2^2 \right)$.

Proof We write temporarily $\boldsymbol{\alpha}^\lambda \in \mathbb{R}^N$ for the coefficient vector of f_λ, satisfying (3.4), so that $\boldsymbol{\alpha}^\lambda$ satisfies

$$\left(\mathbf{A} + \lambda \mathbf{I} \right) \boldsymbol{\alpha}^\lambda = \mathbf{F}. \tag{3.7}$$

We note that $\boldsymbol{\alpha}^\lambda \neq \mathbf{0}$, since $\boldsymbol{\alpha}^\lambda = \mathbf{0}$ would contradict $\mathbf{F} \neq \mathbf{0}$.

Since the matrix $\mathbf{A} + \lambda \mathbf{I}$ is invertible and continuous in λ for all non-negative λ, the vector $\boldsymbol{\alpha}^\lambda$ depends continuously on λ. It follows that so too does

$$J(\lambda) = \sum_{i=1}^{N} \left[f_\lambda(\mathbf{x}_i) - F_i \right]^2 = \left\| \mathbf{A} \boldsymbol{\alpha}^\lambda - \mathbf{F} \right\|_2^2. \tag{3.8}$$

To study the monotonicity of $J(\lambda)$ it is convenient to write, from (3.1) and (3.6),

$$\mu_\lambda(f_\lambda) = J(\lambda) + \lambda K(\lambda), \tag{3.9}$$

where

$$K(\lambda) := \left\| f_\lambda \right\|_\phi^2 = (\boldsymbol{\alpha}^\lambda)^T \mathbf{A} \boldsymbol{\alpha}^\lambda = \left\| \boldsymbol{\alpha}^\lambda \right\|_{\mathbf{A}}^2. \tag{3.10}$$

Now let $0 \leq \lambda_1 < \lambda_2$. We first show that $\boldsymbol{\alpha}^{\lambda_1} \neq \boldsymbol{\alpha}^{\lambda_2}$. For this purpose we write (3.7) first with $\lambda = \lambda_1$ and then with $\lambda = \lambda_2$, that is

$$\left(\mathbf{A} + \lambda_1 \mathbf{I} \right) \boldsymbol{\alpha}^{\lambda_1} = \mathbf{F} \quad \text{and} \quad \left(\mathbf{A} + \lambda_2 \mathbf{I} \right) \boldsymbol{\alpha}^{\lambda_2} = \mathbf{F},$$

and subtract the two equations to obtain

$$\mathbf{A}(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}) + \lambda_1 \boldsymbol{\alpha}^{\lambda_1} - \lambda_2 \boldsymbol{\alpha}^{\lambda_2} = \mathbf{0},$$

or equivalently

$$\mathbf{A}(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}) + \lambda_1 (\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}) = (\lambda_2 - \lambda_1) \boldsymbol{\alpha}^{\lambda_2}. \tag{3.11}$$

If $\boldsymbol{\alpha}^{\lambda_1} = \boldsymbol{\alpha}^{\lambda_2}$ then this reduces to $(\lambda_1 - \lambda_2) \boldsymbol{\alpha}^{\lambda_2} = \mathbf{0}$, giving $\lambda_1 = \lambda_2$, a contradiction.

We next show that $K(\lambda)$ is strictly monotonic decreasing. Multiplying (3.11) from the left by $[\mathbf{A}(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2})]^T$, we obtain

$$\left\| \mathbf{A}(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}) \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2} \right\|_{\mathbf{A}}^2 = (\lambda_2 - \lambda_1) \left(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}, \boldsymbol{\alpha}^{\lambda_2} \right)_{\mathbf{A}}. \tag{3.12}$$

Since the left-hand side of (3.12) is positive (because $\mathbf{A}(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}) \neq \mathbf{0}$ due to $\boldsymbol{\alpha}^{\lambda_1} \neq \boldsymbol{\alpha}^{\lambda_2}$ and the positive definiteness of $\mathbf{A}$) and $\lambda_2 > \lambda_1$, it follows that

$$(\boldsymbol{\alpha}^{\lambda_1} - \boldsymbol{\alpha}^{\lambda_2}, \boldsymbol{\alpha}^{\lambda_2})_{\mathbf{A}} > 0,$$

and hence

$$\|\boldsymbol{\alpha}^{\lambda_2}\|_{\mathbf{A}}^2 < (\boldsymbol{\alpha}^{\lambda_1}, \boldsymbol{\alpha}^{\lambda_2})_{\mathbf{A}} \leq \|\boldsymbol{\alpha}^{\lambda_1}\|_{\mathbf{A}}\|\boldsymbol{\alpha}^{\lambda_2}\|_{\mathbf{A}},$$

and on canceling $\|\boldsymbol{\alpha}^{\lambda_2}\|_{\mathbf{A}}$,

$$\|\boldsymbol{\alpha}^{\lambda_2}\|_{\mathbf{A}} < \|\boldsymbol{\alpha}^{\lambda_1}\|_{\mathbf{A}} \Leftrightarrow K(\lambda_2) < K(\lambda_1).$$

Thus $K(\lambda)$ is strictly monotonic decreasing.

Next, we deduce that $J(\lambda)$ is strictly monotonic increasing. For suppose, to the contrary, that for some $0 < \lambda_1 < \lambda_2$ we have $J(\lambda_2) \leq J(\lambda_1)$. Then we have, from (3.1) and the fact that $K(\lambda)$ is strictly monotonic decreasing,

$$\mu_{\lambda_1}(f_{\lambda_2}) = J(\lambda_2) + \lambda_1 K(\lambda_2) < J(\lambda_1) + \lambda_1 K(\lambda_1) = \mu_{\lambda_1}(f_{\lambda_1}).$$

Since f_{λ_1} is the unique minimizer of μ_{λ_1}, $\mu_{\lambda_1}(f_{\lambda_2}) < \mu_{\lambda_1}(f_{\lambda_1})$ is impossible. Thus $J(\lambda)$ is strictly monotonic increasing for all $\lambda > 0$. The monotonicity then extends to all $\lambda \geq 0$ by continuity.

It only remains to establish the range of J. For $\lambda = 0$ we see from (3.3) that

$$J(0) = \sum_{i=1}^{N}[f_0(\mathbf{x}_i) - F_i]^2 = 0.$$

Since J is continuous and strictly monotonic increasing, we need only establish the limit of $J(\lambda)$ as $\lambda \to \infty$. As λ increases to ∞ we expect the minimizer f_λ of (3.1) to become increasingly close to zero. We now show that

$$\boldsymbol{\alpha}^{\lambda} \to \mathbf{0} \quad \text{as} \quad \lambda \to \infty,$$

from which it will follow using (3.2) that $f_\lambda \to 0$ as $\lambda \to \infty$ and thus $J(\lambda) \to \|\mathbf{F}\|_2^2$ as $\lambda \to \infty$. Using (3.8), (3.9) and (3.10) we can write $\mu_\lambda(f_\lambda)$ as

$$\mu_\lambda(f_\lambda) = J(\lambda) + \lambda K(\lambda) = \|\mathbf{A}\boldsymbol{\alpha}^{\lambda} - \mathbf{F}\|_2^2 + \lambda\|\boldsymbol{\alpha}^{\lambda}\|_{\mathbf{A}}^2 \leq \|\mathbf{F}\|_2^2,$$

where the upper bound in the last step follows from the fact that f_λ is the (unique) minimizer of μ_λ, which justifies the replacement of $\boldsymbol{\alpha}^{\lambda}$ by $\mathbf{0}$ to get an upper bound. It follows that $\lambda\|\boldsymbol{\alpha}^{\lambda}\|_{\mathbf{A}}^2 \leq \|\mathbf{F}\|_2^2$, and in turn that $\boldsymbol{\alpha}^{\lambda} \to \mathbf{0}$ as $\lambda \to \infty$. This completes the proof. □

4 Error of the approximation for four strategies for choosing λ

Now we can formulate the main result of this paper, Theorem 4.1. The theorem gives an L_2 error estimate of the smoothing approximation for four different strategies for choosing the parameter λ, given values of a function $f \in \mathcal{N}_\phi$ corrupted by (deterministic) noise. It was inspired by a similar result in [27] for the case of thin plate splines on a bounded domain in $\mathbb{R}^2$ and by [14] for the third parameter choice strategy.

Theorem 4.1 *Let $d \geq 2$ and $s > d/2$, let ϕ be a strictly positive definite zonal kernel given by (2.6) with $a_\ell \asymp (\ell+1)^{-2s}$ for all $\ell \in \mathbb{N}_0$, and let $\mathcal{N}_\phi$ denote the native space of ϕ with norm $\|\cdot\|_\phi$. For a point set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^d$ of N distinct points and a function $f \in \mathcal{N}_\phi$ with values $\mathbf{f} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_N))^T$ on X, and given a vector of deterministic noise $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T$ and given $\lambda \geq 0$, let $f_\lambda^{\boldsymbol{\varepsilon}}$ denote for $\lambda > 0$ the uniquely determined minimizer in $\mathcal{N}_\phi$ of the quadratic functional*

$$\mu_\lambda(g) := \sum_{i=1}^{N} \left[g(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i\right)\right]^2 + \lambda \|g\|_\phi^2, \quad g \in \mathcal{N}_\phi, \quad (4.1)$$

and for $\lambda = 0$ the uniquely determined minimizer in $\mathcal{N}_\phi$ of (4.1) with minimal norm. Then there exist positive constants $\hat{c}_s$ and c_s' (dependent on d and s) such that, if the point set X satisfies $h_X \leq \hat{c}_s$, then the following results hold true for all $f \in \mathcal{N}_\phi$ and all $\boldsymbol{\varepsilon} \in \mathbb{R}^N$:

(i) *A priori parameter choice: Let $\lambda = \|\boldsymbol{\varepsilon}\|_2^2$. Then*

$$\left\|f_\lambda^{\boldsymbol{\varepsilon}} - f\right\|_{L_2} \leq 2c_s' \left[h_X^s \left(\frac{1}{2} + \|f\|_\phi\right) + h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2 \left(1 + \frac{1}{\sqrt{2}} \|f\|_\phi\right)\right].$$

(ii) *A priori parameter choice: Let $\lambda = \dfrac{\|\boldsymbol{\varepsilon}\|_2^2}{\|f\|_\phi^2}$. Then*

$$\left\|f_\lambda^{\boldsymbol{\varepsilon}} - f\right\|_{L_2} \leq c_s' \left[(1 + \sqrt{2})\, h_X^s \|f\|_\phi + \sqrt{6}\, h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2\right].$$

(iii) *A priori parameter choice: Let $\lambda = \frac{1}{\sqrt{2}} h_X^{s-d/2} \dfrac{\|\boldsymbol{\varepsilon}\|_2}{\|f\|_\phi}$. Then*

$$\left\|f_\lambda^{\boldsymbol{\varepsilon}} - f\right\|_{L_2} \leq 2c_s' \left(h_X^s \|f\|_\phi + 2^{1/4} h_X^{s/2+d/4} \|f\|_\phi^{1/2} \|\boldsymbol{\varepsilon}\|_2^{1/2} + h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2\right)$$

$$\leq (1 + 2^{-3/4}) 2c_s' \left(h_X^s \|f\|_\phi + h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2\right). \quad (4.2)$$

(iv) *Morozov's discrepancy principle: Assume that*

$$0 < \|\boldsymbol{\varepsilon}\|_2 < \left\|\mathbf{f} + \boldsymbol{\varepsilon}\right\|_2. \quad (4.3)$$

Then there exists a unique λ > 0 such that

$$\sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 = \|\boldsymbol{\varepsilon}\|_2^2. \tag{4.4}$$

For this choice of λ, the following L_2 error estimate holds:

$$\left\| f_\lambda^{\boldsymbol{\varepsilon}} - f \right\|_{L_2} \leq 2c_s' \left[h_X^s \|f\|_\phi + h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2 \right].$$

Since by assumption in Theorem 4.1 we have $a_\ell \asymp (\ell + 1)^{-2s}$, the native space $\mathcal{N}_\phi$ can be identified with the Sobolev space H^s. Thus the results hold also for all $f \in H^s$, and the norm $\|f\|_\phi$ in the upper bounds can be replaced by $c\|f\|_{H^s}$.

The theorem is proved in Sect. 6.

The four choices of λ considered in the theorem give similar bounds, with Morozov's discrepancy principle giving marginally the best bound. Section 7 gives a numerical experiment illustrating these parameter choices and discusses their practical calculation.

Remark 4.1 The given L_2 error estimates for the parameter choices (ii) to (iv) are *order-optimal* in the following sense:

– The first term in the L_2 error estimates $h_X^s \|f\|_{H^s}$ has the correct order of the mesh norm h_X for an estimate of the L_2 norm against the H^s norm.
– In the second term, we have

$$h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2 \leq h_X^{d/2} \sqrt{N} \max_{j=1,2,\ldots,N} |\varepsilon_j|.$$

Now we consider a 'well distributed' point set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, such as for example the subsets of given point sets constructed in [18, Proposition 3.2]. For such a point set we have (see [18, Proposition 3.2]) $N \asymp h_X^{-d}$, and hence for such a 'well distributed' point set

$$h_X^{d/2} \|\boldsymbol{\varepsilon}\|_2 \leq h_X^{d/2} \sqrt{N} \max_{j=1,2,\ldots,N} |\varepsilon_j| \asymp \max_{j=1,2,\ldots,N} |\varepsilon_j|.$$

As the error of the approximation can never be smaller than the error in the data, this estimate shows that the second term in the L_2 error estimates is also order-optimal.

5 The L_2 sampling inequality for the sphere

The proof of Theorem 4.1 uses the following L_2 *sampling inequality* for the sphere.

Theorem 5.1 *Let $d \geq 2$ and let $s > d/2$. There exist positive constants $\hat{c}_s$ and $\tilde{c}_s$ (dependent on d and s) such that for any finite point set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of N*

distinct points on $\mathbb{S}^d$ with $h_X \leq \hat{c}_s$ and for any function $g \in H^s$

$$\|g\|_{L_2} \leq \tilde{c}_s \left[h_X^s \|g\|_{H^s} + h_X^{d/2} \left(\sum_{j=1}^N [g(\mathbf{x}_j)]^2 \right)^{1/2} \right]. \qquad (5.1)$$

Remark 5.1 The result in [15, Theorem 3.3] is at first sight similar to Theorem 5.1 above, except that in the second term on the right-hand side of (5.1) the ℓ_2 norm of g is replaced by the ℓ_∞ norm of g. However, there is a critical difference, in that the factor $h_X^{d/2}$ in the second term is missing in [15, Theorem 3.3]. For certain classes of kernels discussed in [11], the paper [10, Proposition 3.6] provides error estimates that at first glance are similar to those in Theorem 5.1 above. However, these estimates do not hold (as Theorem 5.1 does) for all functions in the native space, but instead hold for functions from a finite dimensional space (from a sequence of such approximation spaces). More importantly, the estimates in [10, Proposition 3.6] lack the factor $h_X^{d/2}$. For these two reasons [10, Proposition 3.6] cannot be used instead of Theorem 5.1 in the proof of Theorem 4.1.

Theorem 5.1 follows from [3, Theorem 4.1] which contains the analogous result for bounded connected open domains in $\mathbb{R}^d$ with a Lipschitz-continuous boundary as a special case. Using the ideas from [15] we can lift this result to the sphere and obtain the theorem above. However, as pointed out in Remark 5.1 above the sampling inequality in [15] is not the one that we need and will not be adequate for the proof of Theorem 4.1.

Since we did not find a direct reference for Theorem 5.1 we sketch a proof for the reader's convenience.

First we state without proof a special case of the rather general sampling inequality in [3, Theorem 4.1]. The bounded open domain $\Omega \subset \mathbb{R}^d$ with Lipschitz continuous boundary in Theorem 5.2 below automatically satisfies the required cone property for some radius ρ and angle θ (see [3, page 185] for details). We denote the *mesh norm* of a finite point set $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ in $\overline{\Omega}$ by

$$h_{V,\overline{\Omega}} = \sup_{\mathbf{u} \in \Omega} \min_{\mathbf{v}_j \in V} \|\mathbf{u} - \mathbf{v}_j\|_2.$$

Theorem 5.2 (Arcangéli et al. [3]) *Let $d \geq 2$, let $\Omega \subset \mathbb{R}^d$ be a bounded connected open domain with Lipschitz-continuous boundary, and let $s \in \mathbb{R}$ satisfy $s > d/2$. Then there exist positive constants $\hat{\tilde{c}}_s$ (dependent on θ, ρ, d and s) and $\tilde{\tilde{c}}_s$ (dependent on on Ω, d and s) such that for any finite point set $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ in $\overline{\Omega}$ with $h_{V,\overline{\Omega}} \leq \hat{\tilde{c}}_s$ and any function $g \in W_2^s(\Omega)$*

$$\|g\|_{L_2(\Omega)} \leq \tilde{\tilde{c}}_s \left[h_{V,\overline{\Omega}}^s \|g\|_{W_2^s(\Omega)} + h_{V,\overline{\Omega}}^{d/2} \left(\sum_{j=1}^N [g(\mathbf{v}_j)]^2 \right)^{1/2} \right]. \qquad (5.2)$$

In (5.2) $\| \cdot \|_{L_2(\Omega)}$ is the usual L_2-norm on Ω and, for integer order s, $W_2^s(\Omega)$ is the Sobolev space of those functions in $L_2(\Omega)$ whose distributional derivatives up to (and including) order s are in $L_2(\Omega)$. For the precise definition of the norm and for the definition of $W_2^s(\Omega)$ with fractional order s we refer to [3, Section 2.2]. Note that in [3, Theorem 4.1] the first term on the right-hand side contains a semi-norm for $W_2^s(\Omega)$ which we have estimated above by the corresponding norm. These Sobolev spaces are the same as the ones considered in [15] and in and [17]. It is important that for $g \in W_2^s(\mathbb{R}^d)$ we have $g|_\Omega \in W_2^s(\Omega)$, and the norm $\|g\|_{W_2^s(\Omega)}$ satisfies

$$\|g\|_{W_2^s(\Omega)} \leq c_s \|g\|_{W_2^s(\mathbb{R}^d)}, \tag{5.3}$$

for some $c_s > 0$ (independent of g).

Proof of Theorem 5.1 In order to use Theorem 5.2 we make use of the fact that the Sobolev spaces $H^s = H^s(\mathbb{S}^d)$ can also be defined with the help of charts, giving the same space with an equivalent norm (see [16, Chapter 7.3]). For a better distinction we will write here $W_2^s(\mathbb{S}^d)$ (instead of $H^s(\mathbb{S}^d)$), a space equipped with the equivalent norm $\| \cdot \|_{W_2^s(\mathbb{S}^d)}$. We will only sketch the proof of Theorem 5.1; for any missing details we refer to [15, Section 3].

The sphere $\mathbb{S}^d$ is a compact d-dimensional differentiable manifold without boundary equipped with the atlas $\left\{ C\left(\mathbf{p}_i; \frac{3\pi}{5}\right), \psi_i \right\}_{i=1,2}$, where the chart $\psi_i : C\left(\mathbf{p}_i; \frac{3\pi}{5}\right) \to B(\mathbf{0}; 1)$ is defined as in [15, p.129] with the stereographic projection which maps the open spherical cap

$$C\left(\mathbf{p}_i; \tfrac{3\pi}{5}\right) := \left\{ \mathbf{x} \in \mathbb{S}^d \,\big|\, \mathbf{x} \cdot \mathbf{p}_i > \cos\left(\tfrac{3\pi}{5}\right) \right\} = \left\{ \mathbf{x} \in \mathbb{S}^d \,\big|\, \mathrm{dist}(\mathbf{x}, \mathbf{p}_i) < \tfrac{3\pi}{5} \right\}$$

onto the open ball $B(\mathbf{0}; 1) \subset \mathbb{R}^d$ with radius 1 and center in the origin $\mathbf{0}$. The centers $\mathbf{p}_1$ and $\mathbf{p}_2$ are the north pole $\mathbf{p}_1 = (0, \ldots, 0, 1)^{\mathrm{T}}$ and the south pole $\mathbf{p}_2 = (0, \ldots, 0, -1)^{\mathrm{T}}$, respectively. We define for functions $g : \mathbb{S}^d \to \mathbb{R}$ the projections $\pi_i(g) : \mathbb{R}^d \to \mathbb{R}$ by

$$\pi_i(g)(\mathbf{x}) := \begin{cases} (g \circ \psi_i^{-1})(\mathbf{x}) & \text{for } \mathbf{x} \in B(\mathbf{0}; 1), \\ 0 & \text{otherwise.} \end{cases}$$

Using a partition of unity $\{\chi_i : \mathbb{S}^d \to \mathbb{R}\}_{i=1,2}$ subordinate to the atlas $\left\{ C\left(\mathbf{p}_i; \frac{3\pi}{5}\right), \psi_i \right\}_{i=1,2}$, we define the Sobolev spaces $W_2^s(\mathbb{S}^d)$ for $s \geq 0$ by

$$W_2^s(\mathbb{S}^d) = \left\{ g \in L_2(\mathbb{S}^d) \,\big|\, \pi_i(\chi_i g) \in W_2^s(\mathbb{R}^d) \text{ for } i = 1, 2 \right\},$$

equipped with the norm

$$\|g\|_{W_2^s(\mathbb{S}^d)} = \left(\|\pi_1(\chi_1 g)\|_{W_2^s(\mathbb{R}^d)}^2 + \|\pi_2(\chi_2 g)\|_{W_2^s(\mathbb{R}^d)}^2 \right)^{1/2}.$$

As mentioned before $W_2^s(\mathbb{S}^d) = H^s(\mathbb{S}^d) = H^s$, and the norms $\|\cdot\|_{H^s}$ and $\|\cdot\|_{W_2^s(\mathbb{S}^d)}$ are equivalent. For $s = 2$ we obtain $W_2^0(\mathbb{S}^d) = L_2(\mathbb{S}^d)$, equipped with the norm

$$\|g\|_{W_2^0(\mathbb{S}^d)} = \left(\|\pi_1(\chi_1 g)\|_{L_2(\mathbb{R}^d)}^2 + \|\pi_2(\chi_2 g)\|_{L_2(\mathbb{R}^d)}^2 \right)^{1/2}. \tag{5.4}$$

From [15, Lemma 3.1] we have for $i = 1, 2$

$$\|\psi_i(\mathbf{x}) - \psi_i(\mathbf{y})\|_2 \leq c \, \mathrm{dist}(\mathbf{x}, \mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in C\big(\mathbf{p}_i; \tfrac{3\pi}{5}\big),$$

which implies for a finite point set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ on $\mathbb{S}^d$ that (see [15, Proposition 3.2])

$$h_{\psi_i(X \cap C(\mathbf{p}_i; \frac{3\pi}{5})), \overline{B(\mathbf{0};1)}} \leq c h_X, \quad i = 1, 2. \tag{5.5}$$

From (5.4) it follows that

$$\begin{aligned}
\|g\|_{W_2^0(\mathbb{S}^d)} &\leq \|\pi_1(\chi_1 g)\|_{L_2(\mathbb{R}^d)} + \|\pi_2(\chi_2 g)\|_{L_2(\mathbb{R}^d)} \\
&= \|\pi_1(\chi_1 g)\|_{L_2(B(\mathbf{0};1))} + \|\pi_2(\chi_2 g)\|_{L_2(B(\mathbf{0};1))}, \tag{5.6}
\end{aligned}$$

as the functions $\pi_1(\chi_1 g)$ and $\pi_2(\chi_2 g)$ have compact support in $B(\mathbf{0}; 1)$.

Now let $g \in W_2^s(\mathbb{S}^d)$ with $s > d/2$. Then $\pi_i(\chi_i g) \in W_2^s(\mathbb{R}^n)$ and hence $\pi_i(\chi_i g)|_{B(\mathbf{0};1)} \in W_2^s(B(\mathbf{0};1)), i = 1, 2$. We now consider a term on the right-hand side of (5.6) and apply Theorem 5.2 with $\Omega = B(\mathbf{0}; 1) = \psi_i\big(C(\mathbf{p}_i; \tfrac{3\pi}{5})\big)$ and $V = V_i := \psi_i(X \cap C(\mathbf{p}_i; \tfrac{3\pi}{5}))$:

$$\|\pi_i(\chi_i g)\|_{L_2(B(\mathbf{0};1))}$$
$$\leq \tilde{\tilde{c}}_s \left[h_{V_i, \overline{B(\mathbf{0};1)}}^s \|\pi_i(\chi_i g)\|_{W_2^s(B(\mathbf{0};1))} + h_{V_i, \overline{B(\mathbf{0};1)}}^{d/2} \left(\sum_{j=1}^{N_i} \left[(\pi_i(\chi_i g))(\mathbf{v}_{i,j}) \right]^2 \right)^{1/2} \right],$$

where $V_i = \{\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \ldots, \mathbf{v}_{i,N_i}\}$ and where the point set $X \subset \mathbb{S}^d$ is such that $h_{V_i, \overline{B(\mathbf{0};1)}} \leq \hat{\tilde{c}}_s$.

Using (5.5) and the relationship between the V_i and $X \cap C\big(\mathbf{p}_i; \tfrac{3\pi}{5}\big) \subset X$ yields

$$\|\pi_i(\chi_i g)\|_{L_2(B(\mathbf{0};1))} \leq c' \tilde{\tilde{c}}_s \left[h_X^s \|\pi_i(\chi_i g)\|_{W_2^s(B(\mathbf{0};1))} + h_X^{d/2} \left(\sum_{j=1}^{N} \left[(\chi_i g)(\mathbf{x}_j) \right]^2 \right)^{1/2} \right] \tag{5.7}$$

with $c' := \max\{c^s, c^{d/2}\}$. Applying (5.7) in (5.6) yields for $g \in W_2^s(\mathbb{S}^d)$

$$\|g\|_{W_2^0(\mathbb{S}^d)} \le c' \tilde{\tilde{c}}_s \left(h_X^s \Big[\|\pi_1(\chi_1 g)\|_{W_2^s(B(\mathbf{0};1))} + \|\pi_2(\chi_2 g)\|_{W_2^s(B(\mathbf{0};1))} \Big] \right.$$
$$\left. + h_X^{d/2} \left[\left(\sum_{j=1}^N \big[(\chi_1 g)(\mathbf{x}_j)\big]^2 \right)^{1/2} + \left(\sum_{j=1}^N \big[(\chi_2 g)(\mathbf{x}_j)\big]^2 \right)^{1/2} \right] \right).$$

Now we make use of $|(\chi_i g)(\mathbf{x})| \le |g(\mathbf{x})|$ for all $\mathbf{x} \in \mathbb{S}^d$ for $i = 1, 2$, and $(a + b)^2 \le 2(a^2 + b^2)$ for all $a, b \ge 0$. Using (5.3), we find

$$\|g\|_{W_2^0(\mathbb{S}^d)}$$
$$\le c' \tilde{\tilde{c}}_s \left[\sqrt{2} c_s h_X^s \left(\|\pi_1(\chi_1 g)\|_{W_2^s(\mathbb{R}^d)}^2 + \|\pi_2(\chi_2 g)\|_{W_2^s(\mathbb{R}^d)}^2 \right)^{1/2} + 2 h_X^{d/2} \left(\sum_{j=1}^N \big(g(\mathbf{x}_j)\big)^2 \right)^{1/2} \right]$$
$$\le 2 c' \tilde{\tilde{c}}_s \max\{c_s, 1\} \left[h_X^s \|g\|_{W_2^s(\mathbb{S}^2)} + h_X^{d/2} \left(\sum_{j=1}^N \big(g(\mathbf{x}_j)\big)^2 \right)^{1/2} \right].$$

Since $W_2^0(\mathbb{S}^d)$ and $L_2 = L_2(\mathbb{S}^d)$, and $W_2^s(\mathbb{S}^d)$ and $H^s = H^s(\mathbb{S}^d)$, respectively, are the same space, equipped with equivalent norms, we obtain the estimate in Theorem 5.1. The conditions $h_{V_i, \overline{B(0;1)}} \le \hat{\tilde{c}}_s$ for $i = 1, 2$ translates due to (5.5) to $h_X \le \hat{c}_s$ for some (usually small) constant $\hat{c}_s > 0$. $\qquad\square$

6 Proof of Theorem 4.1

Proof of Theorem 4.1 As an initial step we establish the first result in part (iv) of the theorem. From Lemma 3.1, the inequality (4.3) guarantees that $\|\boldsymbol{\varepsilon}\|_2^2$ is in the range of J, where

$$J(\lambda) := \sum_{i=1}^N \big[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \big(f(\mathbf{x}_i) + \varepsilon_i \big) \big]^2.$$

Since, by Lemma 3.1, J is continuous and strictly monotonically increasing, there exists exactly one $\lambda > 0$ with $J(\lambda) = \|\boldsymbol{\varepsilon}\|_2^2$, that is, there is exactly one λ satisfying (4.4).

Returning now to general λ, we assume that the point set X satisfies $h_X \le \hat{c}_s$, where $\hat{c}_s$ is as in Theorem 5.1. From the assumptions on ϕ, the native space $\mathcal{N}_\phi$ can be identified with H^s, and the norms $\| \cdot \|_\phi$ and $\| \cdot \|_{H^s}$ are equivalent. Thus from applying Theorem 5.1 with the function $g = f_\lambda^{\boldsymbol{\varepsilon}} - f$ in $\mathcal{N}_\phi$ and the point set X and using the equivalence of $\| \cdot \|_\phi$ and $\| \cdot \|_{H^s}$, we obtain

$$\left\|f_\lambda^{\boldsymbol{\varepsilon}} - f\right\|_{L_2} \leq c_s' \left[h_X^s \|f_\lambda^{\boldsymbol{\varepsilon}} - f\|_\phi + h_X^{d/2} \left(\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - f(\mathbf{x}_i) \right]^2 \right)^{1/2} \right]. \quad (6.1)$$

We estimate the two terms in (6.1) separately. In the second term we add and subtract ε_i to obtain

$$\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - f(\mathbf{x}_i) \right]^2 = \sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) + \varepsilon_i \right]^2$$

$$\leq 2 \left(\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 + \sum_{i=1}^N |\varepsilon_i|^2 \right). \tag{6.2}$$

For estimating the right-hand side of (6.2) we proceed differently for the four parameter choices, beginning with the three a priori choices. In these three cases, the right-hand side in (6.2) is estimated by making use of the optimal nature of $f_\lambda^{\boldsymbol{\varepsilon}}$: from (4.1)

$$\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 + \sum_{i=1}^N |\varepsilon_i|^2 \leq \mu_\lambda(f_\lambda^{\boldsymbol{\varepsilon}}) + \sum_{i=1}^N |\varepsilon_i|^2$$

$$\leq \mu_\lambda(f) + \sum_{i=1}^N |\varepsilon_i|^2 = \sum_{i=1}^N |\varepsilon_i|^2 + \lambda \|f\|_\phi^2 + \sum_{i=1}^N |\varepsilon_i|^2 = 2\|\boldsymbol{\varepsilon}\|_2^2 + \lambda \|f\|_\phi^2. \tag{6.3}$$

For $\lambda = \|\boldsymbol{\varepsilon}\|_2^2$ we obtain from (6.3)

$$\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 + \sum_{i=1}^N |\varepsilon_i|^2 \leq \|\boldsymbol{\varepsilon}\|_2^2 \left(2 + \|f\|_\phi^2 \right). \tag{6.4}$$

Hence, from (6.2) and (6.4), for $\lambda = \|\boldsymbol{\varepsilon}\|_2^2$,

$$\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - f(\mathbf{x}_i) \right]^2 \leq 2\|\boldsymbol{\varepsilon}\|_2^2 \left(2 + \|f\|_\phi^2 \right) \leq 4\|\boldsymbol{\varepsilon}\|_2^2 \left(1 + \frac{1}{\sqrt{2}} \|f\|_\phi \right)^2. \tag{6.5}$$

For the second a priori choice $\lambda = \frac{\|\boldsymbol{\varepsilon}\|_2^2}{\|f\|_\phi^2}$, (6.3) gives

$$\sum_{i=1}^N \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 + \sum_{i=1}^N |\varepsilon_i|^2 \leq 3\|\boldsymbol{\varepsilon}\|_2^2. \tag{6.6}$$

Hence, from (6.2) and (6.6), for $\lambda = \frac{\|\boldsymbol{\varepsilon}\|_2^2}{\|f\|_\phi^2}$,

$$\sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - f(\mathbf{x}_i) \right]^2 \leq 6\|\boldsymbol{\varepsilon}\|_2^2. \qquad (6.7)$$

For the third a priori choice we do not yet insert the value for λ but obtain from (6.2) and (6.3) that

$$\sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - f(\mathbf{x}_i) \right]^2 \leq 4 \left(\|\boldsymbol{\varepsilon}\|_2^2 + \frac{\lambda}{2} \|f\|_\phi^2 \right) \leq 4 \left(\|\boldsymbol{\varepsilon}\|_2 + \sqrt{\frac{\lambda}{2}} \|f\|_\phi \right)^2. \quad (6.8)$$

If instead we choose λ by Morozov's discrepancy principle, then

$$\sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 + \sum_{i=1}^{N} |\varepsilon_i|^2 = \|\boldsymbol{\varepsilon}\|_2^2 + \sum_{i=1}^{N} |\varepsilon_i|^2 = 2\|\boldsymbol{\varepsilon}\|_2^2. \qquad (6.9)$$

Then using (6.9) to estimate the right-hand side in (6.2) we obtain

$$\sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - f(\mathbf{x}_i) \right]^2 \leq 4\|\boldsymbol{\varepsilon}\|_2^2. \qquad (6.10)$$

For all four parameter choice strategies we now estimate the first term in (6.1). From the triangle inequality, we have,

$$\left\| f_\lambda^{\boldsymbol{\varepsilon}} - f \right\|_\phi \leq \left\| f_\lambda^{\boldsymbol{\varepsilon}} \right\|_\phi + \left\| f \right\|_\phi, \qquad (6.11)$$

in which the first term can be estimated using the optimality of $f_\lambda^{\boldsymbol{\varepsilon}}$ as follows:

$$\begin{aligned} \left\| f_\lambda^{\boldsymbol{\varepsilon}} \right\|_\phi^2 &= \frac{1}{\lambda} \left(\lambda \left\| f_\lambda^{\boldsymbol{\varepsilon}} \right\|_\phi^2 \right) = \frac{1}{\lambda} \left(\mu_\lambda(f_\lambda^{\boldsymbol{\varepsilon}}) - \sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 \right) \\ &\leq \frac{1}{\lambda} \left(\mu_\lambda(f) - \sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 \right) \\ &= \frac{1}{\lambda} \left(\|\boldsymbol{\varepsilon}\|_2^2 - \sum_{i=1}^{N} \left[f_\lambda^{\boldsymbol{\varepsilon}}(\mathbf{x}_i) - \left(f(\mathbf{x}_i) + \varepsilon_i \right) \right]^2 \right) + \|f\|_\phi^2. \end{aligned} \qquad (6.12)$$

For the three a priori parameter choices the negative term in the last inequality in (6.12) can be omitted. For the a priori parameter choice $\lambda = \|\boldsymbol{\varepsilon}\|_2^2$ this gives

$$\left\| f_\lambda^{\boldsymbol{\varepsilon}} \right\|_\phi \leq \left(1 + \|f\|_\phi^2 \right)^{1/2} \leq 1 + \|f\|_\phi. \qquad (6.13)$$

The result for the first a priori choice then follows from (6.1), (6.5), (6.11), and (6.13).

For the a priori parameter choice $\lambda = \frac{\|\boldsymbol{\varepsilon}\|_2^2}{\|f\|_\phi^2}$, omitting the negative term in (6.12) gives

$$\|f_\lambda^{\boldsymbol{\varepsilon}}\|_\phi \leq \sqrt{2}\|f\|_\phi. \tag{6.14}$$

The result for the second a priori choice then follows from (6.1), (6.7), (6.11), and (6.14).

For the third a priori choice, as we do not yet insert the value for λ, the inequality (6.12) in this case yields

$$\|f_\lambda^{\boldsymbol{\varepsilon}}\|_\phi \leq \left(\frac{1}{\lambda}\|\boldsymbol{\varepsilon}\|_2^2 + \|f\|_\phi^2\right)^{1/2} \leq \frac{1}{\sqrt{\lambda}}\|\boldsymbol{\varepsilon}\|_2 + \|f\|_\phi. \tag{6.15}$$

We then find from (6.1), (6.8), (6.11), and (6.15)

$$\|f_\lambda^{\boldsymbol{\varepsilon}} - f\|_{L_2} \leq 2c_s'\left[h_X^s\left(\frac{1}{2\sqrt{\lambda}}\|\boldsymbol{\varepsilon}\|_2 + \|f\|_\phi\right) + h_X^{d/2}\left(\|\boldsymbol{\varepsilon}\|_2 + \sqrt{\frac{\lambda}{2}}\|f\|_\phi\right)\right]. \tag{6.16}$$

It is easy to see that the parameter choice in part (iii) of the theorem minimizes the right-hand side of (6.16), and that the resulting minimum value is

$$\|f_\lambda^{\boldsymbol{\varepsilon}} - f\|_{L_2} \leq 2c_s'\left(h_X^s\|f\|_\phi + 2^{1/4}h_X^{s/2+d/4}\|f\|_\phi^{1/2}\|\boldsymbol{\varepsilon}\|_2^{1/2} + h_X^{d/2}\|\boldsymbol{\varepsilon}\|_2\right)$$
$$\leq (1 + 2^{-3/4})2c_s'\left(h_X^s\|f\|_\phi + h_X^{d/2}\|\boldsymbol{\varepsilon}\|_2\right),$$

where the last inequality follows from the basic inequality $2ab \leq a^2 + b^2$.

For Morozov's discrepancy principle, the negative term in (6.12) cancels $\|\boldsymbol{\varepsilon}\|_2^2$, giving

$$\|f_\lambda^{\boldsymbol{\varepsilon}}\|_\phi \leq \|f\|_\phi. \tag{6.17}$$

The result for Morozov's discrepancy principle follows now from (6.1), (6.10), (6.11), and (6.17). $\qquad\square$

7 Numerical results

In this section we present a numerical test for $\mathbb{S}^2$ that illustrates the performance of the method and discuss the practical evaluation and performance of the four parameter choices in Theorem 4.1 as well as the L-curve heuristic [12].

The function to be approximated is in Cartesian coordinates given by

$$f(x, y, z) = e^{x+y+z} + 50\big(y - \cos(\pi/3)\big)_+^3, \quad (x, y, z)^T \in \mathbb{R}^3, \tag{7.1}$$

where r_+ is defined to be r if $r \geq 0$ and to be zero otherwise. It can be shown that this function is in H^s for any $s < 3.5$. The function is plotted on the polar coordinate grid in Fig. 1.
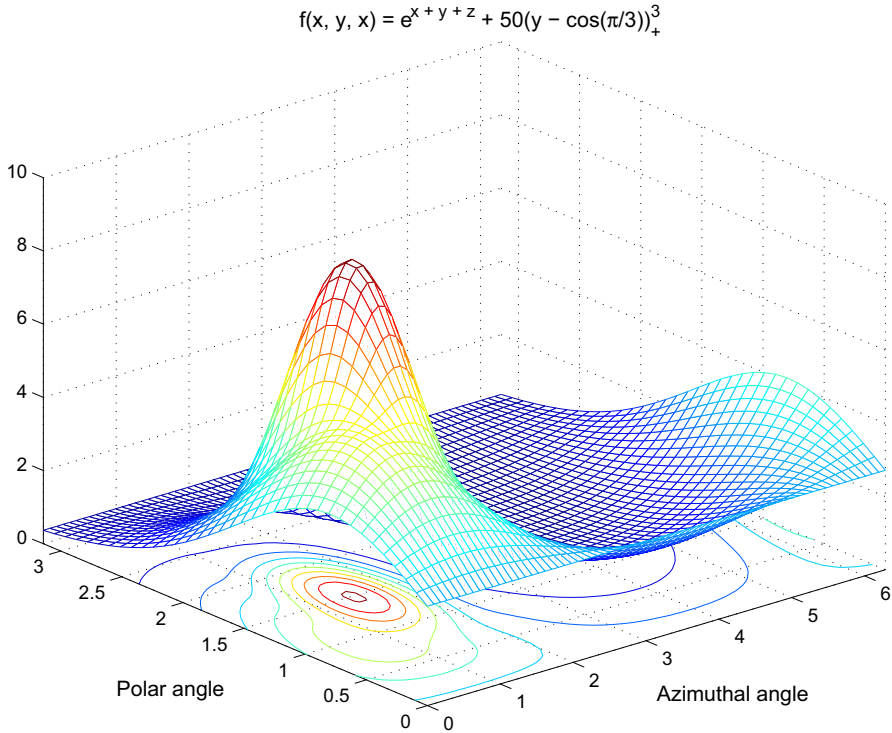
$$f(x, y, x) = e^{x + y + z} + 50(y - \cos(\pi/3))_+^3$$

Fig. 1 The approximated function (7.1)

The point set $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ was a set of $N = 6084$ points chosen to have a polynomial basis matrix with large determinant (see [22]). This provides a uniformly distributed point set on $\mathbb{S}^2$ with mesh norm $h_X = 0.0372$ and separation distance $\min_{i \neq j} \operatorname{dist}(\mathbf{x}_i, \mathbf{x}_j) = 0.0413$. Noise was added to the data by generating independent samples from the standard normal distribution, and multiplying these numbers by the factor 0.1 before they were added to the (exact) function data on X. For our particular samples we find $\|\boldsymbol{\varepsilon}\|_2^2 = 61.8275$.

In the implementation, we used the compactly supported Wendland radial basis function (see [28,30]) $\phi(\mathbf{x}, \mathbf{y}) = \psi(\|\mathbf{x} - \mathbf{y}\|_2)$, where

$$\psi(r) = (4r + 1)(1 - r)_+^4, \quad r \geq 0.$$

The asymptotic behavior of the Fourier coefficients of this RBF is $a_\ell \asymp (\ell + 1)^{-5}$ (see [19]), and hence the native space can be identified with $H^{2.5}$. Since f belongs to a still smoother space, our theory applies with $s = 2.5$. Note that on the unit sphere the support of each unscaled radial basis function covers one quarter of the area of the sphere, affecting the sparsity of $\mathbf{A}$. For the chosen set of $N = 6084$ well distributed points and this Wendland function the sparsity of $\mathbf{A}$ is 24.99% while the 2-norm condition number is $\kappa_2(\mathbf{A}) = 5.9 \times 10^5$.

Table 1 RBF smoothing approximation for a range of smoothing parameters λ

λ	Discrepancy	$\|f_\lambda^{\varepsilon}\|_\phi^2$	L_2 error	L_∞ error
0	7.7740×10^{-11}	6.2983×10^4	3.1079×10^{-1}	3.9720×10^{-1}
3.052×10^{-5}	9.2339×10^{-2}	5.6743×10^4	3.0404×10^{-1}	3.8758×10^{-1}
6.104×10^{-5}	3.3002×10^{-1}	5.1502×10^4	2.9800×10^{-1}	3.7879×10^{-1}
1.221×10^{-4}	1.0756×10^0	4.3219×10^4	2.8759×10^{-1}	3.6323×10^{-1}
2.441×10^{-4}	3.0389×10^0	3.2184×10^4	2.7144×10^{-1}	3.3810×10^{-1}
4.883×10^{-4}	7.1094×10^0	2.0570×10^4	2.4963×10^{-1}	3.0853×10^{-1}
9.766×10^{-4}	1.3543×10^1	1.1245×10^4	2.2420×10^{-1}	2.8145×10^{-1}
1.953×10^{-3}	2.1444×10^1	5.4417×10^3	1.9802×10^{-1}	2.5011×10^{-1}
3.906×10^{-3}	2.9437×10^1	2.4791×10^3	1.7318×10^{-1}	2.1644×10^{-1}
7.813×10^{-3}	3.6552×10^1	1.1531×10^3	1.5065×10^{-1}	1.8215×10^{-1}
1.563×10^{-2}	4.2429×10^1	6.0371×10^2	1.3070×10^{-1}	1.4920×10^{-1}
3.125×10^{-2}	4.7088×10^1	3.8557×10^2	1.1328×10^{-1}	1.2222×10^{-1}
6.250×10^{-2}	5.0700×10^1	3.0095×10^2	9.8313×10^{-2}	1.0326×10^{-1}
1.250×10^{-1}	5.3468×10^1	2.6851×10^2	8.5714×10^{-2}	9.0829×10^{-2}
2.500×10^{-1}	5.5605×10^1	2.5601×10^2	7.5685×10^{-2}	8.1991×10^{-2}
5.000×10^{-1}	5.7404×10^1	2.5080×10^2	6.9974×10^{-2}	7.3962×10^{-2}
1.000×10^0	5.9560×10^1	2.4777×10^2	7.6271×10^{-2}	7.8826×10^{-2}
2.000×10^0	6.4371×10^1	2.4450×10^2	1.1320×10^{-1}	1.3068×10^{-1}
4.000×10^0	8.0267×10^1	2.3917×10^2	2.0712×10^{-1}	2.3671×10^{-1}
8.000×10^0	1.3818×10^1	2.2947×10^2	3.9937×10^{-1}	4.4935×10^{-1}
1.600×10^1	3.4545×10^2	2.1207×10^2	7.6396×10^{-1}	8.5439×10^{-1}
3.200×10^1	1.0326×10^3	1.8306×10^2	1.4132×10^0	1.5596×10^0
6.400×10^1	3.0140×10^3	1.4084×10^2	2.4659×10^0	2.6446×10^0
1.280×10^2	7.6123×10^3	9.1117×10^1	3.9453×10^0	4.0654×10^0

The RBF smoothing approximant was computed for a sequence of increasing smoothing parameters, given by $\lambda = 0$ and $\lambda = 2^\zeta$, for $\zeta = -15, -14.9, \ldots, 6.9, 7$. Table 1 lists a subset corresponding to $\lambda = 0$, $\lambda = 2^j$ for $j = -15, -14, \ldots, 6, 7$, while Fig. 2 plots the full results. For each value of the smoothing parameter λ, we calculate the discrepancy

$$J(\lambda) = \sum_{j=1}^{N} \left[f_\lambda^{\varepsilon}(\mathbf{x}_j) - \left(f(\mathbf{x}_j) + \varepsilon_j \right) \right]^2,$$

$\|f_\lambda^{\varepsilon}\|_\phi^2 = \boldsymbol{\alpha}^T \mathbf{A} \boldsymbol{\alpha}$ where $(\mathbf{A} + \lambda \mathbf{I})\boldsymbol{\alpha} = \mathbf{f} + \boldsymbol{\varepsilon}$, and estimates of the L_2 and the L_∞ norms of the error $f_\lambda^{\varepsilon} - f$. The L_2 error was estimated using a 16474 point spherical 181-design (equal weight quadrature rule exact for all spherical polynomials of degree at most 181), while the L_∞ norm was estimated by using a local maximization routine
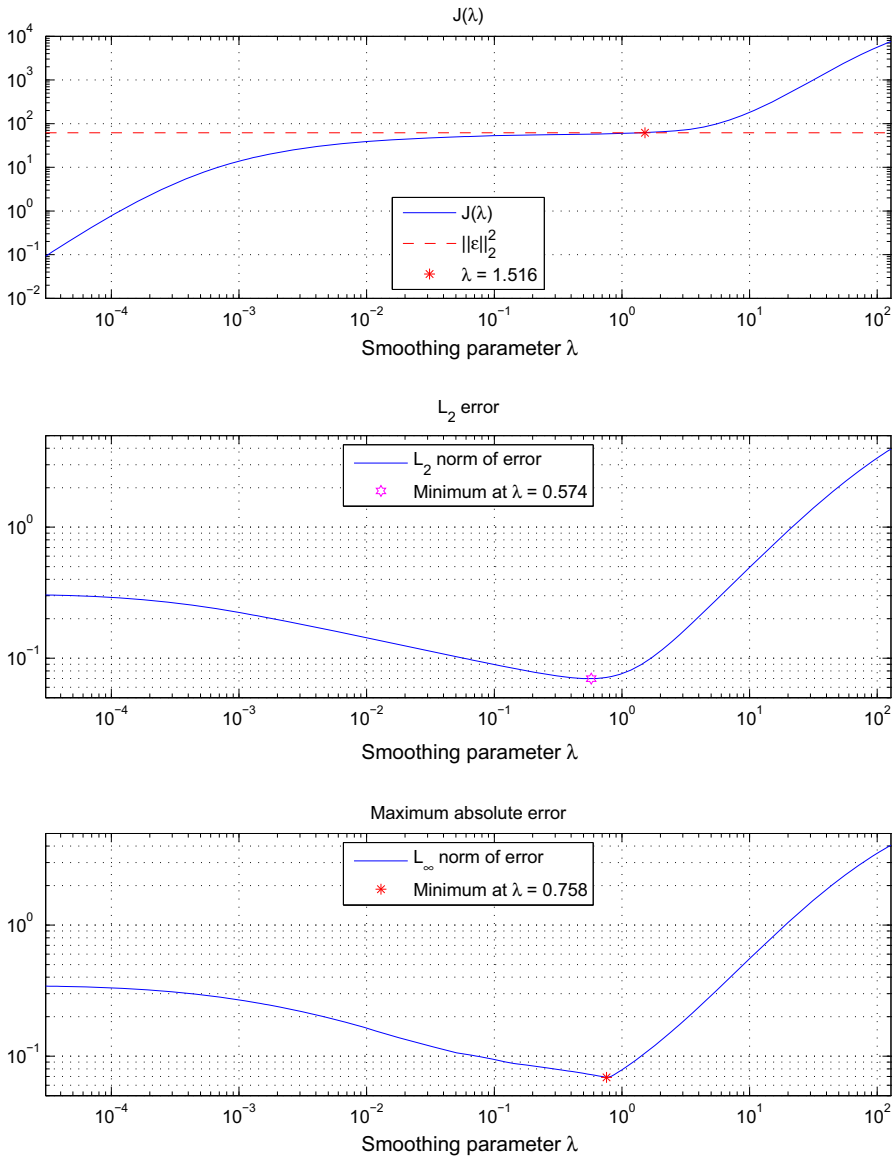
Fig. 2 Variation of $J(\lambda)$, $\|f_\lambda^\varepsilon - f\|_{L_2}$ and $\|f_\lambda^\varepsilon - f\|_C$

starting from the circumcenter of each of the triangles in a Delaunay triangulation of the RBF centers, and checked against the pointwise maximum of the error at the points in the spherical design.

The smallest L_2 error of 6.9841×10^{-2} is obtained for $\lambda = 0.574$, while the smallest L_∞ error of 6.8190×10^{-2} is obtained for $\lambda = 0.758$, as illustrated in Fig. 2.

The a priori parameter choice $\lambda = \|\boldsymbol{\varepsilon}\|_2^2 = 61.8275$ gives an over-smoothed approximation with L_2 error and L_∞ error over 2. A disadvantage of this choice for λ is that it depends on the scaling (units) used for the data.

The second and third a priori choices $\lambda = \frac{\|\boldsymbol{\varepsilon}\|_2^2}{\|f\|_\phi^2}$ and $\lambda = \frac{1}{\sqrt{2}} h_X^{s-d/2} \frac{\|\boldsymbol{\varepsilon}\|_2}{\|f\|_\phi}$ in Theorem 4.1 have a major disadvantage in practice, in that they require a knowledge of the true norm $\|f\|_\phi$, which of course is generally not available. In this test example, where in fact we know f, we can estimate the value of the ϕ norm with reasonable accuracy by the ϕ norm of the minimal-norm interpolant of the exact function values at the data points. (In the language of Section 3 we set $\lambda = 0$ and $\boldsymbol{\varepsilon} = \mathbf{0}$, and use $\|f_0\|_\phi$ to approximate $\|f\|_\phi$.) Using the value $\|f_0\|_\phi = 15.79$, the second a priori choice gives $\lambda \approx \frac{\|\boldsymbol{\varepsilon}\|_2^2}{\|f_0\|_\phi^2} = 0.248$, with L_2 error of 7.58×10^{-2}, just less than the minimal L_2 error (see the second plot in Fig. 2).

The third a priori choice gives $\lambda = 2.5 \times 10^{-3}$ and an L_2 error of 1.81×10^{-1}. This value for λ is lower than those giving the minimal L_2 norm of the errors, resulting in an undersmoothed approximation and again a relatively large error. Moreover, as $\|f_0\|_\phi \leq \|f\|_\phi$, using the true norm $\|f\|_\phi$ will give a still smaller value of λ, making the undersmoothing worse (see Fig. 2). And so far the calculation of the second and third a priori choices is not at all practical, since it assumes a knowledge of exact function values at the data points. In practice perhaps the best a user could do is to approximate $\|f\|_\phi$ by the ϕ norm of a smoothed approximant $f_\lambda^{\boldsymbol{\varepsilon}}$ for some reasonable choice of λ.

Morozov's discrepancy principle requires in practice that we compute $f_\lambda^{\boldsymbol{\varepsilon}}$ for a selection of parameters λ until we have found a value of λ for which (4.4) is satisfied
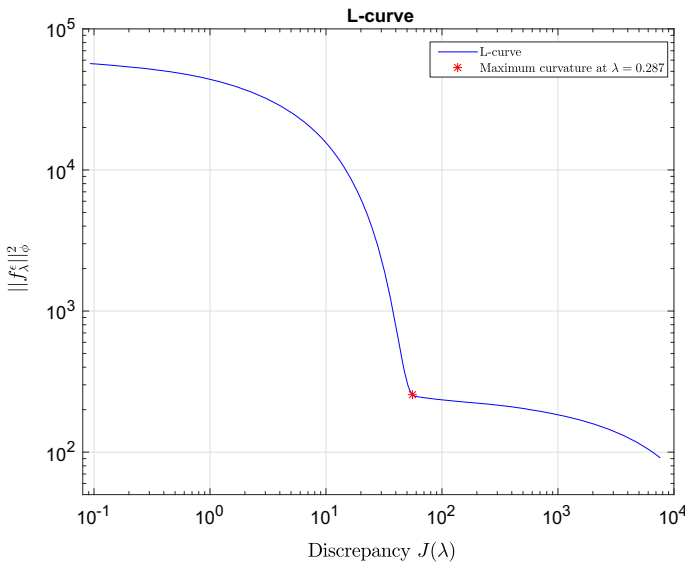

Fig. 3 L-curve method: log-log plot of $\|f_\lambda^{\boldsymbol{\varepsilon}}\|_\phi^2$ against $J(\lambda)$

accurately enough. More specifically the discrepancy principle requires $J(\lambda) = \|\boldsymbol{\varepsilon}\|_2^2$, which for this example is satisfied when $J(\lambda) = \|\boldsymbol{\varepsilon}\|_2^2 = 61.8275$, at the value $\lambda = 1.516$ as illustrated in the first plot of Fig. 2. This value of λ gives an L_2 error of 9.32×10^{-2}, around 35% larger than the minimal L_2 error of 6.98×10^{-2}. Thus the λ predicted by the discrepancy principle is a good choice, even though it is slightly too large and slightly over-smooths.

A heuristic strategy to determine the smoothing parameter is the L-curve method (see [12]): for a range of values of λ, the two parts $\|f_\lambda^{\boldsymbol{\varepsilon}}\|_\phi^2$ and $J(\lambda)$ in the minimized functional (4.1) are plotted against each other in a log-log plot (see Fig. 3). The resulting curve $x(\lambda) = \log J(\lambda)$, $y(\lambda) = \log \|f_\lambda^{\boldsymbol{\varepsilon}}\|_\phi^2$ should then be L-shaped, which is clearly
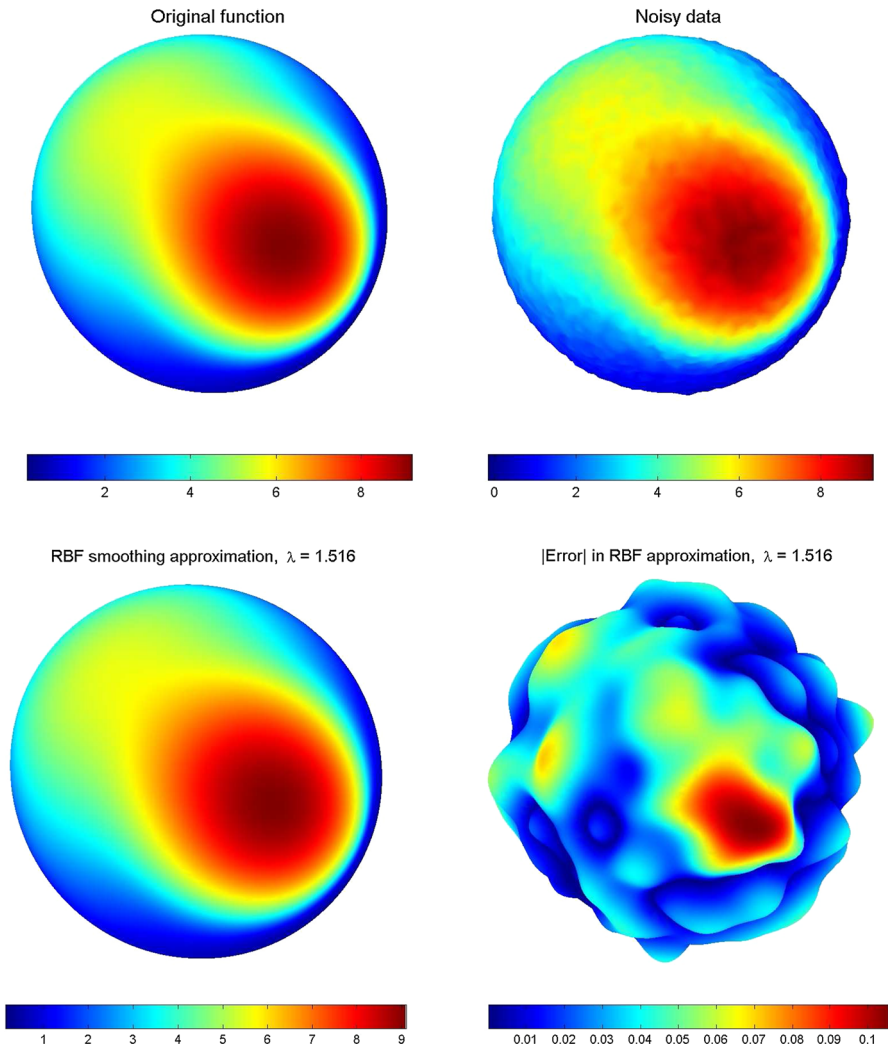

Fig. 4 Plots of test function, noisy data, RBF smoothing approximation and error for $\lambda = 1.516$

the case in Fig. 3. The smoothing parameter is chosen as the value of λ corresponding to the corner of the L, defined as the point with maximal curvature $|x'y'' - y'x''|/(x'^2 + y'^2)^{\frac{3}{2}}$. We used the full range of values of $\lambda = 2^\zeta$, $\zeta = -15, -14.9, \ldots, 6.9, 7$, for the L-curve and then finite differences to estimate the derivatives x', x'', y', y'' and hence the curvature, giving the parameter $\lambda = 0.287$ and an L_2 error of 7.41×10^{-2}. From Table 1 it is clear that this is a fairly good prediction, but one which, unlike the methods considered in this paper, lacks theoretical support.

Finally Fig. 4 gives plots of the test function (7.1), the noisy data, the RBF smoothing approximation and the error for $\lambda = 1.516$, the value predicted by the discrepancy principle.

To summarise, for approximation from noisy scattered data on the sphere, regularized by the square of a native space norm, we have found L_2 error bounds for four different choices of the smoothing parameter. Of the four choices the best result (by a small margin theoretically), and the most practical, is the parameter choice from the discrepancy principle. The actual numerical cost of the discrepancy principle depends on the number of values for the parameter λ for which the approximation has to be computed in order to get a sufficiently good estimate for the smoothing parameter λ of the discrepancy principle.

References

1. Alfeld, P., Neamtu, M., Schumaker, L.L.: Fitting scattered data on sphere-like surfaces using spherical splines. J. Comput. Appl. Math. **73**, 5–43 (1996)
2. An, C., Chen, X., Sloan, I.H., Womersley, R.S.: Regularized least squares approximations on the sphere using spherical designs. SIAM J. Numer. Anal. **50**(3), 1513–1534 (2012). (Corrigendum: SIAM J. Numer. Anal. **52**(4), 2205–2206 (2014).)
3. Arcangéli, R., López de Silanes, M.C., Torrens, J.J.: An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing. Numer. Math. **107**, 181–211 (2007)
4. Bauer, F., Lukas, M.A.: Comparing parameter choice methods for regularization of ill-posed problems. Math. Comput. Simul. **81**, 1795–1841 (2011)
5. Chen, D., Menegatto, V.A., Sun, X.: A necessary and sufficient condition for strictly positive definite functions on spheres. Proc. Am. Math. Soc. **131**, 2733–2740 (2003)
6. Cucker, F., Zhou, D.X.: Learning Theory: An Approximation Theory Viewpoint. Cambridge University Press, Cambridge (2007)
7. Erdélyi, A. (ed.), Magnus, W., Oberhettinger, F., Tricomi, F.G. (research associates): Higher Transcendental Functions, Vol. III. California Institute of Technology, Bateman Manuscript Project, McGraw-Hill Book Company Inc, New York (1953)
8. Freeden, W.: Multiscale Modelling of Spaceborne Geodata. B. G. Teubner, Stuttgart (1999)
9. Freeden, W., Gervens, T., Schreiner, M.: Constructive Approximation on the Sphere (with Applications to Geomathematics). Clarendon Press, Oxford (1998)
10. Hangelbroek, T., Narcowich, F.J., Sun, X., Ward, J.D.: Kernel approximation on manifolds II: the L_∞ norm of the L_2 projector. SIAM J. Math. Anal. **43**(2), 662–684 (2011)
11. Hangelbroek, T., Narcowich, F.J., Ward, J.D.: Kernel approximation on manifolds I: bounding the Lebesgue constant. SIAM J. Math. Anal. **42**(4), 1732–1760 (2010)
12. Hansen, P.C.: Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. **34**(4), 561–580 (1992)

13. Kirsch, A.: An Introduction to the Mathematical Theory of Inverse Problems. Springer, New York (1996)
14. Krebs, J., Louis, A.K., Wendland, H.: Sobolev error estimates and a priori parameter selection for semi-discrete Tikhonov regularization. J. Inverse Ill Posed Probl. **17**, 845–869 (2009)
15. Le Gia, Q.T., Narcowich, F.J., Ward, J.D., Wendland, H.: Continuous and discrete least-squares approximation by radial basis functions on spheres. J. Approx. Theory **143**, 124–133 (2006)
16. Lions, J.L., Magenes, E.: Non-Homogeneous Boundary Value Problems and Applications, vol. 1. Springer, Berlin (1972)
17. McLean, W.: Strongly Elliptic Systems and Boundary Integral Equations. Cambridge University Press, Cambridge (2000)
18. Mhaskar, H.M., Narcowich, F.J., Ward, J.D.: Spherical Marcinkiewicz–Zygmund inequalities. Math. Comput. **70**(235), 1113–1130 (2001). (Corrigendum: Math. Comput. **71**, 453–454 (2002).)
19. Narcowich, F.J., Ward, J.D.: Scattered data interpolation on spheres: error estimates and locally supported basis functions. SIAM J. Math. Anal. **33**(6), 1393–1410 (2002)
20. Reimer, M.: Multivariate Polynomial Approximation. Birkhäuser, Basel (2003)
21. Schoenberg, I.J.: Positive definite functions on spheres. Duke Math. J. **9**, 96–108 (1942)
22. Sloan, I.H., Womersley, R.S.: Extremal systems of points and numerical integration on the sphere. Adv. Comput. Math. **21**, 107–125 (2004)
23. Szegö, G.: Orthogonal Polynomials, vol. XXIII, 4th edn. American Mathematical Society Colloquium Publications, American Mathematical Society, Providence (1975)
24. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
25. von Golitschek, M., Schumaker, L.L.: Data fitting by penalized least squares. In: Algorithms for Approximation. II (Shrivenham, 1988), pp. 210–227. Chapman and Hall, London (1990)
26. Wahba, G.: Spline models for observational data. In: CBMS-NSF regional conference series in applied mathematics, vol. 59, SIAM, Philadelphia (1990)
27. Wei, T., Hon, Y.C., Wang, Y.B.: Reconstruction of numerical derivatives from scattered noisy data. Inverse Problems **21**, 657–672 (2005)
28. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. **4**, 389–396 (1995)
29. Wendland, H.: Error estimates for interpolation by compactly supported radial basis functions of minimal degree. J. Approx. Theory **93**, 258–272 (1998)
30. Wendland, H.: Scattered Data Approximation. Cambridge University Press, Cambridge (2005)
31. Xu, Y., Cheney, E.W.: Strictly positive definite functions on spheres. Proc. Am. Math. Soc. **116**, 977–981 (1992)
32. Zhang, Y., Li, R., Tsai, C.-L.: Regularization parameter selections via generalized information criterion. J. Am. Stat. Assoc. **105**(489), 312–323 (2010)