

# Multiscale Radial Basis Functions

Holger Wendland

## 1 Introduction

For the past few decades, radial basis functions have been established as one of the main tools in multivariate approximation theory. They allow the user to build high-order, meshfree approximation spaces and provide an extremely flexible tool for the reconstruction of functions from scattered data, see [6, 15, 46].

The main ingredient are, of course, radial functions, though many of the results are also true for non-radial functions.

**Definition 1.** A radial function is a function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  of the form  $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|_2)$ ,  $\mathbf{x} \in \mathbb{R}^d$ , where  $\phi : [0, \infty) \rightarrow \mathbb{R}$  is a univariate function and  $\|\mathbf{x}\|_2 = (x_1^2 + \dots + x_d^2)^{1/2}$  denotes the Euclidean norm of  $\mathbf{x} \in \mathbb{R}^d$ .

Sometimes, the univariate function  $\phi$  in the above definition is referred to as the radial function but we will not do this here. We will however make the following general assumption.

*Remark 1.* We will always assume that the univariate function  $\phi : [0, \infty) \rightarrow \mathbb{R}$  is defined on all of  $\mathbb{R}$  by even extension, i.e. by  $\phi(-r) := \phi(r)$  for  $r > 0$ .

We will say that such a  $\phi : [0, \infty)$  generates or induces a multivariate function on  $\mathbb{R}^d$ . This induced function is simply defined to be  $\Phi(\mathbf{x}) := \phi(\|\mathbf{x}\|_2)$ ,  $\mathbf{x} \in \mathbb{R}^d$ .

In this article, we will predominantly be dealing with radial functions  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  having a compact support, i.e. with functions for which

$$\text{supp}\Phi := \overline{\{\mathbf{x} \in \mathbb{R}^d : \Phi(\mathbf{x}) \neq 0\}}$$

---

H. Wendland (✉)

Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany

e-mail: [holger.wendland@uni-bayreuth.de](mailto:holger.wendland@uni-bayreuth.de)

is bounded. In general, we will simply assume that the support of  $\Phi$  is given by the unit ball. In that case, we can introduce scaled versions, i.e.

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

with  $\delta > 0$ , which obviously have support in the ball about the origin with radius  $\delta$ . Besides scaling we can also shift a scaled basis function. This allows us to define our basic approximation spaces. To this end let  $\Omega \subseteq \mathbb{R}^d$  be a bounded domain and let  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$  be a set of distinct points, the *data sites*. Associated to such a point set are two geometric quantities, the *fill distance* or *mesh norm*  $h_{X,\Omega}$  and the separation distance  $q_X$ , which are defined to be

$$h_{X,\Omega} := \sup_{\mathbf{x} \in \Omega} \min_{1 \leq j \leq N} \|\mathbf{x} - \mathbf{x}_j\|_2,$$

$$q_X := \min_{1 \leq j \neq k \leq N} \|\mathbf{x}_j - \mathbf{x}_k\|_2,$$

respectively. The fill distance describes how well the scattered points  $X$  “cover” the domain  $\Omega$  in the following sense. The fill distance  $h_{X,\Omega}$  gives the radius of the largest ball completely contained in  $\Omega$  without a data site within its interior. The fill distance is important for understanding approximation orders. The separation distance defines the smallest distance between the two closest data sites. It becomes important when looking at stability issues.

A scaled, compactly supported kernel and a set of data sites are enough to define a first approximation space via

$$W_X = W_{X,\Phi_\delta} := \text{span}\{\Phi_\delta(\cdot - \mathbf{x}) : \mathbf{x} \in X\}. \quad (1)$$

We will refer to this as a *local approximation space* since we will not deal with only one data set and one associated approximation space, but with a sequence of data sets  $X_1, X_2, \dots$  with mesh norms  $h_j = h_{X_j,\Omega}$ , which are supposed to be 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  $q_j = q_{X_j}$ ,

$$q_j \leq h_j \leq c_q q_j.$$

Having this sequence of data sets  $X_j$ , it is clear that we can build a local approximation space  $W_j = W_{X_j}$  for each  $X_j$ . To this end, we will define a kernel  $\Phi_j : \Omega \times \Omega \rightarrow \mathbb{R}$  for each level. In our application this kernel will be given by the scaled version of a fixed translation invariant radial basis function. To be more precise, we assume that there is a compactly supported function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  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  $\delta_j$  and centre  $\mathbf{y}$ .

As we assume that the data sets become denser and denser, we will also 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  $\mu$  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

$$W_j = \text{span}\{\Phi_j(\cdot, \mathbf{x}) : \mathbf{x} \in X_j\}. \quad (2)$$

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  $n \rightarrow \infty$ .

The rest of the paper is organised as follows. In the next section, we will discuss the relevant material on optimal recovery, reproducing kernel Hilbert spaces and Sobolev spaces. After that, we will discuss compactly supported radial basis functions. In the fourth section, we will then address multiscale interpolation and approximation.

## 2 Optimal Recovery and Reproducing Kernel Hilbert Spaces

### 2.1 The Reconstruction Problem

Our main tool in using discrete approximation spaces  $W_X$  of the form (1) will be optimal recovery and interpolation, which turns out to be the same in this context. Throughout this section, we will consider only local approximation spaces  $W_{X, \phi_\delta}$ . In particular, we want to keep the scaling factor  $\delta$  fixed but might vary the data set  $X$ .

To understand why radial basis functions, or, more generally, positive definite functions are natural tools in multivariate approximation theory, we start with an abstract result from linear algebra.

To this end, let  $H$  be a Hilbert space and denote its dual by  $H^*$ . Suppose  $\Lambda = \{\lambda_1, \dots, \lambda_N\} \subseteq H^*$  is a set of linearly independent functionals on  $H$  and  $f_1, \dots, f_N \in \mathbb{R}$  are certain given values. Then a generalised recovery problem would ask for finding a function  $s \in H$  so that  $\lambda_j(s) = f_j$ ,  $1 \leq j \leq N$ . We will call  $s$  a *generalised interpolant*. Usually, there are an infinite number of possible generalised interpolants and hence there is need to distinguish one of them. Moreover, if the data contain noise, interpolation is not an appropriate way of reconstructing the unknown function  $f$ . Hence, we will also consider a method, which was originally introduced in the context of splines (see [42]).

**Definition 2.** The *norm-minimal generalised interpolant* is the element  $s_0 \in H$  satisfying

$$\|s_0\| = \min\{\|s\| : s \in H, \lambda_j(s) = f_j, 1 \leq j \leq N\}. \tag{3}$$

Given a *smoothing parameter*  $\varepsilon > 0$ , the *optimal recovery* or *smoothing spline* is the element  $s_\varepsilon \in H$  satisfying

$$\min \left\{ \sum_{j=1}^N |\lambda_j(s) - f_j|^2 + \varepsilon \|s\|_H^2 \right\}. \tag{4}$$

The solutions to both problems (3) and (4) are unique. A proof of the following result can be found in [42, 46].

**Theorem 1.** Let  $H$  be a Hilbert space,  $\lambda_1, \dots, \lambda_N \in H^*$  linearly independent functionals with Riesz representers  $v_j$ ,  $1, \leq j \leq N$ , and let  $f_1, \dots, f_N \in \mathbb{R}$  be given. The unique solutions  $s_0$  of (3) and  $s_\varepsilon$  of (4) can be written as

$$s_\varepsilon = \sum_{j=1}^N \alpha_j v_j, \tag{5}$$

$\varepsilon \geq 0$ , where the coefficients  $\{\alpha_j\}$  are the solution of the linear system

$$(A + \varepsilon I)\boldsymbol{\alpha} = \mathbf{f} \tag{6}$$

where  $A = (\lambda_i(v_j))$  and  $\mathbf{f} = (f_1, \dots, f_N)^T$ .

Hence, the optimal recovery problem can constructively be solved once the Riesz representer of the functionals are known. For general Hilbert spaces of functions this is not an easy task. However, for *reproducing kernel Hilbert spaces* this becomes simple once the reproducing kernel is known.

**Definition 3.** A Hilbert space  $H$  of functions  $f : \Omega \rightarrow \mathbb{R}$  is a reproducing kernel Hilbert space if there is a kernel  $\Phi : \Omega \times \Omega \rightarrow \mathbb{R}$  with

1.  $\Phi(\cdot, \mathbf{x}) \in H$  for all  $\mathbf{x} \in \Omega$ ,
2.  $f(\mathbf{x}) = \langle f, \Phi(\cdot, \mathbf{x}) \rangle_H$  for all  $\mathbf{x} \in \Omega$  and all  $f \in H$ .

In such a reproducing kernel Hilbert space, the Riesz representer of any functional  $\lambda \in H^*$  is simply given by applying  $\lambda$  to one of the arguments of the kernel, i.e.  $v_\lambda = \lambda^y \Phi(\cdot, \mathbf{y})$ . In this paper, we will exclusively be concerned with point evaluation functionals, i.e.  $\lambda_j(f) = f(\mathbf{x}_j)$ . Hence, the optimal recovery takes the simple form

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

and the matrix  $A$  has thus entries  $\Phi(\mathbf{x}_i, \mathbf{x}_j)$ .

## 2.2 Sobolev Spaces

The reproducing kernel Hilbert spaces we will be concerned with are Sobolev spaces. For an integer  $\sigma = k \in \mathbb{N}_0$  the Sobolev space  $H^k(\Omega)$  consists of all those functions  $u \in L_2(\Omega)$  which have weak derivatives  $D^\alpha u$  in  $L_2(\Omega)$  for all  $|\alpha| \leq k$ . For non-integer  $\sigma > 0$ , the Sobolev space  $H^\sigma(\Omega)$  can be defined using interpolation theory, see, for example, [1].

The Sobolev embedding theorem guarantees that  $H^\sigma(\Omega)$  is a reproducing kernel Hilbert space for any  $\sigma > d/2$ . However, since the reproducing kernel is uniquely determined by the inner product and also depends upon  $\Omega$  in our setting, it is in general not possible to give an explicit formula for the reproducing kernel. We will circumvent this problem by two means. First of all, we will assume that we can extend our functions defined on  $\Omega$  to all of  $\mathbb{R}^d$ , see [4].

**Proposition 1.** *Let  $\Omega \subseteq \mathbb{R}^d$  be a bounded domain with a Lipschitz boundary. Then, there is a universal extension operator  $E : H^\sigma(\Omega) \rightarrow H^\sigma(\mathbb{R}^d)$  satisfying*

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

for all  $f \in H^\sigma(\Omega)$  and all  $\sigma \geq 0$ .

It is important that though the constant  $C_\sigma > 0$  might depend on  $\sigma$ , the operator  $E$  itself does not. The existence of such an operator means that we can equip  $H^\sigma(\Omega)$  with an equivalent norm  $f \mapsto \|Ef\|_{H^\sigma(\mathbb{R}^d)}$  and thus we can concentrate on determining the reproducing kernel of  $H^\sigma(\mathbb{R}^d)$  for  $\sigma > d/2$ . This becomes particularly easy if we write the norm on  $H^\sigma(\mathbb{R}^d)$  using the Fourier transform  $\widehat{f}$  of  $f \in H^\sigma(\mathbb{R}^d)$ ,  $\sigma > d/2$  defined by

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

With this, we can define the norm on  $H^\sigma(\mathbb{R}^d)$  to be

$$\|f\|_{H^\sigma(\mathbb{R}^d)}^2 = \int_{\mathbb{R}^d} |\widehat{f}(\boldsymbol{\omega})|^2 (1 + \|\boldsymbol{\omega}\|_2^2)^\sigma d\boldsymbol{\omega}. \tag{7}$$

The reproducing kernel can then be written in translation-invariant form, i.e. it satisfies  $\Phi(\mathbf{x}, \mathbf{y}) = \Phi_0(\mathbf{x} - \mathbf{y})$  with  $\Phi_0 : \mathbb{R}^d \rightarrow \mathbb{R}$  for which we will simply write  $\Phi(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y})$  in an abuse of notation. The function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  is then simply the inverse Fourier transform of one over the weight function, i.e.

$$\widehat{\Phi}(\boldsymbol{\omega}) = (1 + \|\boldsymbol{\omega}\|_2^2)^{-\sigma},$$

which is a consequence of the following general result, see [46].

**Proposition 2.** *Suppose  $\Phi \in L_1(\mathbb{R}^d)$  has a Fourier transform satisfying*

$$c_1(1 + \|\omega\|_2^2)^{-\sigma} \leq \widehat{\Phi}(\omega) \leq c_2(1 + \|\omega\|_2^2)^{-\sigma} \tag{8}$$

*Then,  $\Phi$  also defines a reproducing kernel on  $H^\sigma(\mathbb{R}^d)$  with respect to the inner product defined by*

$$\langle f, g \rangle_\Phi := \int_{\mathbb{R}^d} \frac{\widehat{f}(\omega)\overline{\widehat{g}(\omega)}}{\widehat{\Phi}(\omega)} d\omega.$$

*The norm associated to this inner product is equivalent to the Sobolev norm (7).*

This also opens the door to the construction of a variety of kernels, which are also reproducing kernels in  $H^\sigma(\mathbb{R}^d)$ . We will only consider compactly supported  $\Phi$ 's having a Fourier transform with such a decay. However, since for  $t \geq 0$  and  $\sigma \geq 0$ , we have

$$1 + t^\sigma \leq (1 + t)^\sigma \leq 2^\sigma(1 + t^\sigma)$$

we see that the decay property (8) is equivalent to the decay property

$$c_1(1 + \|\omega\|_2^{2\sigma})^{-1} \leq \widehat{\Phi}(\omega) \leq c_2(1 + \|\omega\|_2^{2\sigma})^{-1}, \tag{9}$$

which is slightly easier to handle in what follows. An immediate consequence of these observations is the following one.

**Corollary 1.** *Let  $\Phi \in L_1(\mathbb{R}^d)$  be a reproducing kernel of  $H^\sigma(\mathbb{R}^d)$ ,  $\sigma > d/2$ , i.e. its Fourier transform satisfies (9). For  $\delta \in (0, 1]$  let  $\Phi_\delta := \delta^{-d}\Phi(\cdot/\delta)$ . Then,  $\Phi_\delta$  is also a reproducing kernel of  $H^\sigma(\mathbb{R}^d)$  and for every  $g \in H^\sigma(\mathbb{R}^d)$ , we have the norm equivalence*

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

*Proof.* This follows immediately from  $\widehat{\Phi}_\delta = \widehat{\Phi}(\delta \cdot)$ . □

It is important to note that the Hilbert space in which the scaled kernel  $\Phi_\delta$  is the reproducing kernel is always  $H^\sigma(\mathbb{R}^d)$ , i.e. the space itself does not depend on the scaling parameter. However, the norm associated to the inner product in which  $\Phi_\delta$  is the reproducing kernel does depend on  $\delta > 0$ . Also the equivalence constants depend on  $\delta$ . While we always have  $\|f\|_{\Phi_\delta} \leq C\|f\|_{H^\sigma(\mathbb{R}^d)}$  with  $C > 0$  being independent of  $\delta > 0$ , the constant in the other estimate  $\|f\|_{H^\sigma(\mathbb{R}^d)} \leq C\delta^{-\sigma}\|f\|_{\Phi_\delta}$  tends to infinity with  $\delta \rightarrow 0$ . This is not surprising at all, since  $\|\cdot\|_{\Phi_\delta} \rightarrow \|\cdot\|_{L_2(\mathbb{R}^d)}$  for  $\delta \rightarrow 0$  and the constant has to balance this loss of derivative in the norm.

### 3 Compactly Supported Radial Basis Functions

In this section we will review recent results on compactly supported radial basis functions. We are particularly interested in the construction of such basis functions and in those compactly supported RBFs which are reproducing kernels of Sobolev spaces.

Much of the general theory in this section can be found in [48]. Newer results are based on [8, 9, 22, 35, 37, 43, 50, 51].

#### 3.1 Construction of Compactly Supported RBFs

A famous result of Bochner states that a continuous, integrable function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  is *positive definite* and hence the reproducing kernel of its associated reproducing kernel Hilbert space, if and only if it possesses a non-negative, non-vanishing Fourier transform

$$\widehat{\Phi}(\omega) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \Phi(\mathbf{x}) e^{-i\mathbf{x}^T \omega} d\mathbf{x}, \quad \omega \in \mathbb{R}^d,$$

see [46] for a proof. For a radial, positive definite function, it is well-known that the Fourier transform is also radial and can be reduced to a one-dimensional Hankel transform. This will be essential in what we now want to do. To formulate this, we need to introduce Bessel functions of the first kind.

**Definition 4.** The Bessel function of the first kind of order  $\nu \in \mathbb{C}$  is defined by

$$J_\nu(z) := \sum_{m=0}^{\infty} \frac{(-1)^m (\frac{1}{2}z)^{2m+\nu}}{m! \Gamma(\nu + m + 1)}, \quad z \in \mathbb{C}.$$

This gives the desired representation in the case of a positive definite, radial function.

**Theorem 2.** Suppose  $\Phi \in L_1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$  is radial, i.e.  $\Phi(\mathbf{x}) = \phi(\|\mathbf{x}\|_2)$ ,  $\mathbf{x} \in \mathbb{R}^d$ . Then its Fourier transform  $\widehat{\Phi}$  is also radial, i.e.  $\widehat{\Phi}(\omega) = \mathcal{F}_d \phi(\|\omega\|_2)$  with

$$\mathcal{F}_d \phi(r) = r^{-\frac{d-2}{2}} \int_0^\infty \phi(t) t^{\frac{d}{2}} J_{\frac{d-2}{2}}(rt) dt. \tag{10}$$

Hence,  $\Phi$  is positive definite if and only if  $\mathcal{F}_d \phi \geq 0$  and  $\mathcal{F}_d \phi \not\equiv 0$ .

An interesting consequence of (10) is that the operator  $\mathcal{F}_d$  cannot only be defined for  $d \in \mathbb{N}$  but actually for all  $d \in [0, \infty)$ , which we will from now on do. However, the connection to a multivariate Fourier transform is, of course, only given if  $d \in \mathbb{N}$ . Also, the existence of the integral has to be ensured.

Next, we introduce the key operators to construct compactly supported radial basis functions.

**Definition 5.** 1. Let  $\phi$  be given so that  $t \mapsto \phi(t)t$  is in  $L_1[0, \infty)$ , then we define for  $r \geq 0$ ,

$$(\mathcal{I}\phi)(r) = \int_r^\infty t\phi(t)dt.$$

2. For even  $\phi \in C^2(\mathbb{R})$  we define for  $r \geq 0$ ,

$$(\mathcal{D}\phi)(r) = -\frac{1}{r}\phi'(r).$$

In both cases the resulting functions should be seen as even functions by even extension.

Obviously, these operators are inverse to each other in the sense that  $\mathcal{I}\mathcal{D} = I$  and  $\mathcal{D}\mathcal{I} = I$  whenever these operations are well-defined, where  $I$  denotes the identity operator.

The importance of these operators comes from the fact that they nicely work together with the Fourier transform operator  $\mathcal{F}_d$ . We will formulate this in a moment but before that, we want to generalise them.

**Definition 6.** Let  $\alpha > 0$  and assume that  $t \mapsto t\phi(t)$  is integrable over  $[0, \infty)$ . Then, we define for  $r \geq 0$ ,

$$\mathcal{I}_\alpha\phi(r) := \frac{1}{2^{\alpha-1}\Gamma(\alpha)} \int_r^\infty \phi(t)t(t^2 - r^2)^{\alpha-1} dt$$

and for  $r < 0$  the function  $\mathcal{I}_\alpha\phi$  is defined by even extension.

Note that we can rewrite the integral in the definition as

$$\int_r^\infty \phi(t)t(t+r)^{\alpha-1}(t-r)^{\alpha-1} dt.$$

Hence, for  $\alpha \in (0, 1)$ , the singularity at  $t = r$  is integrable. Since we will be merely be concerned with continuous compactly supported functions in this paper, the operator  $\mathcal{I}_\alpha$  can always be applied to such a function for any  $\alpha > 0$ .

We also see that for  $\alpha = 1$  we have  $\mathcal{I}_1 = \mathcal{I}$ , i.e. both definitions coincide. Finally, we will particularly be interested in the case  $\alpha = 1/2$ . In this case the operator reduces to

$$\mathcal{I}_{1/2}\phi(r) = \sqrt{\frac{2}{\pi}} \int_r^\infty \phi(t)t(t^2 - r^2)^{-1/2} dt.$$



Since we are merely interested in compactly supported functions, it is important for us to note that  $\mathcal{I}_\alpha$  respects a compact support in the following sense.

**Lemma 1.** *Assume that the even function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  has a compact support contained in the interval  $[-R, R]$ . Then, for each  $\alpha > 0$  the function  $\mathcal{I}_\alpha\phi$  has also compact support contained in  $[-R, R]$ .*

We are now going to extend the range of possible operators  $\mathcal{I}_\alpha$  by defining them also for  $\alpha \leq 0$ . We will do this first formally in the following definition.

- Definition 7.**
1. We define  $\mathcal{I}_0$  to be the identity operator, i.e.  $\mathcal{I}_0\phi = \phi$  for all even functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ .
  2. For  $n \in \mathbb{N}$  we define  $\mathcal{I}_{-n} := \mathcal{D}^n$ .
  3. For  $\alpha > 0$  we let  $k := \lceil \alpha \rceil$  and define  $\mathcal{I}_{-\alpha} := \mathcal{I}_{-k}\mathcal{I}_{k-\alpha}$ .

To see to what functions  $\phi$  we can apply these operators, we restrict ourselves to compactly supported and even functions. If such a function possesses  $\ell + 1$  continuous derivatives then we first note that the derivatives must satisfy  $\phi^{(j)}(r) = (-1)^j\phi^{(j)}(-r)$ . In particular, we see that for odd  $j$  the derivative at zero has to vanish. The Taylor expansion about the origin is thus

$$\begin{aligned} \phi(r) &= \sum_{j=0}^{\ell} \frac{\phi^{(j)}(0)}{j!} r^j + \frac{1}{\ell!} \int_0^r (r-t)^\ell \phi^{(\ell+1)}(t) dt \\ &= \sum_{j=0}^{\lfloor \ell/2 \rfloor} \frac{\phi^{(2j)}(0)}{(2j)!} r^{2j} + \frac{1}{\ell!} \int_0^r (r-t)^\ell \phi^{(\ell+1)}(t) dt. \end{aligned}$$

Applying  $\mathcal{I}_{-1} = \mathcal{D}$  to this expansion yields

$$\begin{aligned} \mathcal{I}_{-1}\phi(r) &= -\frac{1}{r} \frac{d}{dr} \phi(r) \\ &= -\sum_{j=1}^{\lfloor \ell/2 \rfloor} \frac{\phi^{(2j)}(0)}{(2j-1)!} r^{2j-2} - \frac{1}{(\ell-1)!} \frac{1}{r} \int_0^r (r-t)^{\ell-1} \phi^{(\ell+1)}(t) dt. \end{aligned}$$

The critical term on the right-hand side behaves like

$$\lim_{r \rightarrow 0} \frac{1}{r} \int_0^r (r-t)^{\ell-1} \phi^{(\ell+1)}(t) dt = \lim_{r \rightarrow 0} (\ell-1) \int_0^r (r-t)^{\ell-2} \phi^{(\ell+1)}(t) dt = 0$$

as long as  $\ell \geq 2$  and like  $\phi^{(2)}(0)$  for  $\ell = 1$ . Hence, for even  $\phi \in C^2(\mathbb{R})$  with compact support  $\mathcal{I}_{-1}\phi$  is well-defined and continuous. We can iterate this process to derive the following result.

**Corollary 2.** *Let  $\alpha > 0$  and  $k = \lceil \alpha \rceil$ . If  $\phi \in C^{2k}(\mathbb{R})$  is even and compactly supported, then  $\mathcal{I}_{-\alpha}\phi$  is well-defined, continuous and compactly supported.*

However, more important to us is the following observation, which can be derived from the above corollary but also straight-forward by induction.

**Corollary 3.** *If  $k \in \mathbb{N}$ , then for every even  $\phi \in C(\mathbb{R})$  with compact support,  $\psi := \mathcal{I}_k\phi$  and  $\mathcal{I}_{-k}\psi$  are well-defined and satisfy*

$$\mathcal{I}_{-k}\mathcal{I}_k\phi = \phi.$$

*Proof.* We already know that  $\psi$  is well-defined. To see that  $\mathcal{I}_{-k}\psi$  is also well-defined, we first recall that the fundamental theorem of calculus immediately yields  $\mathcal{I}_{-1}\mathcal{I}_1\phi = \phi$ . For  $k \geq 2$ , we first compute

$$\begin{aligned} \mathcal{I}_{-1}\mathcal{I}_k\phi(r) &= -\frac{1}{2^{k-1}(k-1)!} \frac{1}{r} \frac{d}{dr} \int_r^\infty \phi(t)t(t^2-r^2)^{k-1} dt \\ &= \frac{2r(k-1)}{2^{k-1}(k-1)!} \frac{1}{r} \int_r^\infty \phi(t)t(t^2-r^2)^{k-2} dt \\ &= \frac{1}{2^{k-2}(k-2)!} \int_r^\infty \phi(t)t(t^2-r^2)^{k-2} dt \\ &= \mathcal{I}_{k-1}\phi(r) \end{aligned}$$

and the rest follows by induction. □

To understand the interaction of the operators  $\mathcal{I}_\alpha$  and  $\mathcal{F}_d$ , we need the following relation on Bessel functions.

**Lemma 2.** *For  $\mu > -1/2$  and  $\nu > -1$  we have*

$$J_{\mu+\nu+1}(t) = \frac{t^{\nu+1}}{2^\nu \Gamma(\nu+1)} \int_0^1 J_\mu(tu)u^{\mu+1}(1-u^2)^\nu du, \quad t > 0.$$

The proof of this lemma starts with the integral on the right-hand side. It uses the definition of the Bessel function as an infinite series, exchanges integration and summation, integrates the resulting terms and interprets the result as a Bessel function again. Details can be found in [38, Lemma 4.13].

**Theorem 3.** *For each  $d, \alpha > 0$  we have*

$$\mathcal{F}_d\mathcal{I}_\alpha\phi = \mathcal{F}_{d+2\alpha}\phi \tag{11}$$

*Hence,  $\phi$  defines a positive definite function on  $\mathbb{R}^{d+2\alpha}$  if and only if  $\mathcal{I}_\alpha$  defines a positive definite function on  $\mathbb{R}^d$ .*

*Proof.* We compute  $\mathcal{F}_d\mathcal{I}_\alpha$  using Theorem 2 and Fubini’s theorem

$$\begin{aligned} \mathcal{F}_d \mathcal{I}_\alpha \phi(r) &= \frac{1}{2^{\alpha-1} \Gamma(\alpha)} r^{-\frac{d-2}{2}} \int_0^\infty \int_t^\infty \phi(s) s (s^2 - t^2)^{\alpha-1} t^{\frac{d}{2}} J_{\frac{d-2}{2}}(rt) ds dt \\ &= \frac{1}{2^{\alpha-1} \Gamma(\alpha)} r^{-\frac{d-2}{2}} \int_0^\infty \phi(s) s \int_0^s (s^2 - t^2)^{\alpha-1} t^{\frac{d}{2}} J_{\frac{d-2}{2}}(rt) dt ds. \end{aligned}$$

After a change of variables  $u := t/s$ , the inner integral can be computed using Lemma 2:

$$\begin{aligned} \int_0^s (s^2 - t^2)^{\alpha-1} t^{\frac{d}{2}} J_{\frac{d-2}{2}}(rt) dt &= \int_0^1 (s^2 - u^2 s^2)^{\alpha-1} (us)^{\frac{d}{2}} J_{\frac{d-2}{2}}(rsu) s du \\ &= s^{2\alpha-1+\frac{d}{2}} \int_0^1 (1 - u^2)^{\alpha-1} u^{\frac{d}{2}} J_{\frac{d-2}{2}}(rsu) du \\ &= s^{2\alpha-1+\frac{d}{2}} 2^{\alpha-1} \Gamma(\alpha) (rs)^{-\alpha} J_{\frac{d-2}{2}+\alpha}(rs). \end{aligned}$$

Inserting this into the above formula for  $\mathcal{F}_d \mathcal{I}_\alpha$  immediately yields

$$\begin{aligned} \mathcal{F}_d \mathcal{I}_\alpha \phi(r) &= r^{-\frac{d-2}{2}-\alpha} \int_0^\infty \phi(s) s s^{\alpha-1+\frac{d}{2}} J_{\frac{d-2}{2}+\alpha}(rs) ds \\ &= r^{-\frac{d+2\alpha-2}{2}} \int_0^\infty \phi(s) s^{\frac{d+2\alpha}{2}} J_{\frac{d+2\alpha-2}{2}}(rs) ds \\ &= \mathcal{F}_{d+2\alpha} \phi(r). \end{aligned}$$

□

This means particularly for the operators  $\mathcal{I} = \mathcal{I}_1$  and  $\mathcal{I}_{1/2}$  the following.

- Corollary 4.** • A function  $\phi$  induces a positive definite function on  $\mathbb{R}^{d+2}$  if and only if  $\mathcal{I} \phi$  induces a positive definite function on  $\mathbb{R}^d$ .
- A function  $\phi$  induces a positive definite function on  $\mathbb{R}^{d+1}$  if and only if  $\mathcal{I}_{1/2} \phi$  induces a positive definite function on  $\mathbb{R}^d$ .

However, (11) has another important consequence. If  $\phi \in C(\mathbb{R})$  has compact support, then we can interpret the classical Fourier transform of  $\phi$  also in the  $L_2$  sense. In the  $L_2$ -sense,  $\phi$  is uniquely determined by its Fourier transform and hence we have the following result.

**Corollary 5.** Let  $\alpha, \beta > 0$ . Assume that  $\phi \in C(\mathbb{R})$  is even and has compact support then we have

$$\mathcal{I}_\alpha \mathcal{I}_\beta \phi = \mathcal{I}_{\alpha+\beta} \phi = \mathcal{I}_\beta \mathcal{I}_\alpha \tag{12}$$

and

$$\mathcal{I}_{-\alpha} \mathcal{I}_\alpha \phi = \phi. \tag{13}$$

*Proof.* Using (11) yields

$$\mathcal{F}_d \mathcal{I}_\alpha \mathcal{I}_\beta \phi = \mathcal{F}_{d+2\alpha} \mathcal{I}_\beta \phi = \mathcal{F}_{d+2\alpha+2\beta} \phi.$$

The other expressions in the first identity have the same Fourier transform and hence all of the stated functions must be the same.

To prove the second equality, let  $k = \lceil \alpha \rceil$ . By definition, we have  $\mathcal{I}_{-\alpha} = \mathcal{I}_{-k} \mathcal{I}_{k-\alpha}$ . Hence,

$$\mathcal{I}_{-\alpha} \mathcal{I}_\alpha \phi = \mathcal{I}_{-k} \mathcal{I}_{k-\alpha} \mathcal{I}_\alpha \phi = \mathcal{I}_{-k} \mathcal{I}_k \phi = \phi,$$

where we first have used (12) and then Corollary 3. □

### 3.2 Wendland Functions and Generalised Wendland Functions

We will now describe a popular class of compactly supported radial basis functions, which is widely used in applications. The starting point is the cut-off power function

$$\phi_\nu(r) = (1 - r)_+^\nu \tag{14}$$

which is known to induce a positive definite function on  $\mathbb{R}^d$  if  $\nu \geq (d + 1)/2$ , see [3, 46].

**Definition 8.** The Wendland function of smoothness  $2k$  for space dimension  $d$  is defined by

$$\begin{aligned} \phi_{d,k}(r) &= \mathcal{I}_k \phi_{\lfloor d/2 \rfloor + k + 1}(r) \\ &= \frac{1}{2^{k-1}(k-1)!} \int_r^1 (1-t)^{\lfloor d/2 \rfloor + k + 1} t(t^2 - r^2)^{k-1} dt, \quad 0 \leq r \leq 1. \end{aligned}$$

The definition shows that we will have the same function for even  $d = 2n$  and odd  $d = 2n + 1$ . We will address this issue later on. However, it is also clear from this definition that  $\phi_{d,k}$  restricted to  $[0, 1]$  is a polynomial of degree  $\lfloor d/2 \rfloor + 3k + 1$ , which can easily be computed. Some of them are, up to a constant factor, listed in Table 1.

The reason why these functions have become quite popular is summarised in the following theorem.

**Theorem 4.** *The functions  $\phi_{d,k}$  are positive definite on  $\mathbb{R}^d$  and are of the form*

$$\phi_{d,k}(r) = \begin{cases} p_{d,k}(r), & 0 \leq r \leq 1, \\ 0, & r > 1, \end{cases}$$

**Table 1** Wendland functions.

Space dimension	Function	Smoothness
$d = 1$	$\phi_{1,0}(r) = (1 - r)_+$	$C^0$
	$\phi_{1,1}(r) \doteq (1 - r)_+^3(3r + 1)$	$C^2$
	$\phi_{1,2}(r) \doteq (1 - r)_+^5(8r^2 + 5r + 1)$	$C^4$
$d \leq 3$	$\phi_{3,0}(r) = (1 - r)_+^2$	$C^0$
	$\phi_{3,1}(r) \doteq (1 - r)_+^4(4r + 1)$	$C^2$
	$\phi_{3,2}(r) \doteq (1 - r)_+^6(35r^2 + 18r + 3)$	$C^4$
	$\phi_{3,3}(r) \doteq (1 - r)_+^8(32r^3 + 25r^2 + 8r + 1)$	$C^6$
$d \leq 5$	$\phi_{5,0}(r) = (1 - r)_+^3$	$C^0$
	$\phi_{5,1}(r) \doteq (1 - r)_+^5(5r + 1)$	$C^2$
	$\phi_{5,2}(r) \doteq (1 - r)_+^7(16r^2 + 7r + 1)$	$C^4$

with a univariate polynomial  $p_{d,k}$  of degree  $\lfloor d/2 \rfloor + 3k + 1$ . They possess continuous derivatives up to order  $2k$ . The degree of  $p_{d,k}$  is minimal for a given space dimension  $d$  and smoothness  $2k$  and are up to a constant factor uniquely determined by this setting.

The above defined functions have been generalised in the following way, see [8, 9, 22, 35].

**Definition 9.** Let  $\alpha > 0$  and  $\nu > -1$  such that  $\alpha + \nu > 0$ . The generalised Wendland functions are defined to be

$$\psi_{\nu,\alpha}(r) := \mathcal{I}_\alpha \phi_\nu(r) = \frac{1}{2^{\alpha-1} \Gamma(\alpha)} \int_r^1 (1-t)^\nu t(t^2 - r^2)^{\alpha-1} dt.$$

Obviously, we have

$$\phi_{d,k} = \psi_{\lfloor d/2 \rfloor + k + 1, k}$$

so that the new functions are indeed a generalisation of the old ones. However, for arbitrary  $\alpha$  and  $\nu$ , we can neither expect  $\psi_{\nu,\alpha}$  to be positive definite on  $\mathbb{R}^d$  nor will  $\psi_{\nu,\alpha}$  in general be representable by a univariate polynomial within its support.

Nonetheless, using the machinery so far, we can compute the Fourier transform of these functions as

$$\begin{aligned} \mathcal{F}_d \psi_{\nu,\alpha}(r) &= \mathcal{F}_{d+2\alpha} \phi_\nu(r) = r^{-(d-2)/2-\alpha} \int_0^1 s^{d/2+\alpha} (1-s)^\nu J_{(d-2)/2+\alpha}(rs) ds \\ &= r^{-d-2\alpha-\nu} \int_0^r t^{d/2+\alpha} (r-t)^\nu J_{(d-2)/2+\alpha}(t) dt. \end{aligned}$$

The latter integral has intensively been studied and can be expressed using hypergeometric series. Recall that such a series is formally defined to be

$$\begin{aligned}
 {}_pF_q(z) &\equiv {}_pF_q[a_1, \dots, a_p; b_1, \dots, b_q; z] \\
 &:= \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n z^n}{(b_1)_n \cdots (b_q)_n n!} =: \sum_{n=0}^{\infty} \frac{(\mathbf{a})_n z^n}{(\mathbf{b})_n n!},
 \end{aligned}$$

where we assume that neither of the  $b_1, \dots, b_q$  is a non-positive integer and use the Pochhammer symbol defined as  $(a)_n := 1$  for  $n = 0$  and  $(a)_n := a(a + 1) \cdots (a + n - 1) = \Gamma(a + n)/\Gamma(a)$  and  $(\mathbf{a})_n := (a_1)_n \cdots (a_p)_n$ . Such a series is known to converge point-wise if  $p \leq q$ , which is the case we are interested in.

A few simple properties and the connection to the integrals we are interested in are collected in the next lemma.

- Lemma 3.** 1. If  $a_p = b_q$ , then  ${}_pF_q = {}_{p-1}F_{q-1}$ .  
 2. For  $\beta + \mu > -1$  and  $\lambda > -1$  we have

$$\begin{aligned}
 \int_0^r (r-t)^\lambda t^\mu J_\beta(t) dt &= \frac{\Gamma(\lambda + 1)\Gamma(\beta + \mu + 1)}{\Gamma(\beta + 1)\Gamma(\beta + \lambda + \mu + 2)} 2^{-\beta} r^{\beta + \lambda + \mu + 1} \\
 &\cdot {}_2F_3 \left[ \frac{\beta + \mu + 1}{2}, \frac{\beta + \mu + 2}{2}; \beta + 1, \frac{\beta + \lambda + \mu + 2}{2}, \frac{\beta + \lambda + \mu + 3}{2}; -r^2/4 \right].
 \end{aligned}$$

3. The derivatives of the hyper-geometric functions satisfy for  $n \in \mathbb{N}$ :

$$\begin{aligned}
 &\frac{d^n}{dz^n} {}_pF_q[a_1, \dots, a_p; b_1, \dots, b_q; z] \\
 &= \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} {}_pF_q[a_1 + n, \dots, a_p + n; b_1 + n, \dots, b_q + n; z].
 \end{aligned}$$

4. For  $r > 0$  we have

$${}_1F_2 \left[ a; a + \frac{b}{2}, a + \frac{b + 1}{2}; -r^2/4 \right] > 0$$

if  $a$  and  $b$  satisfy  $b \geq 2a \geq 0$  or  $b \geq a \geq 1$  or  $0 \leq a \leq 1, b \geq 1$ . The hyper-geometric series cannot be strictly positive if  $0 \leq b < a$  or if  $a = b \in (0, 1)$ .

*Proof.* The first statement is obvious, the third statement can simply be checked by differentiation under the sum. The second statement can be found, for example, in [11, 13.1(56)]. The final statement is a consequence of the findings in [30].  $\square$

Hence, we can continue to compute our Fourier transform. Setting  $\mu := d/2 + \alpha$ ,  $\lambda := \nu$  and  $\beta := (d - 2)/2 + \alpha$ , we see that the second statement of the lemma yields

$$\begin{aligned}
 \mathcal{F}_d \psi_{v,\alpha}(r) &= r^{-d-2\alpha-v} \int_0^r t^{d/2+\alpha} (r-t)^v J_{(d-2)/2+\alpha}(t) dt \\
 &= \frac{\Gamma(v+1)\Gamma(d+2\alpha)}{2^{d/2+\alpha-1}\Gamma(d/2+\alpha)\Gamma(d+2\alpha+v+1)} \\
 &\quad \cdot {}_2F_3 \left[ \frac{d}{2} + \alpha, \frac{d+1}{2} + \alpha; \frac{d}{2} + \alpha, \frac{d+v+1}{2} + \alpha, \frac{d+v+2}{2} + \alpha; -\frac{r^2}{4} \right] \\
 &= C_{v,\alpha}^d {}_1F_2 \left[ \frac{d+1}{2} + \alpha; \frac{d+v+1}{2} + \alpha, \frac{d+v+2}{2} + \alpha; -\frac{r^2}{4} \right].
 \end{aligned}$$

with

$$C_{v,\alpha}^d := \frac{\Gamma(v+1)\Gamma(d+2\alpha)}{2^{d/2+\alpha-1}\Gamma(d/2+\alpha)\Gamma(d+2\alpha+v+1)}. \tag{15}$$

This allows us to make the following general statement on the generalised Wendland functions, which comes from [8].

**Theorem 5.** *Let  $\alpha > 0$  and  $v > -1$  with  $\alpha + v > 0$ . Then, the  $d$ -variate Fourier transform of the generalised Wendland function is given by*

$$\mathcal{F}_d \psi_{v,\alpha}(r) = C_{v,\alpha}^d {}_1F_2 \left[ \frac{d+1}{2} + \alpha; \frac{d+v+1}{2} + \alpha, \frac{d+v+2}{2} + \alpha; -\frac{r^2}{4} \right].$$

Hence,  $\psi_{v,\alpha}$  induces a positive definite function on  $\mathbb{R}^d$  if

$$v \geq \frac{d+1}{2} + \alpha. \tag{16}$$

Furthermore, the Fourier transform satisfies

$$\frac{d}{dr} \mathcal{F}_d \psi_{v,\alpha}(r) = -r \mathcal{F}_d \psi_{v,\alpha+1}(r). \tag{17}$$

*Proof.* Setting  $a = \frac{d+1}{2} + \alpha$  and  $b = v$  shows that (16) is equivalent to  $b \geq a \geq 1$ . Hence, we can conclude from Lemma 3 that  $\psi_{v,\alpha}$  is positive definite. To see (17), we also use Lemma 3, which yields

$$\begin{aligned}
 \frac{d}{dr} \mathcal{F}_d \psi_{v,\alpha}(r) &= C_{v,\alpha}^d \frac{d}{dr} {}_1F_2 \left[ \frac{d+1}{2} + \alpha; \frac{d+v+1}{2} + \alpha, \frac{d+v+2}{2} + \alpha; -\frac{r^2}{4} \right] \\
 &= -\frac{r}{2} C_{v,\alpha}^d \frac{\frac{d+1}{2} + \alpha}{\left(\frac{d+v+1}{2} + \alpha\right) \left(\frac{d+v+2}{2} + \alpha\right)}
 \end{aligned}$$

$$\begin{aligned}
 & \cdot {}_1F_2 \left[ \frac{d+1}{2} + \alpha + 1; \frac{d+v+1}{2} + \alpha + 1, \frac{d+v+2}{2} + \alpha + 1; -\frac{r^2}{4} \right] \\
 &= -r \frac{C_{v,\alpha}^d}{C_{v,\alpha+1}^d} \frac{d+2\alpha+1}{(d+v+2\alpha+1)(d+v+2\alpha+2)} \mathcal{F}_d \psi_{v,\alpha+1}(r) \\
 &= -r \mathcal{F}_d \psi_{v,\alpha+1}(r),
 \end{aligned}$$

where the constant expression in the last but one line can be simplified to 1 using the recurrence formula of the  $\Gamma$ -function.  $\square$

Note that the last property can also be expressed as  $\mathcal{I}_{-1} \mathcal{F}_d \psi_{v,\alpha} = \mathcal{F}_d \psi_{v,\alpha+1}$ . It also has the following interesting consequence.

**Corollary 6.** *The Fourier transform of the generalised Wendland function  $\psi_{v,\alpha}$  is monotonically decreasing if*

$$v \geq \frac{d+1}{2} + \alpha + 1. \tag{18}$$

Since the classical Wendland functions  $\phi_{d,k}$  are a special case of the generalised Wendland functions  $\psi_{v,\alpha}$ , using the parameters  $v = \lfloor d/2 \rfloor + k + 1$  and  $\alpha = k$ , we see that (16) is satisfied, i.e. we have an alternative proof for them being positive definite. However, for  $d \geq n$ , the function  $\phi_{d,k}$  induces not only a positive definite function on  $\mathbb{R}^d$  but also on all  $\mathbb{R}^n$  with  $n \leq d$  and the monotony condition (18) becomes

$$\lfloor d/2 \rfloor + k + 1 \geq \frac{n+1}{2} + k + 1$$

or

$$\lfloor d/2 \rfloor \geq \frac{n+1}{2}.$$

**Corollary 7.** *The Wendland function  $\phi_{d,k}$  induces a positive definite function on  $\mathbb{R}^n$  with a monotonically decreasing Fourier transform for all  $n \leq 2\lfloor d/2 \rfloor - 1$ .*

In [9], there is also a discussion of the decay of the Fourier transform of the generalised Wendland functions. This generalises earlier results from [44].

**Theorem 6.** *The  $d$ -dimensional Fourier transform  $\mathcal{F}_d \psi_{\mu,\alpha}$  of the generalised Wendland function  $\psi_{v,\alpha}$  with  $v \geq \alpha + \frac{d+1}{2}$  and  $\alpha > 0$  satisfies*

$$c_1(1+r^2)^{-\alpha-\frac{d+1}{2}} \leq \mathcal{F}_d \psi_{v,\alpha}(r) \leq c_2(1+r^2)^{-\alpha-\frac{d+1}{2}}, \quad r \in \mathbb{R}$$

with certain constants  $c_1, c_2 > 0$  independent of  $r$ . Hence,  $\psi_{v,\alpha}$  defines a reproducing kernel of

$$H^{\alpha+\frac{d+1}{2}}(\mathbb{R}^d)$$



As mentioned above, this recovers and generalises the decay rate established in [44] for  $\mathcal{F}_d\phi_{d,k}$ , showing that they are reproducing kernels for  $H^\sigma(\mathbb{R}^d)$  with  $\sigma = \frac{d+1}{2} + k$ . While these are integer order Sobolev spaces in odd space dimensions, they are of order integer plus a half in even dimensions. But we are now also in the situation to explicitly state compactly supported RBFs which are reproducing kernels in Sobolev spaces of integer order in even dimensions. We only have to work with  $\alpha = k - 1/2$ ,  $k \in \mathbb{N}$ , and  $\nu \geq k + d/2$  to have a kernel for  $H^{k+d/2}(\mathbb{R}^d)$ . These kernels have a more complicated structure than the original Wendland kernels but are still easily computable. They can best be described by introducing the elementary functions

$$S(r) := \sqrt{1 - r^2}, \quad L(r) := \log\left(\frac{r}{1 + S(r)}\right), \quad r \in (0, 1).$$

Then, some of the functions together with the space dimension  $d$  and the order  $\sigma$  of the Sobolev space in which they are reproducing kernels are, up to a constant factor, given in Table 2. We have always chosen  $\nu = k + d/2 = \sigma$ , since the decay of the Fourier transform and hence the Sobolev space is independent of  $\nu$  as long as  $\nu \geq \alpha + (d + 1)/2 = k + d/2$ .

Yet another consequence is the existence of compactly supported reproducing kernels in Sobolev spaces.

**Corollary 8.** *Each  $H^\sigma(\mathbb{R}^d)$  with  $\sigma > d/2$  possesses a compactly supported, radial reproducing kernel.*

*Proof.* The results of Theorem 6 show that  $H^\sigma(\mathbb{R}^d)$  has such a reproducing kernel, namely  $\psi_{\nu,\alpha}$  as long as  $\sigma = \alpha - \frac{d+1}{2} > \frac{d+1}{2}$ . The case of  $\sigma \in (d/2, (d+1)/2)$  cannot be handled with this technique but follows from another construction technique in [23]. □

**Table 2** The missing Wendland functions.

Space dimension	Function	Sobolev
$d = 2$	$\psi_{2,1/2}(r) \doteq 3r^2L(r) + (2r^2 + 1)S(r)$	$H^2(\mathbb{R}^2)$
	$\psi_{3,3/2}(r) \doteq -15r^4(r^2 + 6)L(r) - (81r^4 + 28r^2 - 4)S(r)$	$H^3(\mathbb{R}^2)$
	$\psi_{4,5/2}(r) \doteq \frac{315r^6(3r^2 + 8)L(r) + (256r^8 + 2639r^6 + 690r^4 - 136r^2 + 16)S(r)}{}$	$H^4(\mathbb{R}^2)$
$d = 4$	$\psi_{3,1/2}(r) \doteq 3r^2(r^2 + 4)L(r) + (13r^2 + 2)S(r)$	$H^3(\mathbb{R}^4)$
	$\psi_{4,3/2}(r) \doteq -105r^4(r^2 + 2)L(r) - (32r^6 + 247r^4 + 40r^2 - 4)S(r)$	$H^4(\mathbb{R}^4)$
	$\psi_{5,5/2}(r) \doteq 315r^6(3r^4 + 60r^2 + 80)L(r) + (9295r^8 + 31670r^6 + 4704r^4 - 688r^2 + 64)S(r)$	$H^5(\mathbb{R}^4)$
$d = 6$	$\psi_{4,1/2}(r) \doteq 15r^2(3r^2 + 4)L(r) + (16r^4 + 83r^2 + 6)S(r)$	$H^4(\mathbb{R}^6)$
	$\psi_{5,3/2}(r) \doteq -105r^4(r^4 + 16r^2 + 16)L(r) - (919r^6 + 2346r^4 + 216r^2 - 16)S(r)$	$H^5(\mathbb{R}^6)$

As a matter of fact, in [23] Johnson constructed compactly supported radial functions  $\phi$  having a  $d$ -variate Fourier transform satisfying (8) for a  $\sigma = k \in \mathbb{N}$  with  $k \geq d/4$  if  $d$  is even and  $k \geq \max((d + 1)/4, 2)$  if  $d$  is odd.

For  $k \in [d/4, d/2]$  for even  $d$  and  $k \in [(d+1)/4, d/2]$  for odd  $d$ , this seems at first to be problematic, since it is well-known that an integrable, continuous function with a non-negative Fourier transform has automatically an integrable Fourier transform (see [46, Corollary 6.12]). The resolution of this seeming contradiction is quite simple, for such values of  $k$  the constructed  $\phi$  in [23] is not continuous on all of  $\mathbb{R}$  any more, it might even have a pole at the origin.

When it comes to the actual computation of the Fourier transform of a generalised Wendland function, it is better to reduce the  $d$ -variate Fourier transform to a univariate one:

$$\begin{aligned} \mathcal{F}_d \psi_{v,\alpha}(r) &= \mathcal{F}_1 \mathcal{I}_{\alpha + \frac{d-1}{2}} \phi_v(r) \\ &= \frac{\sqrt{2/\pi}}{2^{\alpha-1 + \frac{d-1}{2}} \Gamma(\alpha + \frac{d-1}{2})} \int_0^1 \int_s^1 (1-t)^v t(t^2 - s^2)^{\alpha-1 + \frac{d-1}{2}} \cos(rs) dt ds. \end{aligned}$$

Specifying this to the case of  $\phi_{d,k}$  shows that we will naturally have a different Fourier transform for  $d$  even or  $d$  odd. The Fourier transform for odd  $d$  is easily calculated for the classical Wendland functions. As an example, we give the three-dimensional Fourier transform of

$$\phi_{d,1}(r) = \frac{1}{20}(1-r)^4(4r+1).$$

**Lemma 4.** *The Fourier transform of  $\phi_{d,1}$  for  $d = 3$  is given as*

$$\mathcal{F}_3 \phi_{3,1}(r) = -\frac{6\sqrt{2}}{r^8 \sqrt{\pi}} (r^2 - 24) \cos r - 9r \sin r - 4r^2 + 24).$$

Clearly, we see that the Fourier transform decays like  $\mathcal{O}(r^{-6})$ , so that the corresponding Sobolev space is  $H^3(\mathbb{R}^3)$  as predicted. However, note that the numerical evaluation near the origin is highly ill-conditioned because of severe cancellation.

### 3.3 Other Compactly Supported Radial Basis Functions

Besides the functions introduced and discussed in the previous sections, there are plenty of others which can be found in the literature. There is the construction by Buhmann in [5], which produces also “one piece” radial functions, but the decay of their Fourier transform is unknown. There are the earlier constructions by Wu in [50], which are also one piece polynomials but with a higher degree

than those mentioned above, having also a Fourier transform with isolated zeros. Then, there is the construction by Johnson in [2, 23], which produces functions with multiple radial pieces. Each of these pieces is poly-harmonic. The functions are again reproducing kernels in integer order Sobolev spaces.

### 4 Multiscale Interpolation and Approximation

We are now coming to the second main part of this paper, the discussion of multiscale approximation using compactly supported radial basis functions.

Let us first point out why we need a multiscale approach when working with compactly supported radial basis functions. We start with the following negative result (see, for example, [47, 49]).

Recall that the norm-minimal interpolant  $s = s_0 = I_{X, \Phi_\delta}$  to a function  $f \in C(\Omega)$ , which is based upon the data sites  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$  and uses the basis function  $\Phi_\delta : \mathbb{R}^d \rightarrow \mathbb{R}$  can be written as

$$I_{X, \Phi_\delta} f(\mathbf{x}) = \sum_{j=1}^N \alpha_j \Phi_\delta(\mathbf{x} - \mathbf{x}_j), \quad \mathbf{x} \in \Omega.$$

where the coefficients  $\alpha \in \mathbb{R}^N$  are determined by the linear system  $A\alpha = \mathbf{f}$  with the interpolation matrix  $A = (\Phi_\delta(\mathbf{x}_i - \mathbf{x}_j))$ . From now on, we will refer to this norm-minimal interpolant just as *the interpolant*.

**Theorem 7.** *Let  $\Omega \subseteq \mathbb{R}^d$  be a bounded domain with a Lipschitz boundary. Let  $\Phi$  be a reproducing kernel of  $H^\sigma(\mathbb{R}^d)$  with  $\sigma > d/2$ , i.e.  $\widehat{\Phi}$  satisfies (9). Let  $\Phi_\delta = \delta^{-d} \Phi(\cdot/\delta)$ . Finally, let  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$  be given. Then, there is a constant  $C > 0$  such that the error between any  $f \in H^\sigma(\Omega)$  and its interpolant  $s = I_{X, \Phi_\delta} f$  can be bounded by*

$$\|f - I_{X, \Phi_\delta} f\|_{L_2(\Omega)} \leq C \left( \frac{h_{X, \Omega}}{\delta} \right)^\sigma \|f\|_{H^\sigma(\Omega)}. \tag{19}$$

Consequently, if we denote the interpolant to a function  $f$  from our local approximation spaces  $W_j$  from (2) as  $s_j$  and choose the support radii  $\delta_j$  proportional to the fill distances  $h_j = h_{X_j, \Omega}$ , we cannot expect  $s_j$  to converge to  $f$  with  $j \rightarrow \infty$ . We can only expect convergence if  $h_j/\delta_j \rightarrow 0$  for  $j \rightarrow \infty$ , which means that we essentially lose the advantage of the compact support, since the interpolation matrices become more and more dense.

### 4.1 Quasi-Interpolation, Principle Shift-Invariant Spaces and the Strang-Fix Conditions

Even in the more ideal situation of infinite, regular data sites and even if the interpolation process is replaced by a best approximation process the above negative result remains true. This follows from the following short discussion on quasi-interpolation in principle shift-invariant spaces.

For a compactly supported, continuous function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $h > 0$ , we define the space

$$S_h := \text{span} \{ \Phi(h^{-1} \cdot -\alpha) : \alpha \in \mathbb{Z}^d \} = \text{span} \{ \Phi_h(\cdot - h\alpha) : \alpha \in \mathbb{Z}^d \}.$$

Such spaces are called *principal shift-invariant spaces* and they mimic our local approximation spaces in the case of the data sites being a regular grid of grid size  $h$ . These spaces have extensively been studied and their approximation properties are intrinsically connected to polynomial reproduction, i.e. to the question whether  $\pi_m(\mathbb{R}^d) \subseteq S_h$ , or alternatively, to the so-called *Strang-Fix* conditions.

**Definition 10.** An integrable function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfies the Strang-Fix conditions of order  $m$  if  $\widehat{\Phi}(0) \neq 0$  and  $D^\beta \widehat{\Phi}(2\pi\alpha) = 0$  for all  $\alpha \in \mathbb{Z}^d \setminus \{0\}$  and  $\beta \in \mathbb{N}_0^d$  with  $|\beta| \leq m$ .

The following result summarises these ideas. Its proof can be found in [39].

**Theorem 8.** Let  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  be a compactly supported, continuous function. Then, the following statements are equivalent:

1.  $\Phi$  satisfies the Strang-Fix conditions of order  $m$ .
2. For  $\beta \in \mathbb{N}_0^d$  with  $|\beta| \leq m$  the function

$$\mathbf{x} \mapsto \sum_{\alpha \in \mathbb{Z}^d} \alpha^\beta \Phi(\mathbf{x} - \alpha)$$

is a polynomial from  $\pi_{|\alpha|}(\mathbb{R}^d)$ .

3. For each  $f \in H^{m+1}(\mathbb{R}^d)$  and each  $h > 0$  there are weights  $w_\alpha^h$  such that as  $h \rightarrow 0$ ,

$$\left\| f - \sum_{\alpha \in \mathbb{Z}^d} w_\alpha^h \Phi(h^{-1} \cdot -\alpha) \right\|_{H^s(\mathbb{R}^d)} \leq C_s h^{m+1-s} \|f\|_{H^{s+1}(\mathbb{R}^d)}, \quad 0 \leq s \leq p$$

$$\sum_{\alpha \in \mathbb{Z}^d} |w_\alpha^h|^2 \leq K \|f\|_{L_2(\mathbb{R}^d)}^2.$$

The constants  $C_s$  and  $K$  are independent of  $f$ .

Hence, in our terminology, at least if working on an infinite uniform grid, we could get away with just one of the approximation spaces  $S_h = W_{h\mathbb{Z}^d}$  if our compactly supported function was satisfying the Strang-Fix conditions.

However, as we will see now, if the kernel is in addition radial, it will not satisfy the Strang-Fix conditions. To prove this negative result, which goes back to Wu [51], we need two auxiliary results.

The first one is concerned with a question about the density of functions of the form  $\exp(imx)$  in the continuous functions. It was first stated by Pólya as a problem in 1931 ([34]) and then solved by Szász in 1933 ([40]).

**Lemma 5 (Pólya).** *Let the real numbers  $m_1, m_2, \dots$  have the properties  $0 < m_1 < m_2 < \dots$  and*

$$\liminf_{n \rightarrow \infty} \frac{n}{m_n} > \frac{b-a}{2\pi} > 0.$$

*Furthermore, let  $f$  be continuous in the closed interval  $[a, b]$ . Then it will follow from*

$$\int_a^b f(x) \cos(m_n x) dx = \int_a^b f(x) \sin(m_n x) dx = 0$$

*that  $f$  vanishes identically on  $[a, b]$ .*

The second auxiliary result comes from number theory. It deals with the question which natural numbers can be represented as the sum of two squares, see, for example, [24].

**Lemma 6.** *Let  $a_n$  be the  $n$ th natural number which can be expressed as a sum of two integer squares. There are constants  $c_1, c_2 > 0$  and  $n_0 \in \mathbb{N}$  such that*

$$c_1 n < \frac{a_n}{\sqrt{\log(n)}} < c_2 n, \quad n \geq n_0.$$

Using the operators introduced in Section 3.1, we are now able to show that there are no compactly supported radial functions, which satisfy the Strang-Fix conditions in dimensions larger than one.

**Theorem 9.** *For  $d \geq 2$ , there is no non-vanishing, continuous, radial, compactly supported function  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ , which satisfies the Strang-Fix conditions.*

*Proof.* Assume that  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  is such a function, i.e.  $\Phi = \phi(\|\cdot\|_2)$  with an even, compactly supported and continuous function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . Then, its Fourier transform is given by

$$\widehat{\Phi}(\omega) = \mathcal{F}_d \phi(\|\omega\|_2) = \mathcal{F}_1 \mathcal{I}_{\frac{d-1}{2}} \phi(\|\omega\|_2) =: \mathcal{F}_1 \psi(\|\omega\|_2).$$

with a new compactly supported, continuous function  $\psi := \mathcal{I}_{\frac{d-1}{2}}\phi$ . If  $\Phi$  would satisfy the Strang-Fix condition, then we must therefore have

$$0 = \mathcal{F}_1\psi(2\pi\|\alpha\|_2), \quad \alpha \in \mathbb{Z}^d \setminus \{\mathbf{0}\}.$$

Next, let  $a_n$  be the  $n$ th natural number, which can be represented by two squared integers. Since  $d \geq 2$ , for each such  $a_n$  we can therefore pick an  $\alpha_n \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$  such that  $a_n = \|\alpha_n\|_2^2$ . Thus, if we define  $m_n := 2\pi\|\alpha_n\|_2 = 2\pi\sqrt{a_n}$ , then Lemma 6 tells us that

$$\frac{n}{m_n} = \frac{1}{2\pi} \frac{n}{\sqrt{a_n}} > \frac{1}{2\pi\sqrt{c_2}} \frac{n}{\sqrt{n}(\log n)^{1/4}}.$$

This particularly means that  $\liminf_{n \rightarrow \infty} (n/m_n) = \infty$  and hence that according to Lemma 5, we must have that  $\psi = \mathcal{I}_{(d-1)/2}\phi \equiv 0$ . But from (13) we can finally conclude that  $\phi = \mathcal{I}_{-(d-1)/2}\psi = 0$ . □

### 4.2 Multilevel Interpolation

In this section, we want to discuss and analyse the simplest case of a multilevel algorithm, which produces a global approximant from the space  $V_n$ . Let us recall the general setting. We assume that we have a sequence of increasingly finer and finer finite point sets

$$X_1, X_2, \dots, X_n, \dots$$

and a decreasing sequence of support radii

$$\delta_1 \geq \delta_2 \geq \dots \geq \delta_n \geq \dots$$

Then, using a compactly supported RBF  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$  and its scaled versions

$$\Phi_j(\mathbf{x}, \mathbf{y}) := \delta_j^{-d} \Phi((\mathbf{x} - \mathbf{y})/\delta_j), \tag{20}$$

we build, as mentioned in the introduction, local approximation spaces

$$W_j = \text{span}\{\Phi_j(\cdot, \mathbf{x}) : \mathbf{x} \in X_j\}. \tag{21}$$

and global approximation spaces

$$V_n := W_1 + W_2 + \dots + W_n. \tag{22}$$

---

**Algorithm 1:** Multilevel Approximation

---

**Input** : Right-hand side  $f$ , Number of levels  $n$   
**Output** : Approximate solution  $f_n \in W_1 + \dots + W_n$   
 Set  $f_0 = 0, e_0 = f$   
**for**  $j = 1, 2, \dots, n$  **do**  
     Determine a local approximant  $s_j \in W_j$  to  $e_{j-1}$ .  
     Set  $f_j = f_{j-1} + s_j$ .  
     Set  $e_j = e_{j-1} - s_j$ .

---

In this situation, the simplest possible algorithm to compute an approximation  $f_n \in V_n$  to a function  $f \in H$  is a residual correction scheme as described in Algorithm 1.

We will now analyse the approximation properties of this algorithm. To this end, we need a general *sampling inequality*. The following result comes from [32].

**Lemma 7.** *Let  $\Omega \subseteq \mathbb{R}^d$  be a bounded domain with Lipschitz boundary. Let  $\sigma > d/2$ . 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  $f \in H^\sigma(\Omega)$  vanishing on  $X$ , we have*

$$\|f\|_{H^\mu(\Omega)} \leq Ch_{X,\Omega}^{\sigma-\mu} \|f\|_{H^\sigma(\Omega)}.$$

for  $0 \leq \mu \leq \sigma$ .

Using  $s_j = I_{X_j,\Phi_j}e_{j-1}$  as the interpolant to  $e_{j-1}$  on  $X_j$  with kernel  $\Phi_j$ , we have the following theorem.

**Theorem 10.** *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  $\Omega$  with mesh norms  $h_1, h_2, \dots$  satisfying  $h_{j+1} = \mu h_j$  for  $j = 1, 2, \dots$  with fixed  $\mu \in (0, 1)$  and  $h_1 = \mu$  sufficiently small. Let  $\Phi$  be a kernel generating  $H^\sigma(\mathbb{R}^d)$ , i.e. satisfying (9) and let  $\Phi_j$  be defined by (20) with scale factor  $\delta_j = \nu h_j$ . Let the target function  $f$  belong to  $H^\sigma(\Omega)$ . Then, there exists a constant  $C_1 > 0$  such that*

$$\|Ee_j\|_{\Phi_{j+1}} \leq C_1 (\mu^\sigma + \nu^{-\sigma}) \|Ee_{j-1}\|_{\Phi_j} \quad \text{for } j = 1, 2, 3, \dots \tag{23}$$

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

$$\|f - f_n\|_{L_2(\Omega)} \leq C [C_1\mu^\sigma + C_1\nu^{-\sigma}]^n \|f\|_{H^\sigma(\Omega)} \quad \text{for } n = 1, 2, \dots,$$

provided that  $\mu\nu \geq \gamma > 0$  with a constant  $\gamma > 0$ . Thus there are constants  $\mu_0 \in (0, 1)$  and  $\nu_0 > 1$  such that the multiscale approximation  $f_n$  converges linearly to  $f$  in the  $L_2$  norm for all  $\mu \leq \mu_0$  and  $\nu \geq \nu_0$  with  $\mu\nu \geq \gamma$ .

*Proof.* The proof of this theorem can be found in [49]. We will thus only review its critical steps. The first, important auxiliary observation is 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  $X_j \subseteq \Omega$ . Here,  $E$  denotes the extension operator from Proposition 1. From this, it follows that

$$\begin{aligned} \|e_j\|_{H^\sigma(\Omega)} &= \|e_{j-1} - I_{X_j, \delta_j} e_{j-1}\|_{H^\sigma(\Omega)} \\ &= \|Ee_{j-1} - I_{X_j, \delta_j} Ee_{j-1}\|_{H^\sigma(\Omega)} \\ &\leq \|Ee_{j-1} - I_{X_j, \delta_j} Ee_{j-1}\|_{H^\sigma(\mathbb{R}^d)} \\ &\leq C\delta_j^{-\sigma} \|Ee_{j-1} - I_{X_j, \delta_j} Ee_{j-1}\|_{\Phi_j} \\ &\leq C\delta_j^{-\sigma} \|Ee_{j-1}\|_{\Phi_j}, \end{aligned} \tag{24}$$

where we have used Lemma 1 and the fact that the interpolant is norm-minimal with respect to the  $\Phi_j$ -norm.

Then, we have

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

with

$$\begin{aligned} I_1 &:= \int_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}, \\ I_2 &:= \delta_{j+1}^{2\sigma} \int_{\mathbb{R}^d} |\widehat{Ee_j}(\boldsymbol{\omega})|^2 \|\boldsymbol{\omega}\|_2^{2\sigma} d\boldsymbol{\omega}. \end{aligned}$$

Obviously, using Plancharel's theorem, the properties of the extension operator and (19), we see that the first integral can be bounded by

$$\begin{aligned} I_1 &= \|Ee_j\|_{L_2(\mathbb{R}^d)}^2 \leq c \|e_j\|_{L_2(\Omega)}^2 = c \|e_{j-1} - s_j\|_{L_2(\Omega)}^2 \\ &\leq c \left(\frac{h_j}{\delta_j}\right)^{2\sigma} \|e_{j-1}\|_{H^\sigma(\Omega)}^2 \leq c\nu^{-2\sigma} \|Ee_{j-1}\|_{\Phi_j}^2. \end{aligned}$$

Similarly, for the second integral, we can use (24) to derive the bound

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



Piecing these bounds together gives estimate (23). The rest then more or less follows by applying (23) iteratively and the following observation. Since  $e_n = f - f_n$  vanishes on  $X_n$ , we have by Lemma 7 and Lemma 1 that

$$\begin{aligned} \|f - f_n\|_{L_2(\Omega)} &= \|e_n\|_{L_2(\Omega)} \leq Ch_n^\sigma \|e_n\|_{H^\sigma(\Omega)} \leq Ch_n^\sigma \|Ee_n\|_{H^\sigma(\mathbb{R}^d)} \\ &\leq Ch_n^\sigma \delta_{n+1}^{-\sigma} \|Ee_n\|_{\Phi_{n+1}} = C \|Ee_n\|_{\Phi_{n+1}}, \end{aligned} \tag{25}$$

since  $h_n/\delta_{n+1} = h_n/(vh_{n+1}) \leq \frac{1}{v\mu} \leq \frac{1}{\gamma}$ . □

Though we cannot directly determine  $\mu_0$  and  $\nu_0$ , equation (23) gives some insight into the influence of the two critical parameters  $\mu$  and  $\nu$ . On the one hand, the parameter  $\mu$  determines how much we have to refine our data set from level to level. Hence, the smaller  $\mu$  the more points we have to use in the next level.

On the other hand, the parameter  $\nu$  determines the relation between the support radius and the fill distance. Here, a larger  $\nu$  means that we have more non-zero entries per row in the interpolation matrix, which increases the computational cost. Nonetheless, increasing  $\nu$  is less critical than decreasing  $\mu$ .

But there is yet another consequence of this theorem. For simplicity, let us eliminate one of the parameters by setting  $\nu = \mu^{-1}$  so that (23) becomes

$$\|Ee_j\|_{\Phi_{j+1}} \leq C_1 \mu^\sigma \|Ee_{j-1}\|_{\Phi_j} \tag{26}$$

and we have convergence for all  $\mu > 0$  with  $C_1 \mu^\sigma < 1$ . However, we even have convergence if for an arbitrary  $\sigma > \varepsilon > 0$  we have  $C_1 \mu^\varepsilon \leq 1$ . In this case, (26) becomes

$$\|Ee_j\|_{\Phi_{j+1}} \leq \mu^{\sigma-\varepsilon} \|Ee_{j-1}\|_{\Phi_j}. \tag{27}$$

We can also revisit (25) by choosing an  $0 \leq \beta \leq \sigma - \varepsilon$ . Then, Lemma 7 applied in the derivation of (25) yields

$$\|f - f_n\|_{H^\beta(\Omega)} \leq Ch_n^{\sigma-\beta} \|e_n\|_{H^\sigma(\Omega)} \leq Ch_n^{\sigma-\beta} \delta_{n-1}^{-\sigma} \|Ee_n\|_{\Phi_{n+1}} = Ch_n^{-\beta} \|Ee_n\|_{\Phi_{n+1}}.$$

Using (27)  $n$  times yields the estimate

$$\|f - f_n\|_{H^\beta(\Omega)} \leq Ch_n^{-\beta} \mu^{n(\sigma-\varepsilon)} \|f\|_{H^\sigma(\Omega)}.$$

Finally, the fact that  $h_1 = \mu$  and  $h_{j+1} = \mu h_j$  shows that  $h_n = \mu^n$  so that we can rephrase the last estimate as

$$\|f - f_n\|_{H^\beta(\Omega)} \leq Ch_n^{\sigma-\varepsilon-\beta} \|f\|_{H^\sigma(\Omega)},$$

i.e., we have not only derived an estimate also for derivatives but have expressed the error in terms of the fill-distance of the finest data set. The exponent is almost optimal.

**Corollary 9.** *Under the assumptions of Theorem 10 with  $\nu = 1/\mu$  and  $0 \leq \beta \leq \sigma - \varepsilon$  we have the error bound*

$$\|f - f_n\|_{H^\beta(\Omega)} \leq Ch_n^{\sigma - \varepsilon - \beta} \|f\|_{H^\sigma(\Omega)},$$

provided that  $\mu$  is sufficiently small.

There are several possible extensions to this theorem. First of all, the condition  $h_{j+1} = \mu h_j$  can be relaxed to something like  $c\mu h_j \leq h_{j+1} \leq \mu h_j$  with fixed  $\mu, c \in (0, 1)$  without altering the result. Secondly, the algorithm also converges if the target function  $f$  is rougher, say  $f \in H^\tau(\Omega)$  with  $d/2 < \tau < \sigma$ . Details can be found in [49].

From a numerical point of view, the multilevel scheme is extremely efficient. Once, the neighbourhood information are known, i.e. once we know for each level  $\ell$  and each data site  $\mathbf{x}_j^{(\ell)}$  those data sites  $\mathbf{x}_k^{(\ell)}$  which are relevant for the computation, i.e. those with  $\|\mathbf{x}_k^{(\ell)} - \mathbf{x}_j^{(\ell)}\|_2 \leq \delta_\ell$ , we have the following computational cost.

**Corollary 10.** *If the data sets  $X_j$  are quasi-uniform, i.e.  $q_j \approx h_j$ , then we have for the involved linear systems:*

- *The number of non-zero entries per row is independent of the level.*
- *The condition number is independent of the level.*
- *The number of steps required by a non-preconditioned CG method is independent of the level.*
- *The computational cost is linear in each level.*

The neighbourhood information can be assembled in  $\mathcal{O}(N_j \log N_j)$  time using tree-based search structures. Finally, we also have to compute the residuals. If the data sets are nested, we can restrict ourselves to compute residuals only on the finest level. If they are not nested, then we have to do this in step  $j$  for the remaining point sets  $X_{j+1}, \dots, X_n$ . Again, the neighbourhood information can be collected in  $\mathcal{O}(N_j \log N_j)$  time for level  $j$ . Moreover, the number of levels is, because of the uniform refinement, at most  $\mathcal{O}(\log N_n)$ .

Before we come to numerical results, we want to discuss briefly two versions of this algorithm, one which discards unnecessary coefficients and one which only considers necessary data sites.

The first one was introduced in [25] as one of two discarding strategies for multilevel algorithms on the sphere. The general idea is that after computing the local interpolant at level  $j$ , which has a representation of the form

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

to discard all coefficients  $\alpha_k^{(j)}$  which have an absolute value smaller than a given threshold. This threshold can and should be level dependent. Since the discarding

---

**Algorithm 2:** Multilevel Approximation with Dynamical Discarding

---

**Input** : Right-hand side  $f$ , Number of levels  $n$   
**Output** : Approximate solution  $f_n \in V_n = W_1 + \dots + W_n$   
 Set  $\widetilde{f}_0 = 0, \widetilde{e}_0 = f$ ;  
**for**  $j = 1, 2, \dots, n$  **do**  
     Determine the local interpolant  $s_j \in W_j$  to  $\widetilde{e}_{j-1}$ .  
     Drop all coefficients  $|\alpha_k^{(j)}| \leq \text{eps}_j$  in  $s_j$  to define  $\widetilde{s}_j$ .  
     Set  $\widetilde{f}_j = \widetilde{f}_{j-1} + \widetilde{s}_j$ .  
     Set  $e_j = e_{j-1} - \widetilde{s}_j$ .

---

is done during each level, it was named *discarding dynamically* in contrast to the strategy of discarding after all steps have been computed. The Algorithm is formally given in Algorithm 2.

It is possible to show convergence of this algorithm in a very similar way as it has been done in the proof of Theorem 10. Details can be found in [25].

**Theorem 11.** *Let  $\varepsilon > 0$  be given. Assume that the assumption of Theorem 10 are satisfied. Let  $\varepsilon_j \leq c_1 \varepsilon \delta_j^{d/2}$  with a constant  $c_1 > 0$ . Finally, let  $\alpha := C_1(\mu^\sigma + \nu^{-\sigma})$ . Then there is a constant  $C > 0$  such that*

$$\|\widetilde{e}_j\|_{\phi_{j+1}} \leq \alpha \|\widetilde{e}_{j-1}\|_{\phi_j} + C\varepsilon.$$

Hence, after  $n$  steps the error can be bounded by

$$\|f - \widetilde{f}_n\|_{L_2(\Omega)} \leq C\alpha^n \|f\|_{H^\sigma(\Omega)} + C\varepsilon \frac{1 - \alpha^n}{1 - \alpha}.$$

The second variation of our standard multilevel interpolation algorithm is an adaptive version. After computing the local interpolant  $s_j$  on  $X_j$  to  $e_{j-1}$ , we can check the error  $e_j = e_{j-1} - s_j$  on the upcoming data set  $X_{j+1}$ . Then, instead of interpolating this error on all of  $X_{j+1}$ , we will actually only use those points of  $X_{j+1}$  on which  $e_j$  has an absolute value larger than a given threshold  $\varepsilon_j > 0$ . To describe this algorithm in more detail let us denote the initial point sets by  $\widetilde{X}_j$  and let us denote the adaptive point sets which are actually used by  $X_j$ . Then, the algorithm can be stated as in Algorithm 3.

An error analysis of this algorithm is more problematic since it would require to know  $e_j$  on all of  $\widetilde{X}_j$  but we only know  $e_j$  on  $X_j \subseteq \widetilde{X}_j$ . We could avoid this problem by creating an inner loop in which we check  $e_j$  on  $\widetilde{X}_j$  and add those points for which  $e_j$  is still too large. However, in practice this does not seem to be necessary.

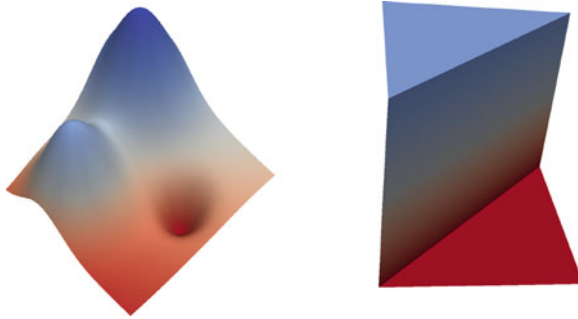
---

**Algorithm 3:** Adaptive Multilevel Approximation
 

---

**Input** : Right-hand side  $f$ , Number of levels  $n$ , thresholds  $\varepsilon_i$ , point sets  $\widetilde{X}_1, \dots, \widetilde{X}_n$   
**Output** : Approximate solution  $f_n \in W_1 + \dots + W_n$   
 Set  $f_0 = 0, e_0 = f, X_1 = \widetilde{X}_1$   
**for**  $j = 1, 2, \dots, n$  **do**  
     Determine  $s_j = I_{X_j, \phi_j} e_{j-1} \in W_j$ .  
     Set  $f_j = f_{j-1} + s_j$ .  
     Set  $e_j = e_{j-1} - s_j$ .  
     **for**  $\mathbf{x} \in \widetilde{X}_{j+1}$  **do**  
         **if**  $|e_j(\mathbf{x})| > \varepsilon_j$  **then**  
              $X_{j+1} = X_{j+1} \cup \{\mathbf{x}\}$

---



**Fig. 1** Franke function (left), step function (right).

### 4.3 Numerical Examples

We will now look at two examples. In the first example, we are interested in understanding the importance of the parameters  $\nu$ , which is responsible for the relation between the support radius and the fill distance on each level, and the parameter  $\mu$ , which is responsible for the refinement of the data sets from level to level.

In this example, we will only be interested in varying the parameter  $\nu$ . We will work on  $\Omega = [0, 1]^2$  and use the Franke function, see, for example, [20] and the left half of Figure 1, as our target function. The RBF is given by  $\phi(r) = (1-r)_+^4(4r+1)$ , i.e. the  $C^2$ -Wendland function in  $\mathbb{R}^2$ . We will work only on equidistant grids  $X_j$  with grid size  $q_j = 2^{-j-2}$ ,  $1 \leq j \leq 8$ , which is also the separation distance of the data set  $X_j$  and hence equivalent to its fill distance. This means that we fix the refinement factor to be  $\mu = 1/2$ . We then vary the overlap factor  $\nu$ , by defining  $\delta_j = \widetilde{\nu} q_j$  and changing  $\widetilde{\nu}$ . Finally, we measure the error on a fine grid of grid size  $q = 2^{-11}$ .

A typical result can be found in Table 3, where we have chosen the overlap factor to be  $\widetilde{\nu} = 3$ , which means that we have at most 25 non-zero entries per row in our matrix, independent of the level. The table contains the number of points per

**Table 3** Approximation of the Franke function. Basis function  $\phi_{2,1} \in C^2, \tilde{\nu} = 3$ .

level	$N$	$l_2$	$l_\infty$	$l_2$ order	$l_\infty$ order	CG
1	81	1.087e-02	7.201e-02			27
2	289	2.464e-03	2.724e-02	2.14	1.40	38
3	1089	5.420e-04	8.947e-03	2.18	1.61	40
4	4225	1.218e-04	2.994e-03	2.15	1.60	40
5	16641	2.775e-05	1.015e-03	2.13	1.56	40
6	66049	6.360e-06	3.466e-04	2.13	1.55	39
7	263169	1.458e-06	1.151e-04	2.12	1.59	39
8	1050625	3.207e-07	3.605e-05	2.19	1.68	39
expected				2.5	1.5	

**Table 4** Approximation of the Franke function. Basis function  $\phi_{2,1} \in C^2$ , various  $\tilde{\nu}$ .

level	$\tilde{\nu} = 5$		$\tilde{\nu} = 7$		$\tilde{\nu} = 9$		$\tilde{\nu} = 11$	
	$l_2$	order	$l_2$	order	$l_2$	order	$l_2$	order
1	6.44e-3		5.45e-3		5.13e-3		5.02e-3	
2	8.33e-4	2.95	4.82e-4	3.50	3.47e-4	3.89	2.81e-4	4.16
3	1.53e-4	2.45	8.12e-5	2.57	5.34e-5	2.70	4.04e-5	2.80
4	2.90e-5	2.40	1.43e-5	2.51	8.89e-6	2.59	6.45e-6	2.65
5	5.56e-6	2.38	2.54e-6	2.49	1.49e-6	2.57	1.04e-6	2.63
6	1.07e-6	2.37	4.53e-7	2.48	2.52e-7	2.57	1.69e-7	2.62
7	2.07e-7	2.37	8.10e-8	2.48	4.26e-8	2.57	2.74e-8	2.63
8	3.82e-8	2.44	1.38e-8	2.55	6.86e-9	2.63	4.23e-9	2.69

level, the discrete  $l_2$ - and  $l_\infty$ -error and the approximation order computed from comparing two successive level. Since  $\phi_{2,1}$  is a reproducing kernel in  $H^\sigma(\mathbb{R}^d)$  with  $\sigma = k + (d + 1)/2 = 2.5$ , which means we would expect an  $L_2$ -approximation order  $\sigma = 2.5$  and an  $L_\infty$ -approximation order of  $\sigma - d/2 = 1.5$ . Finally, the table also contains the number of steps an unpreconditioned CG method requires to compute an approximate solution of the linear system with a relative accuracy of  $10^{-6}$  a higher accuracy does not lead to a significant change in the solutions but only to a larger number CG steps.

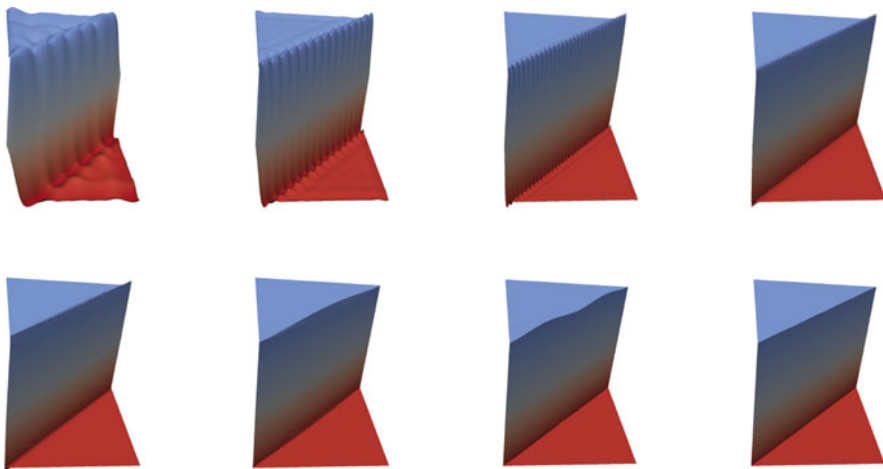
In Table 4 we have collected the error and convergence order estimates for increasing overlaps. It seems that the expected order of 2.5 cannot only be matched but is actually exceeded. The number of non-zero entries per row increases to at most 373 in the case of  $\tilde{\nu} = 11$ . The example also indicates that halving the fill distance, i.e. choosing  $\mu = 1/2$  seems to be an appropriate choice and that the convergence order can be achieved solemnly by increasing the initial support radius.

In our second example, we also want to test the other two algorithms. To this end, we keep the general set-up of our first example but change the test function to a  $C^\infty$  step function given by

$$f(\mathbf{x}) = \frac{1}{2} \tanh(2000(x_2 + x_1 - 1)), \quad \mathbf{x} \in [0, 1]^2, \quad (28)$$

**Table 5** Approximation of the step function (28). Basis function  $\phi_{2,1} \in C^2, \tilde{\nu} = 3$ .

level	$N$	$l_2$	$l_\infty$	$l_2$ order	$l_\infty$ order	CG
1	81	1.329e-01	4.935e-01			19
2	289	9.378e-02	4.869e-01	0.50	0.02	36
3	1089	6.563e-02	4.756e-01	0.51	0.03	41
4	4225	4.520e-02	4.528e-01	0.54	0.07	44
5	16641	3.018e-02	4.203e-01	0.58	0.11	44
6	66049	1.883e-02	3.548e-01	0.68	0.24	44
7	263169	9.965e-03	2.505e-01	0.92	0.50	44
8	1050625	3.398e-03	1.226e-01	1.55	1.03	44
expected				2.5	1.5	



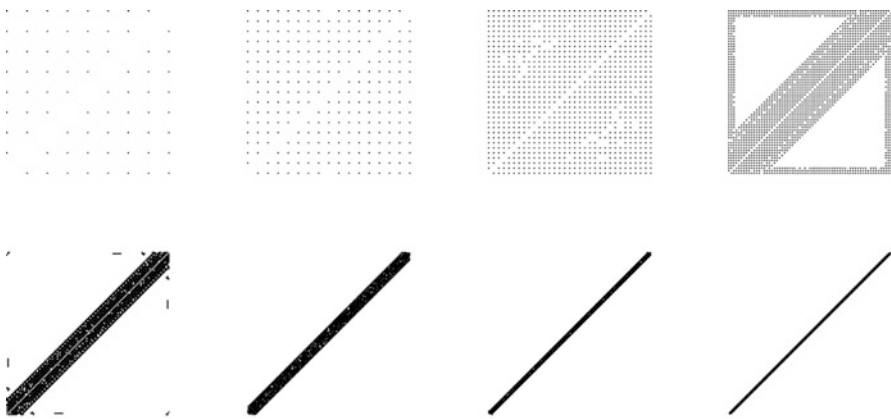
**Fig. 2** Approximation of the step function, levels 0 (upper left) to 7 (lower right).

which is shown in Figure 1 (right). We keep the overlap fixed at  $\tilde{\nu} = 3$ . The results of the standard multilevel algorithm are given in Table 5 and Figure 2. Clearly, convergence is much slower. In particular we have typical overshoots near the edge of the step function, at least in the first few levels.

The result of the dynamically discarding algorithm, Algorithm 2, is given in Table 6. For these results we have set the overall threshold to  $\varepsilon = 10^{-5}$  and have on level  $j$  then discarded all the coefficients with absolute value less than  $\varepsilon \delta_j$ . Clearly, the results are very similar to those of the standard algorithm, Algorithm 1. But the total number of point information used reduces from 1,402,168 in Algorithm 1 to just 31,506 in Algorithm 2, i.e. we use only 2.25% of the original points. As we can see from Figure 3, the reason for this is that only those basis functions centred at data sites near the edge of the step function or near the boundary of the domain have large coefficients.

**Table 6** Approximation of the step function with dynamical discarding. Basis function  $\phi_{2,1} \in C^2, \tilde{\nu} = 3$ .

level	$N$	$l_2$	$l_\infty$	$l_2$ order	$l_\infty$ order	CG
1	72	1.329e-01	4.935e-01			19
2	268	9.378e-02	4.869e-01	0.50	0.02	36
3	1018	6.563e-02	4.756e-01	0.51	0.03	41
4	2080	4.520e-02	4.528e-01	0.54	0.07	44
5	2880	3.018e-02	4.208e-01	0.58	0.11	44
6	4774	1.884e-02	3.548e-01	0.68	0.24	44
7	7134	9.974e-03	2.505e-01	0.92	0.50	44
8	13280	3.453e-03	1.226e-01	1.53	1.03	44
expected				2.5	1.5	

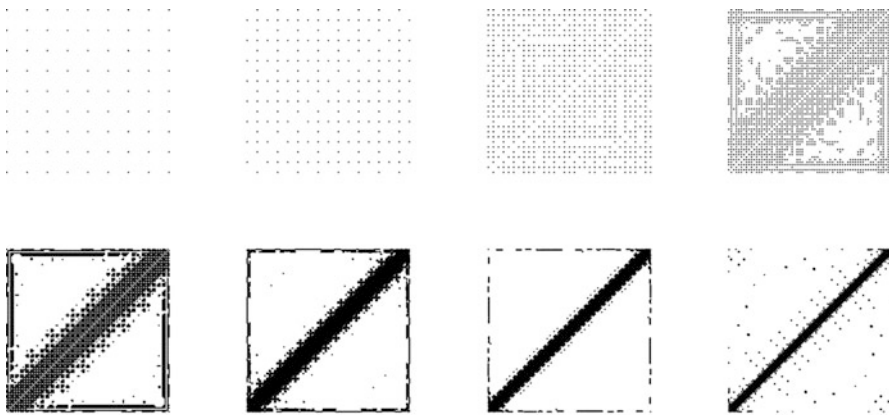


**Fig. 3** Approximation of the step function with dynamical discarding, used data sites for levels 0 (upper left) to 7 (lower right).

The results of the adaptive algorithm are given in Table 7. Here we have used the thresholding strategy that we only considered points on the next level where the error of the current level is larger than  $10^{-3}$  times the maximum error of the current interpolant on the next level. Again, the errors are comparable, though slightly worse. The total number of points used is now 74, 029 out of 1, 402, 168, or 5.3%. Figure 4 shows the data sites which are actually used. The pattern is similar to the one created by the dynamically discarding algorithm, though more points are kept towards the boundary of the computational domain.

**Table 7** Adaptive Approximation of the step function. Basis function  $\phi_{2,1} \in C^2, \tilde{\nu} = 3$ .

level	$N$	$l_2$	$l_\infty$	$l_2$ order	$l_\infty$ order	CG
1	81	1.329e-01	4.935e-01			19
2	200	1.010e-01	4.882e-01	0.40	0.02	26
3	812	6.977e-02	4.786e-01	0.53	0.03	30
4	2318	4.828e-02	4.588e-01	0.53	0.06	31
5	5514	3.249e-02	4.283e-01	0.57	0.10	31
6	10094	2.066e-02	3.699e-01	0.65	0.21	30
7	18480	1.159e-02	2.656e-01	0.83	0.48	30
8	36530	4.934e-03	1.441e-01	1.23	0.88	31
expected				2.5	1.5	



**Fig. 4** Adaptive approximation of the step function, used data sites for levels 0 (upper left) to 7 (lower right).

### 5 Other Multilevel Schemes

So far, we have discussed multilevel and multiscale methods for interpolation problems. It is straight-forward to extend the result to the optimal recovery or smoothing spline approach (4). The results of Theorem 10 and Corollary 9 remain valid if the smoothing parameter  $\varepsilon$  is chosen level dependent satisfying  $\varepsilon_j \leq c(h_j/\delta_j)^{2\sigma}$ , see [49].

In [41], the multilevel scheme is analysed for target functions being zero on the boundary. The centres and the support radii on each level are chosen such that the multilevel interpolant is satisfying zero boundary conditions, as well, i.e. the supports of the basis functions are always contained in the domain  $\Omega$ .

In [13], a multilevel scheme for vector-valued, divergence-free functions is analysed. Here, matrix-valued kernels are employed. In this situation, it is not possible anymore to keep the ratio between fill distance and support radius constant.



The paper [29] follows closely the original papers [26, 27] and discusses convergence orders for interpolation on the sphere, similar to those derived in Corollary 9. It also deals with inverse estimates, i.e. statements that show that a certain convergence rate must lead to a certain smoothness of the target function. These are mainly based upon the results from [36].

Shortly after the introduction of compactly supported RBFs and the initial, purely numerical papers on multilevel interpolation schemes such as [18, 19] and the first theoretical results such as [21, 31], which all differ substantially from the theoretical results given in this paper, also first attempts based upon collocation for solving PDEs numerically with such a multilevel scheme have been derived, see, for example, [14, 16] and [7]. However, only recently, the numerical observations could be underpinned with theoretical proofs in [12, 28]. There have also been first papers on the solution of PDEs based upon a weak formulation and a Galerkin method, see [10, 45]. The first of these two papers is purely numerically, while the second one contains theoretical results, which seem, however, to be too optimistic.

Finally, multiscale schemes like those discussed here have been used in computer graphics (see [33]), in the context of neural networks (see [17]) and in the context of machine learning (see [52]).

## References

1. R.A. Adams, J.J.F. Fournier, *Sobolev Spaces*, 2nd edn. (Academic, New York, 2003)
2. A. Al-Rashdan, M.J. Johnson, Minimal degree univariate piecewise polynomials with prescribed Sobolev regularity. *J. Approx. Theory* **164**, 1–5 (2012)
3. R. Askey, Radial characteristic functions. Technical Report 1262, University of Wisconsin, Mathematics Research Center (1973)
4. S. Brenner, L. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd edn. (Springer, New York, 1994)
5. M.D. Buhmann, Radial functions on compact support. *Proc. Edinb. Math. Soc.* **41**, 33–46 (1998)
6. M.D. Buhmann, *Radial Basis Functions*. Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge, 2003)
7. C.S. Chen, M. Ganesh, M.A. Golberg, A.H.-D. Cheng, Multilevel compact radial functions based computational schemes for some elliptic problems. *Comput. Math. Appl.* **43**, 359–378 (2002)
8. A. Chernih, Multiscale Wendland radial basis functions and applications to solving partial differential equations. PhD thesis, University of New South Wales, Sydney (2013)
9. A. Chernih, S. Hubbert, Closed form representations and properties of the generalised Wendland functions. *J. Approx. Theory* **177**, 17–33 (2014)
10. A. Chernih, Q.T. Le Gia, Multiscale methods with compactly supported radial basis functions for Galerkin approximation of elliptic PDEs. *IMA J. Numer. Anal.* **34**, 569–591 (2014)
11. A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Tables of Integral Transforms*, volume I and II. (McGraw-Hill, New York, 1954)
12. P. Farrell, H. Wendland, RBF multiscale collocation for second order elliptic boundary value problems. *SIAM J. Numer. Anal.* **51**, 2403–2425 (2013)
13. P. Farrell, K. Gillow, H. Wendland, Multilevel interpolation of divergence-free vector fields. *IMA J. Numer. Anal.* **37**, 332–353 (2017)

14. G.E. Fasshauer, Solving differential equations with radial basis functions: multilevel methods and smoothing. *Adv. Comput. Math.* **11**, 139–159 (1999)
15. G. Fasshauer, *Meshfree Approximation Methods with MATLAB* (World Scientific Publishers, Singapore, 2007)
16. G.E. Fasshauer, J.W. Jerome, Multistep approximation algorithms: improved convergence rates through postconditioning with smoothing kernels. *Adv. Comput. Math.* **10**, 1–27 (1999)
17. S. Ferrari, M. Maggioni, N.A. Borhese, Multiscale approximation with hierarchical radial basis functions networks. *IEEE Trans. Neural Netw.* **15**, 178–188 (2004)
18. M.S. Floater, A. Iske, Multistep scattered data interpolation using compactly supported radial basis functions. *J. Comput. Appl. Math.* **73**, 65–78 (1996)
19. M.S. Floater, A. Iske, Thinning and approximation of large sets of scattered data, in *Advanced Topics in Multivariate Approximation*, ed. by F. Fontanella, K. Jetter, P.-J. Laurent (World Scientific, Singapore, 1996), pp. 87–96
20. R. Franke, Scattered data interpolation: tests of some methods. *Math. Comput.* **38**, 181–200 (1982)
21. S.J. Hales, J. Levesley, Error estimates for multilevel approximation using polyharmonic splines. *Numer. Algorithms* **30**, 1–10 (2002)
22. S. Hubbert, Closed form representations for a class of compactly supported radial basis functions. *Adv. Comput. Math.* **36**, 115–136 (2012)
23. M.J. Johnson, Compactly supported, piecewise polyharmonic radial functions with prescribed regularity. *Constr. Approx.* **35**, 201–223 (2012)
24. H. Koch, H. Pieper, *Zahlentheorie: ausgewählte Methoden und Ergebnisse*. VEB Deutscher Verlag der Wissenschaften, Berlin (1976)
25. Q.T. Le Gia, H. Wendland, Data compression on the sphere using multiscale radial basis functions. *Adv. Comput. Math.* **40**, 923–943 (2014)
26. Q.T. Le Gia, I. Sloan, H. Wendland, Multiscale analysis in Sobolev spaces on the sphere. *SIAM J. Numer. Anal.* **48**, 2065–2090 (2010)
27. Q.T. Le Gia, I. Sloan, H. Wendland, Multiscale approximation for functions in arbitrary Sobolev spaces by scaled radial basis functions on the unit sphere. *Appl. Comput. Harmon. Anal.* **32**, 401–412 (2012)
28. Q.T. Le Gia, I. Sloan, H. Wendland, Multiscale RBF collocation for solving PDEs on spheres. *Numer. Math.* **121**, 99–125 (2012)
29. M. Li, F. Cao, Multiscale interpolation on the sphere: Convergence rate and inverse theorem. *Appl. Math Comput.* **263**, 134–150 (2015)
30. D.S. Moak, Completely monotonic functions of the form  $s^{-b}(s^2 + 1)^{-a}$ . *Rocky Mt. J. Math.* **17**, 719–725 (1987)
31. F.J. Narcowich, N. Sivakumar, J.D. Ward, On condition numbers associated with radial-function interpolation. *J. Math. Anal. Appl.* **186**, 457–485 (1994)
32. F.J. Narcowich, J.D. Ward, H. Wendland, Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constr. Approx.* **24**, 175–186 (2006)
33. Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, H.-P. Seidel, Multi-level partition of unity implicits. *ACM Trans. Graphics* **22**, 463–470 (2003)
34. G. Pólya, Aufgabe 108: Ein Bruchteil des Fourierschen Funktionensystems. Jahresbericht der deutschen Mathematiker Vereinigung, **40**, 81 (1931)
35. R. Schaback, The missing Wendland functions. *Adv. Comput. Math.* **34**, 67–81 (2011)
36. R. Schaback, H. Wendland, Inverse and saturation theorems for radial basis function interpolation. *Math. Comput.* **71**, 669–681 (2002)
37. R. Schaback, Z. Wu, Operators on radial functions. *J. Comput. Appl. Math.* **73**, 257–270 (1996)
38. E.M. Stein, G. Weiss, *Fourier Analysis in Euclidean Spaces* (Princeton University Press, Princeton, NJ, 1971)
39. G. Strang, G. Fix, A Fourier analysis of the finite element variational method, in *Constructive Aspects of Functional Analysis*, ed. by G. Geymonat. C.I.M.E. II, Ciclo 1971 (Springer, Berlin, 1973), pp. 793–840

40. O. Szász, Lösung zu Aufgabe 108: Ein Bruchteil des Fourierschen Funktionensystems. Jahresbericht der deutschen Mathematiker Vereinigung **43**Part 2, 20–23 (1933)
41. A. Townsend, H. Wendland, Multiscale analysis in sobolev spaces on bounded domains with zero boundary values. IMA J. Numer. Anal. **33**, 1095–1114 (2013)
42. G. Wahba, *Spline Models for Observational Data*. CBMS-NSF, Regional Conference Series in Applied Mathematics (SIAM, Philadelphia, 1990)
43. H. Wendland, Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. **4**, 389–396 (1995)
44. H. Wendland, Error estimates for interpolation by compactly supported radial basis functions of minimal degree. J. Approx. Theory **93**, 258–272 (1998)
45. H. Wendland, Numerical solutions of variational problems by radial basis functions. in *Approximation Theory IX, Volume 2: Computational Aspects*, ed. by C.K. Chui, L.L. Schumaker (Vanderbilt University Press, Nashville, 1998), pp. 361–368
46. H. Wendland, *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge, 2005)
47. H. Wendland, Computational aspects of radial basis function approximation, in *Topics in Multivariate Approximation and Interpolation*, ed. by K. Jetter, M.D. Buhmann, W. Haussmann, R. Schaback, J. Stöckler. Studies in Computational Mathematics, vol. 12 (Elsevier, Amsterdam, 2006), pp. 231–256
48. H. Wendland, Divergence-free kernel methods for approximating the Stokes problem. SIAM J. Numer. Anal. **47**, 3158–3179 (2009)
49. H. Wendland, Multiscale analysis in Sobolev spaces on bounded domains. Numer. Math. **116**, 493–517 (2010)
50. Z. Wu, Compactly supported positive definite radial functions. Adv. Comput. Math. **4**, 283–292 (1995)
51. Z. Wu, Compactly supported radial functions and the Strang-Fix condition. Appl. Math. Comput. **84**, 115–124 (1997)
52. B. Xu, S. Lu, M. Zhong, Multiscale support vector regression method in Sobolev spaces on bounded domains. Appl. Anal. **94**, 548–569 (2015)