Numer. Math. (2010) 116:493–517
DOI 10.1007/s00211-010-0313-8

Numerische Mathematik

Multiscale analysis in Sobolev spaces on bounded domains

Holger Wendland

Received: 14 November 2009 / Revised: 22 March 2010 / Published online: 21 May 2010
© Springer-Verlag 2010

Abstract We study a multiscale scheme for the approximation of Sobolev functions on bounded domains. Our method employs scattered data sites and compactly supported radial basis functions of varying support radii at scattered data sites. The actual multiscale approximation is constructed by a sequence of residual corrections, where different support radii are employed to accommodate different scales. Convergence theorems for the scheme are proven, and it is shown that the condition numbers of the linear systems at each level are independent of the level, thereby establishing for the first time a mathematical theory for multiscale approximation with scaled versions of a single compactly supported radial basis function at scattered data points on a bounded domain.

Mathematics Subject Classification (2000) 65J10 · 46E35

1 Introduction

Radial basis functions form an important tool in modern approximation. They allow the easy construction of approximation spaces in arbitrary dimensions and with arbitrary smoothness. These discretization spaces are known to lead to approximations which exhibit both excellent approximation properties but, unfortunately, also poor conditioning and a high computational cost. This is a well-known "trade-off" principle which is related to the smoothness of the basis function and can briefly be summarized as smoothness implies high convergence order plus ill conditioning [10].

Since the advent of compactly supported radial basis functions in 1995 in [18] and particularly in [15], there was always the hope to overcome this "trade-off" principle.

H. Wendland (✉)
Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, England
e-mail: holger.wendland@maths.ox.ac.uk

However, early numerical tests (see for example [11]) indicated that an additional trade-off principle comes into play, now depending on the choice of the support radius. The general observation was that a small support radius leads to a very well-conditioned sparse system but also a poor approximation rate, while a larger support radius yields excellent approximation at the price of a badly conditioned, full system.

Consequently, some effort has been spent on finding the "optimal" support radius, as it has similarly been spent on finding the optimal shape parameter for Gaussians and (inverse) Multiquadrics. However, it seems unnatural to employ only one support radius for all possible data sites. It seems more natural to allow different support radii to capture, for example, multiscale phenomena within the data.

While it is, up to now, an open problem whether interpolation in scattered data sites using a different support radius at every site leads always to an invertible interpolation matrix, another approach came up quite early [2,9,11], which reflects the idea of possibly having multiple scales represented within the data. A multilevel type algorithm has been employed, which proceeds in the following way. First, a coarse data set and a large support radius are chosen and the target function is interpolated in this data set. This is supposed to capture the large-scale variations of the target function. This means, however, that the residual between the target function and this first interpolant lives on a finer scale. This obviously leads to the strategy, next to choose a finer data set and a smaller support radius and to interpolate the residual at the finer data set, capturing now the finer details. The sum of both interpolants obviously interpolates the target function at the data sites of the finer data set, which was one motivation, in the original work [2,9,11] for choosing the data sites nested. Obviously, the process does not have to stop after two such levels, but can be iterated accordingly.

As simple as this approach is, besides numerical evidence of convergence, no proof has been given until very recently, where it has been shown that this multiscale scheme converges in the special situation that the data sites are located on a sphere, see [6]. Earlier papers by Narcowich, Schaback and Ward [7] and Hales and Levesley [5] are dealing with completely different situations than the one we are interested in here.

It is the goal of this paper to show that the convergence analysis also holds for bounded domains, giving a rigorous proof in this case for the very first time. Furthermore, in contrast to [6] we do not restrict ourselves to the case where the smoothness of the target function is coupled to the smoothness of the radial basis function. We also do not restrict ourselves to the case of interpolation but also study "smoothed approximations", which is important, for example, for noisy data. We show that even in this situation the multiscale algorithm converges, provided the smoothing parameter is chosen carefully.

The paper is organized as follows. In the next section we introduce the concept of scaling and the multiscale approximation algorithm. In section three we discuss function spaces relevant to this paper. The fourth section is devoted to the analysis of the multiscale approximation algorithm for the interpolation of smooth target functions, while the fifth section deals with the interpolation of rough target functions. In section six, we replace interpolation with a penalized least-squares algorithm and show convergence of the multiscale algorithm in this situation. The final section deals with numerical efficiency and gives examples.

2 Multiscale approximation

In this section we collect necessary material on what we mean by multiscale approximation. We also introduce the necessary notation.

We start by introducing discrete approximation spaces. There are two ingredients in building these approximation spaces: point sets and a scaled radial basis function.

Let $\Omega \subseteq \mathbb{R}^d$ be given. Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ be a finite point set in Ω. We will associate to this point set the following two measures:

$$h_{X,\Omega} := \sup_{\mathbf{x}\in\Omega} \min_{\mathbf{x}_j\in X} \|\mathbf{x}-\mathbf{x}_j\|_2,$$
$$q_X := \min_{j\neq k} \|\mathbf{x}_j-\mathbf{x}_k\|_2.$$

The first one is usually called fill distance or mesh norm and measures how well the points in X cover the region Ω. It is the radius of the largest ball without a point from X, entirely contained in Ω. The second one is called separation distance and measures, together with $h_{X,\Omega}$ uniformity of the data set.

We will, however, not deal with only one data set, but with a sequence of data sets $X_1, X_2, \dots$ with mesh norms $h_j = h_{X_j,\Omega}$, which are monotonically decreasing. To ensure a certain uniformity in decrease, we will assume that $h_{j+1} \approx \mu h_j$ for some fixed $\mu \in (0, 1)$. For some of our results we will require that the sequence is quasi-uniform, which means that there is a constant c_q such that, with $q_j = q_{X_j}$,

$$q_j \le h_j \le c_q q_j.$$

Next, we will pick a kernel $\Phi_j : \Omega \times \Omega \to \mathbb{R}$ for each level. In our application this kernel will be given by the scaled version of a translation invariant function. To be more precise, we assume that there is a compactly supported function $\Phi : \mathbb{R}^d \to \mathbb{R}$ with support in the unit ball $B(\mathbf{0}, 1)$ and that, for each level, there is a scaling parameter $\delta_j > 0$ such that we can define

$$\Phi_j(\mathbf{x}, \mathbf{y}) = \delta_j^{-d}\Phi((\mathbf{x}-\mathbf{y})/\delta_j).$$

If we consider $\mathbf{y}$ to be fixed, it follows that the function $\Phi_j(\cdot, \mathbf{y})$ has support in $B(\mathbf{y}, \delta_j)$, the ball with radius δ_j and center $\mathbf{y}$.

As we assume that the data sets become denser and denser, we will assume that the support radii become smaller and smaller, usually in the same way, i.e., we will assume that $\delta_j = \nu h_j$ with a constant $\nu > 0$, which we will link to μ later on. Note that this also leads to $\delta_{j+1} \approx \mu\delta_j$.

With the data sets and the associated kernels at hand, we can build approximation spaces of the form

$$W_j = \operatorname{span}\{\Phi_j(\cdot, \mathbf{x}) : \mathbf{x} \in X_j\}.$$

representing details on level j. Thus, the approximation of our function will come from the sum of these spaces, i.e., we have to investigate the approximation power of

$$V_n := W_1 + W_2 + \cdots + W_n$$

for $n \to \infty$.

So far, we have deliberately followed the path of a classical multiresolution analysis, though our sets W_j are, by construction, finite dimensional, which is a major difference to the classical wavelet theory. Moreover, to determine a concrete approximation from $V_n = W_1 + \cdots + W_n$, we will proceed differently. We will use the algorithm outlined in Algorithm 1.

Algorithm 1: Multiscale approximation

Input : Right-hand side f, Number of levels n
Output : Approximate solution $f_n \in V_n = W_1 + \cdots + W_n$

Set $f_0 = 0$, $e_0 = f$
for $j = 1, 2, \ldots, n$ **do**
  Determine a local approximant $s_j \in W_j$ to e_{j-1}.
  Set $f_j = f_{j-1} + s_j$.
  Set $e_j = e_{j-1} - s_j$.

The algorithm does exactly what we explained in the introduction. It starts with a coarse approximation, forms the residual, approximates the residual on a finer scale, forms the residual and so on.

Note that, by induction, Algorithm 1 introduces local approximations s_j, current approximations f_j and residuals e_j, which are connected via

$$f_j = s_1 + \cdots + s_j, \tag{1}$$
$$e_j = f - (s_1 + \cdots + s_j). \tag{2}$$

It is interesting to see what this algorithm means in matrix form. Since we assume that each W_j is finite dimensional, we can represent the approximant $s_j \in W_j$ as a linear combination of the basis functions in W_j. To explain this in more detail, let us write $X_j = \{\mathbf{x}_1^{(j)}, \ldots, \mathbf{x}_{N_j}^{(j)}\}$. Then, we have a representation of the form

$$s_j = \sum_{k=1}^{N_j} \alpha_k^{(j)} \Phi_j(\cdot, \mathbf{x}_k^{(j)}).$$

Let us assume that the process to determine the local approximant is a linear one. Then, there exists a matrix $A_j \in \mathbb{R}^{N_j \times N_j}$ such that

$$\boldsymbol{\alpha}^{(j)} = A_j^{-1}(e_{j-1}|X_j). \tag{3}$$

Furthermore, if we introduce the evaluation matrices

$$B_{jk} := (\Phi_k(\mathbf{x}_i^{(j)}, \mathbf{x}_\ell^{(k)})) \in \mathbb{R}^{N_j \times N_k},$$

the representation (2) restricted to X_j gives

$$A_j \boldsymbol{\alpha}^{(j)} = \mathbf{f}^{(j)} - \sum_{k=1}^{j-1} B_{jk} \boldsymbol{\alpha}^{(k)},$$

with $\mathbf{f}^{(j)} = f|X_j \in \mathbb{R}^{N_j}$. Hence, if we collect all these equations in one linear system, we obtain a triangular system of the form

$$\begin{pmatrix} A_1 & & & & \\ B_{21} & A_2 & & & \\ B_{31} & B_{32} & A_3 & & \\ \vdots & \vdots & \dots & \ddots & \\ B_{n1} & B_{n2} & \cdots & B_{n,n-1} & A_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}^{(1)} \\ \boldsymbol{\alpha}^{(2)} \\ \boldsymbol{\alpha}^{(3)} \\ \vdots \\ \boldsymbol{\alpha}^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(2)} \\ \mathbf{f}^{(3)} \\ \vdots \\ \mathbf{f}^{(n)} \end{pmatrix},$$

and Algorithm 1 simply solves this system by mimicking the standard forward substitution for classical lower triangular systems. However, since the diagonal elements are now matrices, in each step we have to solve a linear system with a growing number of unknowns. We will see that this can be done in linear time since the matrices A_j are positive definite, sparse, having roughly the same number of non-zero entries per row, and their condition number is independent of the level.

In this paper, we will investigate two different types of approximation processes to determine the local approximations s_j, interpolation and a penalized least-squares approach sometimes referred to as smoothing splines or smoothed approximation.

3 Function spaces

It is now time to specify the analytic setting in which our approximation scheme will work.

We will assume that our functions are defined on a bounded domain Ω with a Lipschitz boundary. This gives rise to an extension operator for functions defined in Sobolev spaces. The following result comes from [13, Theorem 5, Section 3.1] for integer τ and extends to real $\tau \geq 0$ by standard interpolation in Sobolev spaces, see also [1].

Proposition 1 *Suppose $\Omega \subseteq \mathbb{R}^d$ is open and has a Lipschitz boundary. Let $\tau \geq 0$. Then, there exists a linear operator $E : H^\tau(\Omega) \to H^\tau(\mathbb{R}^d)$, such that, for all $f \in H^\tau(\Omega)$,*

1. $Ef|\Omega = f|\Omega$,
2. $\|Ef\|_{H^\tau(\mathbb{R}^d)} \leq C_\tau \|f\|_{H^\tau(\Omega)}$,

i.e., E is a bounded extension operator. Furthermore, the same operator E can be used for every $\tau \geq 0$.

It is important to see that only the constant may depend on τ but not the extension operator, meaning that for a fixed $\Omega \subseteq \mathbb{R}^d$ the same extension operator can be used for all Sobolev spaces $H^\tau(\Omega)$, $\tau \geq 0$.

This allows us to identify a function $f \in H^\tau(\Omega)$ with its extension $Ef \in H^\tau(\mathbb{R}^d)$ and, since we obviously have,

$$\|f\|_{H^\tau(\Omega)} \leq \|Ef\|_{H^\tau(\mathbb{R}^d)} \leq C_\tau \|f\|_{H^\tau(\Omega)},$$

we see that $f \mapsto \|Ef\|_{H^\tau(\mathbb{R}^d)}$ effectively defines a norm on $H^\tau(\Omega)$, which is equivalent to any other standard norm on $H^\tau(\Omega)$. This interplay between f and its extension Ef will be of quite some importance in our analysis of the multiscale approximation scheme.

In particular, it allows us to restrict quite a large part of our analysis to $H^\tau(\mathbb{R}^d)$, where, for $g \in H^\tau(\mathbb{R}^d)$, we have the alternative norm

$$\|g\|^2_{H^\tau(\mathbb{R}^d)} = \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 (1 + \|\boldsymbol{\omega}\|_2^2)^\tau d\boldsymbol{\omega}$$

using the Fourier transform

$$\widehat{g}(\boldsymbol{\omega}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} g(\mathbf{x}) e^{-i\mathbf{x}^T\boldsymbol{\omega}} d\mathbf{x}.$$

This alternative way of characterizing functions from $H^\tau(\mathbb{R}^d)$ gives us the necessary leeway to work with compactly supported radial basis functions. There exist such basis functions (see [12, 15]), which have a Fourier transform satisfying

$$c_1(1 + \|\boldsymbol{\omega}\|_2^2)^{-\tau} \leq \widehat{\Phi}(\boldsymbol{\omega}) \leq c_2(1 + \|\boldsymbol{\omega}\|_2^2)^{-\tau}, \quad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{4}$$

with two fixed constants $0 < c_1 \leq c_2$.

This means that the associated reproducing kernel Hilbert space (or native space) $\mathcal{N}_\Phi(\mathbb{R}^d)$ of Φ which consists of all functions $g \in L_2(\mathbb{R}^d)$ with

$$\|g\|_\Phi^2 = \int_{\mathbb{R}^d} \frac{|\widehat{g}(\boldsymbol{\omega})|^2}{\widehat{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega} < \infty$$

is norm-equivalent to the Sobolev space $H^\tau(\mathbb{R}^d)$. We will scale this basis function in the following way. Let Φ_δ be defined by

$$\Phi_\delta(\mathbf{x}) := \delta^{-d} \Phi(\mathbf{x}/\delta), \tag{5}$$

such that it has a Fourier transform $\widehat{\Phi_\delta}(\boldsymbol{\omega}) = \widehat{\Phi}(\delta\boldsymbol{\omega})$ satisfying

$$c_1 \left(1 + \delta^2 \|\boldsymbol{\omega}\|_2^2\right)^{-\tau} \leq \widehat{\Phi_\delta}(\boldsymbol{\omega}) \leq c_2 \left(1 + \delta^2 \|\boldsymbol{\omega}\|_2^2\right)^{-\tau}, \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{6}$$

Using a scaled basis function yields the following norm equivalence.

Lemma 1 *For every $\delta \in (0, 1]$ we have $\mathcal{N}_{\Phi_\delta}(\mathbb{R}^d) = H^\tau(\mathbb{R}^d)$ and for every $g \in H^\tau(\mathbb{R}^d)$, we have the norm equivalence*

$$c_1^{1/2} \|g\|_{\Phi_\delta} \leq \|g\|_{H^\tau(\mathbb{R}^d)} \leq c_2^{1/2} \delta^{-\tau} \|g\|_{\Phi_\delta},$$

Proof For $\delta \leq 1$ we have

$$\left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau = \delta^{-2\tau} \left(\delta^2 + \delta^2 \|\boldsymbol{\omega}\|_2^2\right)^\tau \leq \delta^{-2\tau} \left(1 + \delta^2 \|\boldsymbol{\omega}\|_2^2\right)^\tau.$$

Hence, the definition of the Sobolev and native space norms and (6) yield for $\delta \leq 1$ that

$$\begin{aligned}
\|g\|_{H^\tau(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} \\
&\leq \delta^{-2\tau} \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1 + \delta^2 \|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} \\
&\leq c_2 \delta^{-2\tau} \int_{\mathbb{R}^d} \frac{|\widehat{g}(\boldsymbol{\omega})|^2}{\widehat{\Phi_\delta}(\boldsymbol{\omega})} d\boldsymbol{\omega} \\
&= c_2 \delta^{-2\tau} \|g\|_{\Phi_\delta}^2.
\end{aligned}$$

This gives the upper bound. The lower bound simply follows from

$$\begin{aligned}
\|g\|_{H^\tau(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} \\
&\geq \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1 + \delta^2 \|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} \\
&\geq c_1 \|g\|_{\Phi_\delta}^2,
\end{aligned}$$

which obviously holds for $\delta \in (0, 1]$. □

Let us remark that the condition $\delta \le 1$ can be relaxed to $\delta \le \delta_0$ for any $\delta_0 > 0$ at the price of having constants depending on δ_0.

4 Multiscale interpolation of smooth functions

We will now start our analysis of the multiscale approximation algorithm. In this section, we will assume that the local approximants s_j are interpolants using the scaled kernel

$$\Phi_j(\mathbf{x}, \mathbf{y}) = \delta_j^{-d} \Phi((\mathbf{x} - \mathbf{y})/\delta_j), \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \tag{7}$$

where Φ satisfies (4). This means, we can write

$$s_j(\mathbf{x}) = I_{X_j,\delta_j} e_{j-1}(\mathbf{x}) = \sum_{k=1}^{N_j} \alpha_k^{(j)} \Phi_j(\mathbf{x}, \mathbf{x}_k^{(j)}), \quad \mathbf{x} \in \mathbb{R}^d,$$

introducing the interpolation operator I_{X_j,δ_j}, which maps a continuous function, known at the data sites X_j to its interpolant based on the data set X_j and the scaled basis function Φ_j.

The coefficient vector $\boldsymbol{\alpha}^{(j)}$ is determined by solving the linear system (3) with

$$A_j = (\Phi_j(\mathbf{x}_i^{(j)}, \mathbf{x}_k^{(j)}))_{1 \le i,k \le N_j}.$$

Though we only need f to be continuous to make interpolation possible, we will here, in contrast to the next section, assume that $f \in H^\tau(\Omega)$, i.e., $Ef \in \mathcal{N}_{\Phi_j}(\mathbb{R}^d) = H^\tau(\mathbb{R}^d)$.

It is important to note that, since we assume $X_j \subseteq \Omega$, we also have that the interpolant to $e_{j-1} \in H^\tau(\Omega)$ is identical to the interpolant to $Ee_{j-1} \in H^\tau(\mathbb{R}^d)$, meaning

$$I_{X_j,\delta_j} e_{j-1} = I_{X_j,\delta_j} E e_{j-1}.$$

Furthermore, optimality of the interpolation process guarantees

$$\|Ee_{j-1} - I_{X_j,\delta_j} Ee_{j-1}\|_{\Phi_j} \le \|Ee_{j-1}\|_{\Phi_j}.$$

Finally, we need the following result from [8].

Lemma 2 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let $\tau > d/2$. Let $X \subseteq \Omega$ be a finite point set with sufficiently small mesh norm $h_{X,\Omega}$. Then, there is a constant $C > 0$, independent of X, such that for all $f \in H^\tau(\Omega)$ vanishing on X, we have*

$$\|f\|_{H^\mu(\Omega)} \le C h_{X,\Omega}^{\tau-\mu} \|f\|_{H^\tau(\Omega)}.$$

for $0 \le \mu \le \tau$.

The next theorem is our main convergence result for target functions from the associated reproducing kernel Hilbert space.

Theorem 1 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let X_1, $X_2, \ldots$ be a sequence of point sets in Ω with mesh norms $h_1, h_2, \ldots$ satisfying $c\mu h_j \le h_{j+1} \le \mu h_j$ for $j = 1, 2, \ldots$ with fixed $\mu \in (0, 1)$, $c \in (0, 1]$ and h_1 sufficiently small. Let Φ be a kernel generating $H^\tau(\mathbb{R}^d)$, i.e., satisfying (4) and let Φ_j be defined by (7) with scale factor $\delta_j = \nu h_j$. Assume $1/h_1 \ge \nu \ge \gamma/\mu$ with a fixed $\gamma > 0$. Let the target function f belong to $H^\tau(\Omega)$. Then, there exists a constant $C_1 = C_1(\gamma) > 0$ such that, with $\alpha = C_1\mu^\tau$,*

$$\|Ee_j\|_{\Phi_{j+1}} \le \alpha \|Ee_{j-1}\|_{\Phi_j} \quad \text{for } j = 1, 2, 3, \ldots$$

and hence there exists a constant $C > 0$ such that

$$\|f - f_n\|_{L_2(\Omega)} \le C\alpha^n \|f\|_{H^\tau(\Omega)} \quad \text{for } n = 1, 2, \ldots$$

Thus the multiscale approximation f_n converges linearly to f in the L_2 norm if $\alpha = C_1\mu^\tau < 1$.

Proof First of all note that $\delta_{j+1} = \nu h_{j+1} \le \nu\mu h_j = \mu\delta_j < \delta_j$ and hence $\delta_{j+1} < \delta_1 = \nu h_1 \le 1$ such that we can apply Lemma 1 in the following. Using the extension operator : $E : H^\tau(\Omega) \to H^\tau(\mathbb{R}^d)$ we have

$$\|Ee_j\|^2_{\Phi_{j+1}} \le \frac{1}{c_1} \int\limits_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \delta^2_{j+1}\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} =: \frac{1}{c_1}(I_1 + I_2)$$

with

$$I_1 := \int\limits_{\|\boldsymbol{\omega}\|_2 \le \frac{1}{\delta_{j+1}}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \delta^2_{j+1}\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega},$$

$$I_2 := \int\limits_{\|\boldsymbol{\omega}\|_2 \ge \frac{1}{\delta_{j+1}}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \delta^2_{j+1}\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega}.$$

We are now going to estimate both integrals separately. To this end, it is helpful to observe that the interpolant at X_j to e_{j-1} is the same as the interpolant to Ee_{j-1} since both functions coincide on $X_j \subseteq \Omega$. From this, it follows that

$$\begin{aligned}
\|e_j\|_{H^\tau(\Omega)} &= \|e_{j-1} - I_{X_j,\delta_j} e_{j-1}\|_{H^\tau(\Omega)} \\
&= \|Ee_{j-1} - I_{X_j,\delta_j} Ee_{j-1}\|_{H^\tau(\Omega)} \\
&\le \|Ee_{j-1} - I_{X_j,\delta_j} Ee_{j-1}\|_{H^\tau(\mathbb{R}^d)} \\
&\le C\delta_j^{-\tau} \|Ee_{j-1} - I_{X_j,\delta_j} Ee_{j-1}\|_{\Phi_j} \\
&\le C\delta_j^{-\tau} \|Ee_{j-1}\|_{\Phi_j}, \qquad (8)
\end{aligned}$$

where we have used Lemma 1 and the fact that the interpolant is norm-minimal with respect to the Φ_j-norm. Consequently, the constant appearing here depends only on the equivalence constants given in Lemma 1.

For the first integral we use $\delta_{j+1}\|\boldsymbol{\omega}\|_2 \le 1$ and then Lemma 2, because e_j vanishes on X_j. This yields

$$
\begin{aligned}
I_1 &\le 2^\tau \int\limits_{\|\boldsymbol{\omega}\|_2 \le 1/\delta_{j+1}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \\
&\le 2^\tau \int\limits_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} = 2^\tau \|Ee_j\|^2_{L_2(\mathbb{R}^d)} \\
&\le 2^\tau C_0 \|e_j\|^2_{L_2(\Omega)} \le C h_j^{2\tau} \|e_j\|^2_{H^\tau(\Omega)} \\
&\le C \left(\frac{h_j}{\delta_j}\right)^{2\tau} \|Ee_{j-1}\|^2_{\Phi_j} \\
&\le C\mu^{2\tau} \|Ee_{j-1}\|^2_{\Phi_j},
\end{aligned}
$$

where the last two steps follow from (8) and since $h_j/\delta_j = h_j/(\nu h_j) = 1/\nu \le \mu/\gamma$. Note that the constant appearing in the last line depends on the smoothness τ, the extension constant from Proposition 1, the equivalence constants from Lemma 1, the constant from Lemma 2 and on γ, i.e., it is of the form $C = C(\Phi, \Omega, \tau, \gamma) = C(\Phi, \Omega, \tau)\gamma^{-2\tau}$. The estimate itself is valid if h_j is sufficiently small.

For the second integral, I_2, we observe that $\delta_{j+1}\|\boldsymbol{\omega}\|_2 \ge 1$ implies

$$
\left(1 + \delta_{j+1}^2 \|\boldsymbol{\omega}\|_2^2\right)^\tau \le 2^\tau \delta_{j+1}^{2\tau} \|\boldsymbol{\omega}\|_2^{2\tau} \le 2^\tau \delta_{j+1}^{2\tau} \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau,
$$

such that, by (8),

$$
\begin{aligned}
I_2 &\le 2^\tau \delta_{j+1}^{2\tau} \int\limits_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} = 2^\tau \delta_{j+1}^{2\tau} \|Ee_j\|^2_{H^\tau(\mathbb{R}^d)} \\
&\le C\delta_{j+1}^{2\tau} \|e_j\|^2_{H^\tau(\Omega)} \le C \left(\frac{\delta_{j+1}}{\delta_j}\right)^{2\tau} \|Ee_{j-1}\|^2_{\Phi_j} \\
&\le C\mu^{2\tau} \|Ee_{j-1}\|^2_{\Phi_j}.
\end{aligned}
$$

This time the constant depends only on τ, the extension constant from Proposition 1 and the equivalence constant from Lemma 1, i.e., it is of the form $C = C(\Phi, \Omega, \tau)$.

Putting things together, we see that we have

$$
\|Ee_j\|_{\Phi_{j+1}} \le C_1 \mu^\tau \|Ee_{j-1}\|_{\Phi_j} \le \alpha \|Ee_{j-1}\|_{\Phi_j},
$$

provided $C_1\mu^\tau \le \alpha$, which is the first statement of our theorem. Note that the final constant takes the form $C_1 = C_1(\Phi, \Omega, \tau, \gamma) = C_1(\Phi, \Omega, \tau)(1 + \gamma^{-2\tau})$. Finally, since $e_n = f - f_n$ vanishes on X_n, we have by Lemma 2 and Lemma 1 that

$$\begin{aligned}\|f - f_n\|_{L_2(\Omega)} = \|e_n\|_{L_2(\Omega)} &\le Ch_n^\tau \|e_n\|_{H^\tau(\Omega)} \\ &\le Ch_n^\tau \|Ee_n\|_{H^\tau(\mathbb{R}^d)} \\ &\le Ch_n^\tau \delta_{n+1}^{-\tau} \|Ee_n\|_{\Phi_{n+1}} \\ &= C\|Ee_n\|_{\Phi_{n+1}},\end{aligned}$$

since $h_n/\delta_{n+1} = h_n/(\nu h_{n+1}) \le 1/(c\gamma)$. Now we can apply the first step n times to derive

$$\|f - f_n\|_{L_2(\Omega)} \le C\alpha^n \|Ef\|_{\Phi_1} \le C\alpha^n \|Ef\|_{H^\tau(\mathbb{R}^d)} \le C\alpha^n \|f\|_{H^\tau(\Omega)},$$

which is the stated bound. □

It is now a good time to have another look at the context of multiresolution analysis. As pointed out in Sect. 2, we define spaces

$$\begin{aligned} W_j &= \operatorname{span}\{\Phi_j(\cdot, \mathbf{x}) : \mathbf{x} \in X_j\}, \\ V_j &= W_1 + \cdots + W_j, \end{aligned}$$

where the W_j take over the role of the detail spaces of a classical multiresolution analysis. Furthermore, we have the following properties, which can be proven exactly in the same way, as it has been done in [6].

Proposition 2 *Let $V_0 = \{0\}$. Then, for every $j \ge 0$, V_j is a closed and finite dimensional subspace of $L_2(\Omega)$, actually even of $H^\tau(\Omega)$, and*

1. $V_j \subseteq V_{j+1}$,
2. $\overline{\cup_{j=0}^\infty V_j} = L_2(\Omega)$, *where the closure is with respect to the $L_2(\Omega)$ norm.*

Furthermore, the following sum is direct for every $j \ge 1$,

$$V_j = V_{j-1} \oplus W_j,$$

which also means that

$$\bigoplus_{j=1}^\infty W_j = L_2(\Omega).$$

Obviously, because of the fact that we are working with scattered data sites, we cannot expect a property of the form $f \in V_j$ if and only if $f(\cdot/\delta_j) \in V_1$. This will also not be true for regular data sites and scaling, since compactly supported radial basis functions usually do not satisfy any kind of refinement equation.

However, a refinement equation is only important for a specific algorithm to compute best L_2 approximations from V_j, which is not our goal. Instead, we are using Algorithm 1.

5 Multiscale interpolation of rough functions

It is now our goal to extend the convergence results of Theorem 1 to the situation of rougher target functions. To be more precise, throughout this section we will assume, as in the last section, that our interpolants are formed using a reproducing kernel Φ of $H^\tau(\mathbb{R}^d)$ but now the target function f belongs only to $H^\beta(\mathbb{R}^d)$ with $\tau \geq \beta > d/2$. The assumption $\beta > d/2$ is still necessary since we are talking about interpolation and hence need continuous point evaluations. We will denote the reproducing kernel of $H^\beta(\mathbb{R}^d)$ by $\Psi : \mathbb{R}^d \to \mathbb{R}$, which means that we have

$$\widehat{\Psi}(\boldsymbol{\omega}) = (1 + \|\boldsymbol{\omega}\|_2^2)^{-\beta}, \quad \boldsymbol{\omega} \in \mathbb{R}^d.$$

We will scale this kernel exactly as we did it with Φ, i.e., we define $\Psi_\delta(\mathbf{x}) = \delta^{-d}\Psi(\mathbf{x}/\delta)$ and

$$\Psi_j(\mathbf{x}, \mathbf{y}) = \delta_j^{-d}\Psi((\mathbf{x} - \mathbf{y})/\delta_j).$$

In deriving our convergence result, we will use the following essential result from [8], which guarantees the existence of an interpolating and almost best approximating band-limited function. Let $B(\mathbf{0}, \sigma) \subseteq \mathbb{R}^d$ denote the ball with radius σ and center $\mathbf{0}$ in $\mathbb{R}^d$. Let

$$\mathcal{B}_\sigma = \{f \in L_2(\mathbb{R}^d) : \operatorname{supp}(\widehat{f}) \subseteq B(\mathbf{0}, \sigma)\}$$

and

$$\operatorname{dist}_{H^\beta(\mathbb{R}^d)}(f, \mathcal{B}_\sigma) = \inf_{g \in \mathcal{B}_\sigma} \|f - g\|_{H^\beta(\mathbb{R}^d)}.$$

Then, the following result holds.

Lemma 3 *Let $X \subseteq \mathbb{R}^d$ be a finite point set. Let $\beta, t \in \mathbb{R}$, $\beta > d/2$, $t \geq 0$. If $f \in H^{\beta+t}(\mathbb{R}^d)$ then there exists an $f_\sigma \in \mathcal{B}_\sigma$ such that $f_\sigma|X = f|X$ and*

$$\|f - f_\sigma\|_{H^\beta(\mathbb{R}^d)} \leq 5 \cdot \operatorname{dist}_{H^\beta(\mathbb{R}^d)}(f, \mathcal{B}_\sigma) \leq 5\kappa^{-t} q_X^t \|f\|_{H^{\beta+t}(\mathbb{R}^d)},$$

with $\sigma = \kappa/q_X$, where $\kappa \geq 1$ depends only on d and β.

A consequence of this result is, for example, that we have

$$\|f_\sigma\|_{H^\beta(\mathbb{R}^d)} \leq 6\|f\|_{H^\beta(\mathbb{R}^d)}.$$

We will now derive a series of auxiliary results. We start with a version of Lemma 3 for scaled Sobolev spaces.

Lemma 4 *Let $f \in H^\beta(\mathbb{R}^d)$, $\beta > d/2$ and let $X \subseteq \mathbb{R}^d$ be a finite subset with separation distance q_X. Let $\delta \in (0, 1]$ be given and $\sigma = \kappa\delta/q_X$ with $\kappa \geq 1$ from Lemma 3. Then, there exists a function $f_{\sigma/\delta} \in \mathcal{B}_{\sigma/\delta}$ with $f_{\sigma/\delta}|X = f|X$ and*

$$\|f - f_{\sigma/\delta}\|_{\Psi_\delta} \leq 5\|f\|_{\Psi_\delta}.$$

Proof Let us define $g \in H^\beta(\mathbb{R}^d)$ by $g(\mathbf{x}) = \delta^{d/2} f(\delta\mathbf{x})$. Then we have $\widehat{g}(\boldsymbol{\omega}) = \delta^{-d/2}\widehat{f}(\boldsymbol{\omega}/\delta)$ and hence

$$\begin{aligned}
\|g\|_{H^\beta(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d} |\widehat{g}(\boldsymbol{\omega})|^2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\beta d\boldsymbol{\omega} \\
&= \delta^{-d} \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{\omega}/\delta)|^2 \left(1 + \|\boldsymbol{\omega}\|_2^2\right)^\beta d\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{\omega})|^2 \left(1 + \delta^2\|\boldsymbol{\omega}\|_2^2\right)^\beta d\boldsymbol{\omega} \\
&= \|f\|_{\Psi_\delta}^2,
\end{aligned}$$

i.e., $\|g\|_{H^\beta(\mathbb{R}^d)} = \|f\|_{\Psi_\delta}$. Let us define the set $Y = X/\delta = \{\mathbf{x}/\delta : \mathbf{x} \in X\}$, which obviously has separation distance $q_Y = q_X/\delta$. Lemma 3 guarantees the existence of a $g_\sigma \in \mathcal{B}_\sigma$ with $\sigma = \kappa/q_Y = \delta\kappa/q_X$ and $g_\sigma|Y = g|Y$ and $\|g - g_\sigma\|_{H^\beta(\mathbb{R}^d)} \leq 5\|g\|_{H^\beta(\mathbb{R}^d)}$. We can now define $f_{\sigma/\delta}(\mathbf{x}) := \delta^{-d/2} g_\sigma(\mathbf{x}/\delta)$, which has Fourier transform $\widehat{f_{\sigma/\delta}}(\boldsymbol{\omega}) = \delta^{d/2}\widehat{g_\sigma}(\delta\boldsymbol{\omega})$. Since $g_\sigma \in \mathcal{B}_\sigma$ this means in particular $f_{\sigma/\delta} \in \mathcal{B}_{\sigma/\delta}$. Furthermore, we have, for $\mathbf{x}_j \in X$,

$$\begin{aligned}
f_{\sigma/\delta}(\mathbf{x}_j) &= \delta^{-d/2} g_\sigma(\mathbf{x}_j/\delta) = \delta^{-d/2} g_\sigma(\mathbf{y}_j) = \delta^{-d/2} g(\mathbf{y}_j) = f(\delta\mathbf{y}_j) \\
&= f(\mathbf{x}_j),
\end{aligned}$$

i.e., $f_{\sigma/\delta}|X = f|X$. Finally, since $\|f_{\sigma/\delta}\|_{\Psi_\delta} = \|g_\sigma\|_{H^\beta(\mathbb{R}^d)}$ and $\|g - g_\sigma\|_{H^\beta(\mathbb{R}^d)} = \|f - f_{\sigma/\delta}\|_{\Psi_\delta}$, Lemma 3 yields also the stated bound. □

The next lemma is a sort of inverse inequality for our scaled spaces.

Lemma 5 *Let $f \in H^\beta(\mathbb{R}^d)$ and $\delta \in (0, 1]$. Let $X \subseteq \mathbb{R}^d$ be a finite set with $q_X \leq \delta$. Choose κ, $\sigma = \kappa\delta/q_X$ and $f_{\sigma/\delta} \in \mathcal{B}_{\sigma/\delta}$ according to Lemma 4. Then,*

$$\|f_{\sigma/\delta}\|_{\Phi_\delta} \leq C\sigma^{\tau-\beta}\|f\|_{\Psi_\delta}.$$

Proof Lemma 4 and $\sigma = \delta\kappa/q_X \geq 1$ give

$$\begin{aligned}
\|f_{\sigma/\delta}\|_{\Phi_\delta}^2 &= \int\limits_{\|\boldsymbol{\omega}\|_2 \leq \sigma/\delta} |\widehat{f_{\sigma/\delta}}(\boldsymbol{\omega})|^2 \left(1+\delta^2\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} \\
&= \int\limits_{\|\boldsymbol{\omega}\|_2 \leq \sigma/\delta} |\widehat{f_{\sigma/\delta}}(\boldsymbol{\omega})|^2 (1+\delta^2\|\boldsymbol{\omega}\|_2^2)^\beta (1+\delta^2\|\boldsymbol{\omega}\|_2^2)^{\tau-\beta} d\boldsymbol{\omega} \\
&\leq (1+\sigma^2)^{\tau-\beta}\|f_{\sigma/\delta}\|_{\Psi_\delta}^2 \\
&\leq 36(1+\sigma^2)^{\tau-\beta}\|f\|_{\Psi_\delta}^2 \\
&\leq 36\cdot 2^{\tau-\beta}\sigma^{2\tau-2\beta}\|f\|_{\Psi_\delta}^2,
\end{aligned}$$

as required. □

The next lemma mimics estimate (8) in the situation of rougher target functions. We need now, however, that our data sets are quasi-uniform.

Lemma 6 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary. Let $X_j \subseteq \Omega$ with $q_j \leq h_j \leq c_q q_j$ and $q_j \leq \delta_j$. Then,*

$$\|e_j\|_{H^\beta(\Omega)} \leq C\delta_j^{-\beta}\|Ee_{j-1}\|_{\Psi_j}.$$

Proof The idea is to use a band-limited interpolating approximant for Ee_{j-1}. Hence, we start by splitting the error in the form

$$\begin{aligned}
\|e_j\|_{H^\beta(\Omega)} &= \|e_{j-1} - I_{X_j,\delta_j}e_{j-1}\|_{H^\beta(\Omega)} \\
&= \|Ee_{j-1} - I_{X_j,\delta_j}Ee_{j-1}\|_{H^\beta(\Omega)} \\
&\leq \|Ee_{j-1} - (Ee_{j-1})_{\sigma_j/\delta_j}\|_{H^\beta(\Omega)} \\
&\quad + \|(Ee_{j-1})_{\sigma_j/\delta_j} - I_{X_j,\delta_j}(Ee_{j-1})_{\sigma_j/\delta_j}\|_{H^\beta(\Omega)},
\end{aligned}$$

where we chose $\sigma_j = \kappa\delta_j/q_j$ according to Lemma 4. The first term can be bounded via

$$\begin{aligned}
\|Ee_{j-1} - (Ee_{j-1})_{\sigma_j/\delta_j}\|_{H^\beta(\Omega)} &\leq \|Ee_{j-1} - (Ee_{j-1})_{\sigma_j/\delta_j}\|_{H^\beta(\mathbb{R}^d)} \\
&\leq \delta_j^{-\beta}\|Ee_{j-1} - (Ee_{j-1})_{\sigma_j/\delta_j}\|_{\Psi_j} \\
&\leq 5\delta_j^{-\beta}\|Ee_{j-1}\|_{\Psi_j},
\end{aligned}$$

using Lemma 1 and Lemma 4. For the second term we use Lemma 2, Lemma 1, the optimality of the interpolant in the $\|\cdot\|_{\Phi_j}$ norm, Lemma 5, and the definition of σ_j, respectively, to conclude that

$$\begin{aligned}
&\|(Ee_{j-1})_{\sigma_j/\delta_j} - I_{X_j,\delta_j}(Ee_{j-1})_{\sigma_j/\delta_j}\|_{H^\beta(\Omega)} \\
&\quad\le Ch_j^{\tau-\beta}\|(Ee_{j-1})_{\sigma_j/\delta_j} - I_{X_j,\delta_j}(Ee_{j-1})_{\sigma_j/\delta_j}\|_{H^\tau(\Omega)} \\
&\quad\le Ch_j^{\tau-\beta}\delta_j^{-\tau}\|(Ee_{j-1})_{\sigma_j/\delta_j} - I_{X_j,\delta_j}(Ee_{j-1})_{\sigma_j/\delta_j}\|_{\Phi_j} \\
&\quad\le Ch_j^{\tau-\beta}\delta_j^{-\tau}\|(Ee_{j-1})_{\sigma_j/\delta_j}\|_{\Phi_j} \\
&\quad\le Ch_j^{\tau-\beta}\delta_j^{-\tau}\sigma_j^{\tau-\beta}\|Ee_{j-1}\|_{\Psi_j} \\
&\quad= C\delta_j^{-\beta}(\kappa h_j/q_j)^{\tau-\beta}\|Ee_{j-1}\|_{\Psi_j}.
\end{aligned}$$

Taking both bounds together finally gives

$$\|e_j\|_{H^\beta(\Omega)} \le C\delta_j^{-\beta}\left[1 + \left(\kappa\frac{h_j}{q_j}\right)^{\tau-\beta}\right]\|Ee_{j-1}\|_{\Psi_j},$$

which is the desired result since $h_j \le c_q q_j$. □

Now we are in a position to prove that the multilevel algorithm also converges for target functions from $H^\beta(\Omega)$.

Theorem 2 *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let $X_1, X_2, \ldots$ be a sequence of point sets in Ω with mesh norms $h_1, h_2, \ldots$ and separation distances $q_1, q_2, \ldots$ satisfying*

1. *$c\mu h_j \le h_{j+1} \le \mu h_j$ for $j = 1, 2, \ldots$ with $\mu \in (0, 1)$, $c \in (0, 1]$,*
2. *$q_j \le h_j \le c_q q_j$ for $j = 1, 2, \ldots$ with $c_q > 0$.*

Let Φ be a kernel generating $H^\tau(\mathbb{R}^d)$, i.e., satisfying (4) and let Φ_j be defined by (7) with scale factor $\delta_j = \nu h_j$ with $1/h_1 \ge \nu \ge \gamma/\mu \ge 1$ with a fixed $\gamma > 0$. Let Ψ be the kernel generating $H^\beta(\mathbb{R}^d)$ with $\tau \ge \beta > d/2$ and let Ψ_j be the scaled version using the same scale factor δ_j. Assume that the target function f belongs to $H^\beta(\Omega)$. Then, there exists a constant $C_1 = C_1(\gamma) > 0$ such that, with $\alpha = C_1\mu^\beta$,

$$\|Ee_j\|_{\Psi_{j+1}} \le \alpha\|Ee_{j-1}\|_{\Psi_j} \quad \text{for } j = 1, 2, 3, \ldots$$

and hence there exists a constant $C > 0$ such that

$$\|f - f_n\|_{L_2(\Omega)} \le C\alpha^n\|f\|_{H^\beta(\Omega)} \quad \text{for } n = 1, 2, \ldots$$

Thus the multiscale approximation f_n converges linearly to f in the L_2 norm if $\alpha = C_1\mu^\beta < 1$.

Proof The proof is now very much the same as the proof of Theorem 1. As in the proof of Theorem 1 we can conclude that $\delta_j \le 1$ for all j. Moreover, we have $\delta_j \ge h_j \ge q_j$.

Thus, we can use Lemma 6. We start again with

$$\|Ee_j\|_{\Psi_{j+1}}^2 = \int_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 (1+\delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2)^\beta d\boldsymbol{\omega} =: I_1 + I_2$$

with

$$I_1 := \int_{\|\boldsymbol{\omega}\|_2 \le \frac{1}{\delta_{j+1}}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1+\delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2\right)^\beta d\boldsymbol{\omega},$$

$$I_2 := \int_{\|\boldsymbol{\omega}\|_2 \ge \frac{1}{\delta_{j+1}}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1+\delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2\right)^\beta d\boldsymbol{\omega}.$$

For the first integral we use $\delta_{j+1}\|\boldsymbol{\omega}\|_2 \le 1$ and then Lemma 2, because e_j vanishes on X_j, and finally Lemma 6, to derive

$$\begin{aligned}
I_1 &\le 2^\beta \int_{\|\boldsymbol{\omega}\|_2 \le 1/\delta_{j+1}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \\
&\le 2^\beta \int_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} = 2^\beta \|Ee_j\|_{L_2(\mathbb{R}^d)}^2 \\
&\le 2^\beta C_0 \|e_j\|_{L_2(\Omega)}^2 \le C h_j^{2\beta} \|e_j\|_{H^\beta(\Omega)}^2 \\
&\le C \left(\frac{h_j}{\delta_j}\right)^{2\beta} \|Ee_{j-1}\|_{\Psi_j}^2 \\
&\le C\mu^{2\beta} \|Ee_{j-1}\|_{\Psi_j}^2.
\end{aligned}$$

For the second integral, I_2, we observe that $\delta_{j+1}\|\boldsymbol{\omega}\|_2 \ge 1$ implies

$$\left(1+\delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2\right)^\beta \le 2^\beta \delta_{j+1}^{2\beta} \|\boldsymbol{\omega}\|_2^{2\beta} \le 2^\beta \delta_{j+1}^{2\beta} \left(1+\|\boldsymbol{\omega}\|_2^2\right)^\beta,$$

so that, again with Lemma 6,

$$\begin{aligned}
I_2 &\le 2^\beta \delta_{j+1}^{2\beta} \int_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1+\|\boldsymbol{\omega}\|_2^2\right)^\beta d\boldsymbol{\omega} = 2^\beta \delta_{j+1}^{2\beta} \|Ee_j\|_{H^\beta(\mathbb{R}^d)}^2 \\
&\le C\delta_{j+1}^{2\beta} \|e_j\|_{H^\beta(\Omega)}^2 \le C \left(\frac{\delta_{j+1}}{\delta_j}\right)^{2\beta} \|Ee_{j-1}\|_{\Psi_j}^2 \\
&\le C\mu^{2\beta} \|Ee_{j-1}\|_{\Psi_j}^2
\end{aligned}$$

Putting things together, we see that we have

$$\|Ee_j\|_{\Psi_{j+1}} \le C_1\mu^\beta \|Ee_{j-1}\|_{\Psi_j} \le \alpha\|Ee_{j-1}\|_{\Psi_j},$$

provided that $C_1\mu^\beta \le \alpha$, which is the first statement of our theorem. Finally, since $e_n = f - f_n$ vanishes on X_n, we have by Lemmas 2 and 1 that

$$\begin{aligned}\|f - f_n\|_{L_2(\Omega)} = \|e_n\|_{L_2(\Omega)} &\le Ch_n^\beta \|e_n\|_{H^\beta(\Omega)}\\ &\le Ch_n^\beta \|Ee_n\|_{H^\beta(\mathbb{R}^d)}\\ &\le Ch_n^\beta \delta_{n+1}^{-\beta}\|Ee_n\|_{\Psi_{n+1}}\\ &= C\|Ee_n\|_{\Psi_{n+1}},\end{aligned}$$

since $h_n/\delta_{n+1} = h_n/(\nu h_{n+1}) \le 1/(c\gamma)$. Now we can apply the first step n times to derive

$$\|f - f_n\|_{L_2(\Omega)} \le C\alpha^n\|Ef\|_{\Psi_1} \le C\alpha^n\|Ef\|_{H^\beta(\mathbb{R}^d)} \le C\alpha^n\|f\|_{H^\beta(\Omega)},$$

as stated. □

6 Multiscale approximation with smoothing

In the case of non-exact data, interpolation usually is not the best choice. Instead, one often tries to solve a penalized least-squares problem of the form

$$\min\left\{\sum_{j=1}^N |s(\mathbf{x}_j) - f(\mathbf{x}_j)|^2 + \epsilon\|s\|^2_{H^\tau(\mathbb{R}^d)} : s \in H^\tau(\mathbb{R}^d)\right\}$$

for given data $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and $f|X = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^T$ and with a smoothing parameter $\epsilon > 0$. It is well known that the solution to this problem also has the form

$$s_\epsilon = \sum_{j=1}^N \alpha_j \Phi(\cdot, \mathbf{x}_j),$$

with the reproducing kernel Φ of $H^\tau(\mathbb{R}^d)$. This time, however, the coefficients $\boldsymbol{\alpha}$ are determined by the linear system

$$(A + \epsilon I)\boldsymbol{\alpha} = f|X$$

with $A = (\Phi(\mathbf{x}_i, \mathbf{x}_j))$ and I the identity on $\mathbb{R}^N$, see for example [14]. An immediate consequence of this is that, as in the case of interpolation, the solution s_ϵ depends on the target function f only via its values at X, meaning particularly that the smoothed approximation to f is the same as the smoothed approximation to Ef.

Of course, it is possible to employ this reconstruction in the multiscale algorithm. This means that the local reconstruction s_j is given as the solution of

$$\min\left\{\sum_{\mathbf{x}\in X_j}|e_{j-1}(\mathbf{x})-s(\mathbf{x})|^2+\epsilon_j\|s\|^2_{\Phi_j}: s\in H^\tau(\mathbb{R}^d)\right\}. \tag{9}$$

Since $s=Ee_{j-1}$ is a feasible candidate with $s|X_j=e_{j-1}|X_j$, a direct consequence of this is that the new approximation s_j and the new residual $e_j=e_{j-1}-s_j$ satisfy the bounds

$$\|e_j\|_{L_\infty(X_j)}=\|e_{j-1}-s_j\|_{L_\infty(X_j)}\le\sqrt{\epsilon_j}\|Ee_{j-1}\|_{\Phi_j}, \tag{10}$$

$$\|s_j\|_{\Phi_j}\le\|Ee_{j-1}\|_{\Phi_j}. \tag{11}$$

We also need the following *sampling inequality*, which comes from [17] and generalizes Lemma 2.

Lemma 7 *Let $\Omega\subseteq\mathbb{R}^d$ be a bounded domain with a Lipschitz boundary. Let $\tau>d/2$. Then, there is a constant $C>0$ such that for all finite sets $X\subseteq\Omega$ with sufficiently small mesh norm h and all $f\in H^\tau(\Omega)$ we have*

$$\|f\|_{L_2(\Omega)}\le C\left(h^\tau\|f\|_{H^\tau(\Omega)}+\|f\|_{L_\infty(X)}\right).$$

With these ingredients, it is now easy to show that, assuming a certain choice of the smoothing parameters, the multiscale algorithm converges also in this situation.

Theorem 3 *Let $\Omega\subseteq\mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let $X_1,X_2,\dots$ be a sequence of point sets in Ω with mesh norms $h_1,h_2,\dots$ satisfying $c\mu h_j\le h_{j+1}\le\mu h_j$ for $j=1,2,\dots$ with fixed $\mu\in(0,1)$, $c\in(0,1]$ and h_1 sufficiently small. Let Φ be a kernel generating $H^\tau(\mathbb{R}^d)$, i.e., satisfying (4) and let Φ_j be defined by (7) with scale factor $\delta_j=\nu h_j$ with $1/h_1\ge\nu\ge\gamma/\mu$ with a fixed $\gamma>0$. Assume that the target function f belongs to $H^\tau(\Omega)$. Let s_j be the reconstruction from (9) and assume that the smoothing parameters satisfy $\epsilon_j\le\kappa(h_j/\delta_j)^{2\tau}$ with a fixed constant $\kappa>0$. Then, there exists a constant $C_1=C_1(\gamma)>0$ such that, with $\alpha=C_1\mu^\tau$,*

$$\|Ee_j\|_{\Phi_{j+1}}\le\alpha\|Ee_{j-1}\|_{\Phi_j}\quad \textit{for } j=1,2,3,\dots$$

and hence there exists a constant $C>0$ such that

$$\|f-f_n\|_{L_2(\Omega)}\le C\alpha^n\|f\|_{H^\tau(\Omega)}\quad \textit{for } n=1,2,\dots$$

Thus the multiscale approximation f_n converges linearly to f in the L_2 norm if $\alpha=C_1\mu^\tau<1$.

Proof We proceed once again as in the proof of Theorem 1. Since we have the same assumptions on h_j and δ_j we in particular have $\delta_j \leq 1$.

First, we derive the following bound, using the now familiar extension operator E, Lemma 1 and (11),

$$\begin{aligned} \|e_j\|_{H^\tau(\Omega)} &= \|e_{j-1} - s_j\|_{H^\tau(\Omega)} = \|Ee_{j-1} - s_j\|_{H^\tau(\Omega)} \\ &\leq C\delta_j^{-\tau}\|Ee_{j-1} - s_j\|_{\Phi_j} \\ &\leq C\delta_j^{-\tau}\|Ee_{j-1}\|_{\Phi_j}. \end{aligned} \tag{12}$$

As in the proof of Theorem 1, we split

$$\|Ee_j\|_{\Phi_{j+1}}^2 \leq \frac{1}{c_1}\int_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega} =: \frac{1}{c_1}(I_1 + I_2)$$

with

$$\begin{aligned} I_1 &:= \int_{\|\boldsymbol{\omega}\|_2 \leq \frac{1}{\delta_{j+1}}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega}, \\ I_2 &:= \int_{\|\boldsymbol{\omega}\|_2 \geq \frac{1}{\delta_{j+1}}} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \left(1 + \delta_{j+1}^2\|\boldsymbol{\omega}\|_2^2\right)^\tau d\boldsymbol{\omega}. \end{aligned}$$

As in the proof of Theorem 1, we estimate both integrals separately. For I_1, we use Lemma 7, the bounds (10) and (12), Lemma 1, the upper bound on ϵ_j and the general assumption on μ, h_j, δ_j, respectively, to derive

$$\begin{aligned} I_1 &\leq C\|Ee_j\|_{L_2(\mathbb{R}^d)}^2 \leq C\|e_j\|_{L_2(\Omega)}^2 \\ &\leq C\left(h_j^\tau\|e_j\|_{H^\tau(\Omega)} + \|e_j\|_{L_\infty(X_j)}\right)^2 \\ &\leq C\left(\left(\frac{h_j}{\delta_j}\right)^\tau\|Ee_{j-1}\|_{\Phi_j} + \sqrt{\epsilon_j}\|Ee_{j-1}\|_{\Phi_j}\right)^2 \\ &= C\left(\left(\frac{h_j}{\delta_j}\right)^\tau + \sqrt{\epsilon_j}\right)^2\|Ee_{j-1}\|_{\Phi_j}^2 \\ &\leq C\mu^{2\tau}\|Ee_{j-1}\|_{\Phi_j}^2. \end{aligned}$$

For I_2, we follow the previous estimates and then use (12) to derive

$$\begin{aligned} I_2 &\leq C\delta_{j+1}^{2\tau}\|Ee_j\|_{H^\tau(\mathbb{R}^d)}^2 \leq C\delta_{j+1}^{2\tau}\|e_j\|_{H^\tau(\Omega)}^2 \\ &\leq C\left(\frac{\delta_{j+1}}{\delta_j}\right)^{2\tau}\|Ee_{j-1}\|_{\Phi_j}^2 \\ &\leq C\mu^{2\tau}\|Ee_{j-1}\|_{\Phi_j}^2. \end{aligned}$$

Taking both bounds on I_1 and I_2 together yields the first statement of the theorem.

The second part follows as for the upper bound of I_1. We use Lemma 7 and bounds (10) and (12) for $e_n = f - f_n$ to derive

$$\begin{aligned}\|f - f_n\|_{L_2(\Omega)} &= \|e_n\|_{L_2(\Omega)} \le C\left(h_n^\tau \|e_n\|_{H^\tau(\Omega)} + \|e_n\|_{L_\infty(X_n)}\right)\\ &\le C\left(\left(\frac{h_n}{\delta_n}\right)^\tau \|Ee_{n-1}\|_{\Phi_n} + \sqrt{\epsilon_n}\|Ee_{n-1}\|_{\Phi_n}\right)\\ &\le C\mu^\tau \|Ee_{n-1}\|_{\Phi_n}\\ &= \alpha\|Ee_{n-1}\|_{\Phi_n}.\end{aligned}$$

Now we can apply the first step $n-1$ times to derive

$$\|f - f_n\|_{L_2(\Omega)} \le C\alpha^n \|Ef\|_{\Phi_1} \le C\alpha^n \|Ef\|_{H^\tau(\mathbb{R}^d)} \le C\alpha^n \|f\|_{H^\tau(\Omega)},$$

which is the stated result. □

7 Computational aspects and numerical examples

In this section we will look at the computational cost of the multiscale method and provide examples. However, since the analysis of the computational cost is exactly the same as it has been in the sphere case [6], we will only cite the corresponding result. Moreover, since numerical evidence of the scheme, including real-world applications, has already been given in [2] and [16], we will focus on verifying approximation orders. To this end, we will rewrite the convergence results of Theorem 1 in terms of the fill distance h_n of the finest data set. So far, the resulting bound is not optimal, since it contains a factor C_1^n, where C_1 is the constant in $\alpha = C_1\mu^\tau$.

The results on the computational complexity can be summarized as follows.

Theorem 4 *Suppose that the data sets used in each step of the multiscale algorithm are quasi-uniform. Then, the condition number of each interpolation matrix can be bounded independently of the level. Hence, for a given precision, the conjugate gradient method can solve the linear systems on each level in linear time after a preprocessing step, which collects the necessary information about the non-zero entries in the interpolation matrix and which costs $\mathcal{O}(N_j \log N_j)$ time, where N_j denotes the number of points at that level.*

From now on, we assume that we have a fine data set $X \subseteq \Omega$, on which the target function f is known. To derive the different scales for our multiscale algorithm, we create a, possibly nested, sequence of point sets

$$X_1, X_2, X_3, \ldots, X_n = X,$$

where, for simplicity, we will assume that $h_1 = \mu$ and for $1 \le j \le n$,

$$\delta_j = \nu h_j \quad \text{and} \quad h_{j+1} = \mu h_j$$

with $\nu = \gamma/\mu$. Then, for every $f \in H^\tau(\Omega)$, we are, for example, interested in the approximation problem

$$\min\left\{\|f - v\|_{L_2(\Omega)} : v \in W_1 + \cdots + W_n\right\}.$$

In the situation of Sect. 4, we have by Lemma 2 and the proof of Theorem 1 an upper bound of the form

$$\begin{aligned}\|f - f_n\|_{L_2(\Omega)} &\le C h_n^\tau \|f - f_n\|_{H^\tau(\Omega)}\\ &\le C h_n^\tau \delta_{n+1}^{-\tau} \|E e_n\|_{\Phi_{n+1}}\\ &\le C\alpha^n \|f\|_{H^\tau(\Omega)}.\end{aligned}$$

To analyze this estimate, it is helpful to express it entirely in terms of h_n. Now, since $\alpha = C_1\mu^\tau$ with a constant $C_1 > 0$, of which we do not know very much, and $h_n = \mu^n$, we have $\mu = C_1^{-1/\tau}\alpha^{1/\tau}$ and hence

$$h_n = \mu^n = C_1^{-n/\tau}\alpha^{n/\tau}, \quad h_n^\tau = C_1^{-n}\alpha^n,$$

such that we can rewrite the previous estimate as

$$\|f - f_n\|_{L_2(\Omega)} \le C C_1^n h_n^\tau \|f\|_{H^\tau(\Omega)}. \tag{13}$$

This shows that we have roughly the same convergence as in the non-stationary setting if we consider the number n of levels fixed and let h_n tend to zero. Otherwise, the constant C_1^n will spoil the convergence order in the following sense. From $h_n = \mu^n$ we have that $\log h_n = n\log\mu$ and hence

$$C_1^n = C_1^{\frac{\log h_n}{\log\mu}} = e^{\log C_1 \frac{\log h_n}{\log\mu}} = h_n^{\frac{\log C_1}{\log\mu}} =: h_n^{-\sigma}$$

with $\sigma = -\log C_1/\log\mu \ge 0$. This leads to an estimate of the form

$$E_n := \|f - f_n\|_{L_2(\Omega)} \le C h_n^{\tau-\sigma}\|f\|_{H^\tau(\Omega)}.$$

If we assume equality here, we can estimate $\tau - \sigma$ from two consecutive computations via

$$\frac{E_{j+1}}{E_j} = \left(\frac{h_{j+1}}{h_j}\right)^{\tau-\sigma}$$

and hence

$$(\tau - \sigma) = \frac{\log\frac{h_{j+1}}{h_j}}{\log\frac{E_{j+1}}{E_j}}. \tag{14}$$

Table 1 Number of points in each level and h_j

Level	1	2	3	4	5	6	7	8	9	10
N	9	25	81	289	1089	4225	16641	66049	263169	1050625
h	0.5	0.25	0.125	0.0625	0.03125	0.015625	0.0078125	0.0039063	0.0019531	0.0009766

Table 2 Basis function = $\phi_{2,1} \in C^2$, expected order = 2.5

Level	$\gamma = 1.5$			$\gamma = 2$		
	Error	Order	CG	Error	Order	CG
1	1.774e−01		7	1.841e−01		7
2	2.691e−02	2.72	17	2.818e−02	2.71	20
3	4.957e−03	2.44	31	5.085e−03	2.47	45
4	3.994e−04	3.63	38	3.428e−04	3.89	67
5	8.270e−05	2.27	41	6.565e−05	2.38	84
6	1.866e−05	2.15	40	1.340e−05	2.29	85
7	4.261e−06	2.13	40	2.757e−06	2.28	86
8	9.772e−07	2.12	39	5.695e−07	2.28	85
9	2.241e−07	2.12	39	1.176e−07	2.28	84
10	4.929e−08	2.18	39	2.321e−08	2.34	84

A closer look at the proof of Theorem 1 shows that the constant C_1 is of the form

$$C_1 = \widetilde{C_1}\left(1 + \gamma^{-2\tau}\right)^{1/2},$$

where $\gamma = \nu\mu$ controls the size of the support radius relatively to the fill distance. Hence, we can reduce this constant C_1 by choosing a larger constant γ at the cost of more non-zero entries in each row.

In our first example, we will test the influence of γ on the approximation order $\tau - \sigma$ and hence on C_1. To this end, we use a smooth target function, defined on the unit square $\Omega = [0, 1]^2$ and we will distribute our points on a regular grid of size $h_j = 2^{-j}$, i.e., we choose $X_j = h_j\mathbb{Z}^2 \cap \Omega$. Table 1 lists the number of points and h_j.

Our target function is the standard Franke function, which is a sum of four exponential terms, see [3]. To build our approximants, we use $\Phi = \phi_{2,1}(\|\cdot\|_2) \in C^2(\mathbb{R}^2)$, which is the standard C^2 Wendland function, generating $H^{2.5}(\mathbb{R}^d)$. We stopped the iterations of the non-preconditioned CG method if the relative error falls under 10^{-6}. We tested even higher accuracies but the approximation errors stayed approximately the same, only the number of CG iterations increased.

We ran a series of tests with varying γ. The results are shown in Tables 2, 3, 4, where we have given, for each γ, the $L_2(\Omega)$ error, discretized on a fine 2049×2049 grid, the resulting approximation order according to (14) and the number of CG iterations required.

Table 3 Basis function = $\phi_{2,1} \in C^2$, expected order = 2.5

Level	$\gamma = 2.5$			$\gamma = 3$		
	Error	Order	CG	Error	Order	CG
1	1.888e−01		7	1.918e−01		7
2	2.861e−02	2.72	20	2.894e−02	2.73	23
3	5.174e−03	2.47	53	5.113e−03	2.50	61
4	3.332e−04	3.96	96	3.287e−04	3.96	125
5	5.893e−05	2.50	132	5.510e−05	2.58	197
6	1.125e−05	2.39	141	1.009e−05	2.45	227
7	2.170e−06	2.37	144	1.862e−06	2.44	229
8	4.201e−07	2.37	143	3.448e−07	2.43	231
9	8.126e−08	2.37	142	6.379e−08	2.43	231
10	1.502e−08	2.44	142	1.127e−08	2.50	231

Table 4 Basis function = $\phi_{2,1} \in C^2$, expected order = 2.5

Level	$\gamma = 3.5$			$\gamma = 4$		
	Error	Order	CG	Error	Order	CG
1	1.939e−01		8	1.954e−01		8
2	2.923e−02	2.73	24	2.948e−02	2.73	25
3	5.140e−03	2.51	68	5.175e−03	2.51	74
4	3.120e−04	4.04	147	3.166e−04	4.03	172
5	5.101e−05	2.61	260	4.965e−05	2.67	336
6	9.004e−06	2.50	329	8.514e−06	2.54	438
7	1.606e−06	2.49	347	1.474e−06	2.53	480
8	2.876e−07	2.48	347	2.561e−07	2.52	484
9	5.145e−08	2.48	345	4.447e−08	2.53	486
10	8.788e−09	2.55	346	7.367e−09	2.59	484

It seems that we can get the expected approximation order, or even better, if we just choose γ sufficiently large. The number of CG iterations necessary to achieve the desired accuracy becomes constant with a constant obviously depending on γ. It is remarkable that even in the case of a larger γ like $\gamma = 4$, which leads already to a maximum number of 289 non-zero entries per row, the number of iterations required is still very low.

In our second example, we want to study the behavior for non-smooth functions. To this end, we employ an example often used in the context of elliptic problems. We use the L-shaped domain $\Omega = (-1/2, 1/2) \times (-1/2, 1, 2) \setminus [0, 1/2) \times [0, 1/2)$, on which we introduce polar coordinates $x = r\cos\phi$, $y = r\sin\phi$ with $r \geq 0$ and $\phi \in [\pi/2, 2\pi]$. Our target function is a solution to $\Delta u = 0$ on Ω given by

Table 5 L-shaped domain, basis function $\phi_{2,0} \in C^0$

Level	N	l_2	l_∞	l_2 order	l_∞ order	CG
1	21	1.383194e−02	4.904724e−02			14
2	65	4.660419e−03	3.134300e−02	1.57	0.65	32
3	225	1.476477e−03	1.983347e−02	1.66	0.66	58
4	833	4.517391e−04	1.243790e−02	1.71	0.67	89
5	3201	1.406569e−04	7.834898e−03	1.68	0.67	109
6	12545	4.469146e−05	4.855611e−03	1.65	0.69	114
7	49665	1.451548e−05	2.999504e−03	1.62	0.69	115
8	197633	4.552532e−06	1.493392e−03	1.67	1.01	115
Theory				1.5	0.5	

$u(r, \phi) = -r^{2/3} \sin((2\phi - \pi)/3)$, which belongs to $H^s(\Omega)$ for $s < 1 + 2/3$, see for example [4].

Again, we choose $h_j = 2^{-j}$ and $X_j = h_j\mathbb{Z}^2 \cap \Omega$. Due to the reduced smoothness of the target function, we take Φ to be the kernel $\Phi = \phi_{2,0}(\|\cdot\|_2)$, where $\phi_{2,0}$ is the $C^0(\mathbb{R}^2)$ compactly supported function from [15]. Furthermore, we choose $\mu = 1/2$ and $\nu = 3$. We stop the unpreconditioned CG method if the error drops under a relative bound of 10^{-8}. The error is measured on a fine grid of width $h = 2^{-9}$.

This time, we also look at the L_∞ error. The results are summarized in Table 5. Again, after the first few levels, the number of steps necessary for the CG method to achieve the desired precision is constant. Interestingly, in this example, we have even for a small γ already a better order than theoretically predicted. This is the case for both the L_2 and the L_∞ order.

This example leads to the natural question of local refinement. It seems not to be necessary to choose the data sites for the next level distributed on all of Ω. An immediate adaptive strategy would be to evaluate the error on a finer data set but to include only those points, on which the error is larger than a certain threshold. If in addition a certain quasi-uniformity is maintained, convergence of the CG method should remain unaltered. However, this will be part of future research.

Acknowledgments The author is extremely grateful to Professor Ian H. Sloan and Dr. Q. Thong LeGia from the University of New South Wales, Sydney. They both have revived the author's interest in multiscale approximation using compactly supported radial basis functions during a visit of the author to the University of New South Wales which was supported by the Australian Research Council. Furthermore, both have contributed valuable comments and remarks to earlier versions of this paper.

References

1. DeVore, R.A., Sharpley, R.C.: Besov spaces on domains in $\mathbb{R}^d$. Trans. AMS **335**, 843–864 (1993)
2. Floater, M.S., Iske, A.: Multistep scattered data interpolation using compactly supported radial basis functions. J. Comput. Appl. Math. **73**, 65–78 (1996)
3. Franke, R.: Scattered data interpolation: tests of some methods. Math. Comput. **38**, 181–200 (1982)
4. Hackbusch, W.: Theorie und Numerik elliptischer Differentialgleichungen. Teubner, Stuttgart (1986)

5. Hales, S.J., Levesley, J.: Error estimates for multilevel approximation using polyharmonic splines. Numer. Algorithms **30**, 1–10 (2002)
6. Le Gia, Q.T., Sloan, I., Wendland, H.: Multiscale analysis in Sobolev spaces on the sphere. Preprint Sussex/Sydney (2009)
7. Narcowich, F.J., Schaback, R., Ward, J.D.: Multilevel interpolation and approximation. Appl. Comput. Harmon. Anal. **7**, 243–261 (1999)
8. Narcowich, F.J., Ward, J.D., Wendland, H.: Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. Constr. Approx. **24**, 175–186 (2006)
9. Schaback, R.: Creating surfaces from scattered data using radial basis functions. In: Dæhlen, M., Lyche, T., Schumaker, L.L. (eds.) Mathematical Methods for Curves and Surfaces, pp. 477–496. Vanderbilt University Press, Nashville (1995)
10. Schaback, R.: Error estimates and condition number for radial basis function interpolation. Adv. Comput. Math. **3**, 251–264 (1995)
11. Schaback, R.: On the efficiency of interpolation by radial basis functions. In: Le Méhauté, A., Rabut, C., Schumaker, L.L. (eds.) Surface Fitting and Multiresolution Methods, pp. 309–318. Vanderbilt University Press, Nashville (1997)
12. Schaback, R.: The missing Wendland functions. Adv. Comput. Math. (2009). doi:10.1007/s10444-009-9142-7
13. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton University Press, Princeton (1971)
14. Wahba, G.: Spline Models for Observational Data. CBMS-NSF, Regional Conference Series in Applied Mathematics. Siam, Philadelphia
15. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. **4**, 389–396 (1995)
16. Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge (2005)
17. Wendland, H., Rieger, C.: Approximate interpolation with applications to selecting smoothing parameters. Numer. Math. **101**, 643–662 (2005)
18. Wu, Z.: Characterization of positive definite functions. In: Dæhlen, M., Lyche, T., Schumaker, L.L. (eds.) Mathematical Methods for Curves and Surfaces, pp. 573–578. Vanderbilt University Press, Nashville (1995)