

A Data Driven Approach to Multiscale Approximation with Anisotropically Scaled Radial Basis Functions

R. Kempf ¹ and H. Wendland ^{1,*}

¹*Applied and Numerical Analysis, Department of Mathematics, University of Bayreuth*

Received: 16/07/2025 – Published: 24/02/2026

Communicated by: A. Sommariva

Dedicated to Prof. Ed Kansa

Abstract

We study elliptic basis functions, which are a generalization of the class of well-known radial basis functions. We derive a purely data-driven way to determine the scaling matrix by using the principle component analysis method and combine the elliptic basis functions with a multilevel approximation scheme. We derive some results on the native spaces of such functions and the stability of the interpolation process. We give some numerical examples, showing that multilevel methods using elliptic basis functions can be superior to standard multilevel methods.

Keywords: Kernel-based learning, anisotropic basis functions, elliptic support, multilevel approximation (MSC2020: 65D05, 65D12, 65D15)

1 Introduction

Kernel methods have become more and more prominent in a wide field of applications such as scattered data approximation [7, 24, 25, 31], solving partial differential equations [9, 10, 14, 15, 17, 18, 33] and machine learning [12, 26, 27], where the list of applications and citations is by no means exhaustive.

In approximation, most often one chooses *radial basis functions (RBFs)*, i.e., kernels whose values only depend on the distance of their arguments. Popular examples of these kinds of kernels are Gaussians, (inverse) multiquadrics or Matérn functions [21] who all have in common that they are globally supported. There is also the option of using compactly supported Wendland functions [30]. All these choices have in common that they allow an easy construction

* Corresponding author: Holger.Wendland@uni-bayreuth.de

of approximation spaces in any dimension and of arbitrary smoothness. Although these discretizations exhibit excellent approximation properties, it is well-known that they lead to poor conditioning and a high computational cost if the number of data grows [23].

One way to overcome this trade-off principle is to scale the chosen basis function. One possibility is to choose a different scaling parameter for every point, see, e.g., [4], which somewhat reflects Kansa's ideas of unsymmetric collocation or approximation. However, this leads to unsymmetric matrices, which might even become singular, see [13].

Another approach [8, 22] follows a multilevel residual correction scheme. First, a coarse set and a large support radius are chosen in order to capture the large-scale variations of the target function. This implies that the residual between the target function and the first interpolant lives on a finer scale. In the second step of the scheme, one then chooses a finer data set and a smaller support radius and interpolates the residual from the first level, capturing the finer details. The sum of both interpolants then interpolates the target function at the fine data set. This procedure can be repeated multiple times, if necessary. This method is well-studied [19, 20, 32] and the list of applications grows, see, e.g., [2, 6, 16, 29].

All cited articles have in common that isotropically scaled RBFs are used, i.e., the level lines of the functions are still hyper-circles. However, this may not be the best choice for the given data. If the data, or domain, exhibits anisotropy it may be better to scale the RBF accordingly, for example, one would "compress" the support of the basis function in the direction of a larger spread and would "expand" the support in a direction with less spread, leading to level curves being hyper-ellipses. There are already several publications using such *elliptic basis functions (EBFs)*, see for example [1, 3, 5], and employing EBFs in a multiscale setting [11, 28]. We follow these ideas and study a data driven way to construct EBFs by performing a *principle component analysis* on the given sets of data in order to find the scaling and rotation parameters.

The paper is organized as follows. In [Section 2](#) we briefly introduce the kernel-based multilevel method in a general setting. Then we study elliptic basis functions and derive some important properties such as positive definiteness and a relation of the native space of an EBF to a classical Sobolev space. We also derive lower bounds on the smallest eigenvalue of the interpolation matrix, which is known to dominate the condition number. The latter two points are direct generalizations of known results for isotropically scaled basis functions. In [Section 2.3](#), we introduce our way for determining the scaling matrix A and finally in [Section 2.4](#) we collect these ideas to derive the *EBF multilevel method*. We justify the new ideas by numerical examples in [Section 3](#).

2 Elliptic Basis Function Multilevel

2.1 Kernel-based Multilevel Method

Let $\Omega \subseteq \mathbb{R}^n$ be a bounded domain and let $X_1, X_2, \dots \subseteq \Omega$ be a family of sets of sites with N_j elements, i.e., we can write $X_j = \{\mathbf{x}_{j,1}, \dots, \mathbf{x}_{j,N_j}\}$. The index j denotes the level to which the set is associated with. In general, these sets need not to be nested, however, we assume that $N_{j+1}/N_j \sim \mu^{-n} \in (0, 1)$, independent of the level.

Next, we fix a kernel $K_j : \Omega \times \Omega \rightarrow \mathbb{R}$ for each level j . Most often, this will be done by choosing one compactly supported RBF $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ with support $B_1(\mathbf{0})$ and scaling it level-dependently, with a factor $\delta_j > 0$, yielding kernel functions $\Phi_j := \Phi(\cdot/\delta_j)$ and $K_j(\mathbf{x}, \mathbf{y}) := \Phi_j(\mathbf{x} - \mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \Omega$.

The kernel-based multilevel method is now a simple residual correction scheme. Defining detail spaces $W_j := \text{span}\{K_j(\cdot, \mathbf{x}_{j,i}) : \mathbf{x}_{j,i} \in X_j\}$ and (global) approximation spaces $V_L = W_1 +$

$\dots + W_L$ we follow the easy **Algorithm 1** to compute the approximation $f_L \in V_L$ to a function $f \in C(\Omega)$.

Data: Right-hand side f , number of levels L

Result: Approximate solution f_L

Set $f_0 = 0, e_0 = f$;

for $j = 1, 2, \dots, L$ **do**

Compute $s_j = \sum_{i=1}^{N_j} \alpha_{j,i} K_j(\cdot, \mathbf{x}_{j,i})$ by solving the interpolation problem

$s_j|_{X_j} = e_{j-1}|_{X_j}$;

Set $f_j = f_{j-1} + s_j$;

Set $e_j = e_{j-1} - s_j$;

end

Algorithm 1: Multilevel approximation

2.2 Elliptic Basis Functions

In standard kernel-based methods, particularly those employing radial basis functions, the kernel is typically isotropic and scaled by a single scalar shape parameter. That is, for a radial basis function $\phi : [0, \infty) \rightarrow \mathbb{R}$, the kernel centered at a point $\mathbf{x}_i \in \mathbb{R}^n$ is expressed as

$$K_\delta(\mathbf{x}, \mathbf{x}_i) = \phi\left(\frac{\|\mathbf{x} - \mathbf{x}_i\|_2}{\delta}\right),$$

where $\delta > 0$ controls the spatial extent of the scaled basis function. While effective in many scenarios, this isotropic scaling becomes a significant limitation when the underlying data or solution exhibits anisotropic features. In such cases, using a scalar scaling factor can lead to poor approximation quality, inefficient representations, or ill-conditioned interpolation systems.

To address this limitation, we propose an alternative scaling approach in which the radial kernel is transformed using a symmetric positive definite matrix, though, in many situations an invertible matrix A is sufficient. To avoid problems with well-definedness, we assume from now on that our kernel K is either *translation invariant*, i.e. it is of the form $K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y})$ with a given $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ or even *radial*, i.e. it is of the form $K(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|_2)$ with a given $\phi : [0, \infty) \rightarrow \mathbb{R}$. Then, we define the *elliptic kernel* K_A by

$$K_A(\mathbf{x}, \mathbf{x}_i) = \Phi(A(\mathbf{x} - \mathbf{x}_i)) = \phi(\|A(\mathbf{x} - \mathbf{x}_i)\|_2),$$

where $A \in \mathbb{R}^{n \times n}$ is a symmetric and positive definite matrix that governs the shape and orientation of the kernel's support. The norm $\|A(\mathbf{x} - \mathbf{x}_i)\|_2$ defines an ellipsoidal level set centered at \mathbf{x}_i , in contrast to the spherical level set defined by the standard Euclidean norm. This form allows the kernel to stretch or contract along preferred directions, enabling better alignment with anisotropic structures in the data.

The transformation $\mathbf{x} \mapsto A(\mathbf{x} - \mathbf{x}_i)$ induces an ellipsoidal distortion of the Euclidean space. The level sets of the kernel become ellipsoids centered at \mathbf{x}_i , with principal axes determined by the eigenvectors of $A^T A$, and axis lengths inversely proportional to the square roots of the corresponding eigenvalues. This anisotropic shape enables the kernel to better capture directional trends or structures in the data, for example, elongated features, gradients, or interfaces aligned along specific directions.

In particular, if $A = \frac{1}{8}I$, we recover the standard isotropic case. However, in the more general setting, A is constructed to encode directional information, with each direction scaled according to its importance or variance in the data. This idea forms the basis for the data-driven scaling approach we introduce in [Section 2.3](#).

We find the following important properties of elliptic kernels. We can formulate and prove the result even for invertible A .

Proposition 2.1. *Let $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be positive definite on \mathbb{R}^n . Then $\Phi_A := \Phi(A \cdot)$ defines a positive definite kernel $K_A : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ for any invertible $A \in \mathbb{R}^{n \times n}$.*

Proof. We have to check if the quadratic form

$$\sum_{i,j=1}^N \alpha_i \alpha_j \Phi_A(\mathbf{x}_i - \mathbf{x}_j)$$

is positive for any pairwise disjoint $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \mathbb{R}^n$ and $\boldsymbol{\alpha} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$. Since A is invertible, the transformed points $\tilde{\mathbf{x}}_j := A\mathbf{x}_j$, $1 \leq j \leq N$, are also pairwise disjoint. Thus, with the assumption that Φ is positive definite, we have

$$\begin{aligned} 0 &< \sum_{i,j=1}^N \alpha_i \alpha_j \Phi(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j) = \sum_{i,j=1}^N \alpha_i \alpha_j \phi(\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2) \\ &= \sum_{i,j=1}^N \alpha_i \alpha_j \phi(\|A(\mathbf{x}_i - \mathbf{x}_j)\|_2) = \sum_{i,j=1}^N \alpha_i \alpha_j \Phi_A(\mathbf{x}_i - \mathbf{x}_j). \end{aligned}$$

□

It is well-known that positive definite functions give rise to special reproducing kernel Hilbert spaces, the *native space* \mathcal{N}_Φ of Φ . The native space of the elliptic kernel Φ_A can be understood as the pull-back of the original native space under A , i.e.,

$$f \in \mathcal{N}_{\Phi_A} \quad \text{if and only if} \quad f \circ A^{-1} \in \mathcal{N}_\Phi.$$

Consequently, for the native space norm holds

$$\|f\|_{\mathcal{N}_{\Phi_A}} = \|f \circ A^{-1}\|_{\mathcal{N}_\Phi}.$$

These facts have already been investigated in [\[3\]](#).

In applications, kernels whose native space is norm-equivalent to Sobolev spaces are of particular interest. We have the following new result.

Theorem 2.2. *Let Φ be a reproducing kernel of $H^s(\mathbb{R}^n)$, $s > n/2$, i.e., its Fourier transform satisfies*

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

with constants $c_1, c_2 > 0$. Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite, i.e., there is an orthonormal matrix Q and a diagonal matrix $D = \text{diag}(d_1, \dots, d_n)$ such that $A = QDQ^T$. Assume that $d_i \geq 1$, $1 \leq i \leq n$. Then the norm of the native space of Φ_A is equivalent to $\|\cdot\|_{H^s(\mathbb{R}^n)}$, in particular,

$$\left(\frac{c_1}{\prod_{i=1}^n d_i}\right)^{1/2} \|g\|_{\mathcal{N}_{\Phi_A}} \leq \|g\|_{H^s(\mathbb{R}^n)} \leq \left(\frac{c_2}{\prod_{i=1}^n d_i}\right)^{1/2} \left(\max_{1 \leq i \leq n} d_i\right)^s \|g\|_{\mathcal{N}_{\Phi_A}} \quad (2)$$

holds for all $g \in \mathcal{N}_{\Phi_A}$.

Proof. The Fourier transform of Φ_A is given by $\widehat{\Phi}_A = |\det A|^{-1} \widehat{\Phi}(A^{-1}\cdot)$. Using this in (1) yields immediately that $\widehat{\Phi}_A$ satisfies

$$\frac{|\det A|}{c_2} (1 + \|A^{-1} \boldsymbol{\omega}\|_2^2)^s \leq \frac{1}{\widehat{\Phi}_A(\boldsymbol{\omega})} = \frac{|\det A|}{\widehat{\Phi}(A^{-1} \boldsymbol{\omega})} \leq \frac{|\det A|}{c_1} (1 + \|A^{-1} \boldsymbol{\omega}\|_2^2)^s. \quad (3)$$

Next, using the decomposition $A = QDQ^T$ with diagonal D and orthonormal Q , we have, setting $\boldsymbol{\eta} := Q^T \boldsymbol{\omega}$,

$$\|A^{-1} \boldsymbol{\omega}\|_2^2 = \|D^{-1} Q^T \boldsymbol{\omega}\|_2^2 = \|D^{-1} \boldsymbol{\eta}\|_2^2 = \sum_{i=1}^n \left(\frac{\eta_i}{d_i} \right)^2,$$

which yields the two bounds

$$\begin{aligned} \|A^{-1} \boldsymbol{\omega}\|_2^2 &\leq \frac{1}{d_{\min}^2} \|Q^T \boldsymbol{\omega}\|_2^2 = \frac{1}{d_{\min}^2} \|\boldsymbol{\omega}\|_2^2 \leq \|\boldsymbol{\omega}\|_2^2, \\ \|A^{-1} \boldsymbol{\omega}\|_2^2 &\geq \frac{1}{d_{\max}^2} \|Q^T \boldsymbol{\omega}\|_2^2 = \frac{1}{d_{\max}^2} \|\boldsymbol{\omega}\|_2^2, \end{aligned}$$

where d_{\min} and d_{\max} denote the minimum and maximum value of the d_i . Inserting this into (3) yields on the one hand the bound

$$\frac{1}{\widehat{\Phi}_A(\boldsymbol{\omega})} \leq \frac{\prod_{i=1}^n d_i}{c_1} (1 + \|\boldsymbol{\omega}\|_2^2)^s$$

and thus

$$\|g\|_{\mathcal{N}_{\Phi_A}} \leq \left(\frac{\prod_{i=1}^n d_i}{c_1} \right)^{1/2} \|g\|_{H^s(\mathbb{R}^n)},$$

which is the stated lower bound on the Sobolev norm. On the other hand, we find

$$\begin{aligned} \frac{1}{\widehat{\Phi}_A(\boldsymbol{\omega})} &\geq \frac{\prod_{i=1}^n d_i}{c_2} \left(1 + \frac{1}{d_{\max}^2} \|\boldsymbol{\omega}\|_2^2 \right)^s = \frac{\prod_{i=1}^n d_i}{c_2} \frac{1}{d_{\max}^{2s}} (d_{\max}^2 + \|\boldsymbol{\omega}\|_2^2)^s \\ &\geq \frac{\prod_{i=1}^n d_i}{c_2} \frac{1}{d_{\max}^{2s}} (1 + \|\boldsymbol{\omega}\|_2^2)^s, \end{aligned}$$

using again that all the eigenvalues are assumed to be at least one. This then leads to

$$\|g\|_{\mathcal{N}_{\Phi_A}} \geq \left(\frac{\prod_{i=1}^n d_i}{c_2} \frac{1}{d_{\max}^{2s}} \right)^{1/2} \|g\|_{H^s(\mathbb{R}^n)},$$

which yields the stated upper bound. □

Remark 2.3. If $D = \text{diag}(\frac{1}{\delta}, \dots, \frac{1}{\delta})$, $\delta \leq 1$, we recover the standard norm equivalence for the scaled RBF $\Phi_\delta = \Phi(\cdot/\delta)$, cf. [32, Lemma 1].

It is fairly easy to see how the *scaling matrix* A influences the condition number of the interpolation matrix. We discuss this by looking only at the smallest eigenvalue, as it is this eigenvalue that dominates the condition number. Noticing once again that the interpolation matrix $\Phi_A(\mathbf{x}_i - \mathbf{x}_j)$ is nothing but the standard interpolation matrix $\Phi(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)$ in the data sites $\tilde{\mathbf{x}}_i = A\mathbf{x}_i$, $1 \leq i \leq N$, we can employ the standard theory on lower bounds for the smallest

eigenvalue. To be more precise, [31, Theorem 12.3] gives for a Sobolev kernel, i.e. a kernel that satisfies (1), the lower bound

$$\lambda_{\min}(\Phi_A(\mathbf{x}_i - \mathbf{x}_j)) = \lambda_{\min}(\Phi(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)) \geq Cq_{\tilde{X}}^{2s-n}$$

using the *separation radius* $q_{\tilde{X}} = \frac{1}{2} \min_{i \neq j} \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2$. As in the proof of Theorem 2.2, we can proceed to express this separation radius using the matrix using the decomposition $A = QDQ^T$ of the matrix A . As a matter of fact, defining $\mathbf{y}_j = Q^T \mathbf{x}_j$, we have

$$\|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_j\|_2^2 = \|\mathbf{D}\mathbf{y}_i - \mathbf{D}\mathbf{y}_j\|_2^2 \geq d_{\min}^2 \|\mathbf{x}_i - \mathbf{x}_j\|_2^2,$$

showing

$$\lambda_{\min}(\Phi_A(\mathbf{x}_i - \mathbf{x}_j)) \geq Cd_{\min}^{2s-d} q_X^{2s-d},$$

which is an improvement over the lower bound of the smallest eigenvalue of the unmodified interpolation matrix $(\Phi(\mathbf{x}_i - \mathbf{x}_j))$, as long as we have $d_{\min} > 1$. This result can slightly be further improved by employing some of the estimates derived in the proof of Theorem 2.2.

Corollary 2.4. *Let Φ be a reproducing kernel of $H^s(\mathbb{R}^n)$, $s > n/2$, i.e., its Fourier transform satisfies (1). Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite, i.e., there is a decomposition $A = QDQ^T$ with an orthonormal matrix Q and a diagonal matrix $D = \text{diag}(d_1, \dots, d_n)$. Assume that $d_i \geq 1$, $1 \leq i \leq n$. Given a set $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ of pairwise distinct data sites, the smallest eigenvalue of the interpolation matrix satisfies the lower bound*

$$\lambda_{\min}(\Phi_A(\mathbf{x}_i - \mathbf{x}_j)) \geq C \frac{(\min_{1 \leq i \leq n} d_i)^{2s}}{\prod_{i=1}^n d_i} q_X^{2s-d}.$$

Proof. From the proof of Theorem 2.2 we have the lower bound

$$\widehat{\Phi}_A(\boldsymbol{\omega}) \geq \frac{c_1}{\prod_{i=1}^n d_i} \left(1 + \frac{1}{d_{\min}^2} \|\boldsymbol{\omega}\|_2^2\right)^{-s}.$$

Using this lower bound in [31, Theorem 12.3], yields

$$\lambda_{\min}(\Phi_A(\mathbf{x}_i - \mathbf{x}_j)) \geq \frac{C}{\prod_{i=1}^n d_i} (1 + d_{\min}^{-2} q_X^{-2})^{-s} = C \frac{d_{\min}^{2s}}{\prod_{i=1}^n d_i} q_X^{2s-d},$$

which is the stated bound. □

Remark 2.5. *The assumption $d_i \geq 1$ in Theorem 2.2 and Theorem 2.4 corresponds to a similar assumption $\delta \leq 1$ in the isotropic case $\Phi(\cdot/\delta)$. It is not a restriction, as we can always scale the original basis function Φ or the matrix A or, equivalently, the data sites such that $d_i \geq 1$ is satisfied.*

2.3 PCA-Based Determination of Scaling Matrices

To construct anisotropic kernels that adapt to the local geometry of the data, we require a method to determine appropriate scaling matrices that capture the dominant directions and scales present in the point cloud or function to be approximated. Principal Component Analysis (PCA) provides a natural and efficient framework for this purpose.

PCA is a classical technique in statistical data analysis that identifies the directions of maximal variance in a dataset. For a given local neighborhood of data points, PCA computes a set

of orthogonal axes (principal components) along which the data is most spread out. These axes are determined by the eigenvectors of the sample covariance matrix, and the corresponding eigenvalues quantify the variance along each direction.

Consider a point $\mathbf{x}_i \in X$. We define a local neighborhood $\mathcal{U}_i \subset X$ about \mathbf{x}_i , typically by including the k nearest neighbors or all points within a given radius. The sample mean of the neighborhood is given by

$$\bar{\mathbf{x}}_i = \frac{1}{|\mathcal{U}_i|} \sum_{\mathbf{x}_m \in \mathcal{U}_i} \mathbf{x}_m,$$

and the local (empirical) covariance matrix is

$$C_i = \frac{1}{|\mathcal{U}_i|} \sum_{\mathbf{x}_m \in \mathcal{U}_i} (\mathbf{x}_m - \bar{\mathbf{x}}_i)(\mathbf{x}_m - \bar{\mathbf{x}}_i)^\top.$$

The matrix $C_i \in \mathbb{R}^{n \times n}$ is symmetric and positive semi-definite. Its eigendecomposition

$$C_i = Q_i \Lambda_i Q_i^\top,$$

provides a matrix Q_i of orthonormal eigenvectors and a diagonal matrix $\Lambda_i = \text{diag}(\lambda_{i,1}, \dots, \lambda_{i,d})$ of non-negative eigenvalues. The eigenvectors represent the local principal directions, and the eigenvalues quantify the local directional variances.

We use this information to construct a local scaling matrix A_i for the kernel centered at \mathbf{x}_i . Specifically, we define

$$A_i := Q_i D_i Q_i^\top, \tag{4}$$

where $D_i = \nu \cdot \text{diag}(d_{i,1}, \dots, d_{i,n})$, where $d_{i,k} = \lambda_{i,k} / \min_{1 \leq j \leq n} \lambda_{i,j}$ are the normalized eigenvalues of C_i and $\nu > 0$ is a freely choosable scaling parameter. This construction ensures that the scaled kernel

$$\Phi_{A_i}(\mathbf{x}, \mathbf{x}_i) = \phi(\|A_i(\mathbf{x} - \mathbf{x}_i)\|_2)$$

has elliptic level sets aligned with the local data geometry: elongated in directions of low variance and compressed in directions of high variance. Intuitively, this allows the kernel to “stretch out” along sparse directions and concentrate more closely where the data is dense or exhibits strong local variation.

In practice, the parameter ν can be chosen empirically or based on a bound on the minimum variance deemed meaningful. The role of the parameter ν is similar to its role in the isotropic case, where it defines the relation between the fill distance and the support radius. However, due to the presence of the matrix A , its influence is less obvious. Hence, so far, we only employed heuristics from the isotropic case, which means that a parameter ν too large leads to better approximation results but worse stability, conditioning and runtime, while a parameter too small has the exact opposite effect. More research in the anisotropic situation is necessary. Moreover, the eigenvalues can be truncated after a few dominant components, resulting in a low-rank approximation of the local geometry, especially useful in high-dimensional settings.

For multilevel methods, a separate PCA-based matrix $A_{j,i}$ can be constructed at each level j and at each node $\mathbf{x}_{j,i} \in X_j$. This allows the kernels to be locally and level-wise adapted to anisotropic structures, which is particularly advantageous in problems with directional features that vary across scales, such as those arising in PDEs with anisotropic coefficients or interpolation of elongated manifolds.

This PCA-driven scaling strategy integrates naturally with the matrix-scaled kernel framework and leads to enhanced approximation performance in anisotropic settings, as we demonstrate in the numerical experiments in [Section 3](#).

In [Fig. 1](#) we show a comparison between isotropic and PCA-based anisotropic scaled kernel supports. We generated a two-dimensional point cloud centered at $(0,0)$ with 100 data points, drawn from a Gaussian distribution. The data points were then transformed to introduce anisotropy: scaling by a factor of 3 along the first principal axis and 1 along the second, followed by a 30° rotation. We performed the PCA as described above by using all 100 points. The mean \bar{x} is indicated in black. We also give the unit circle in blue and the ellipse determined by the PCA in red, centered about this mean. The figure illustrates how PCA correctly identifies the principal components of the data, adjusting the orientation and scaling of the ellipse accordingly.

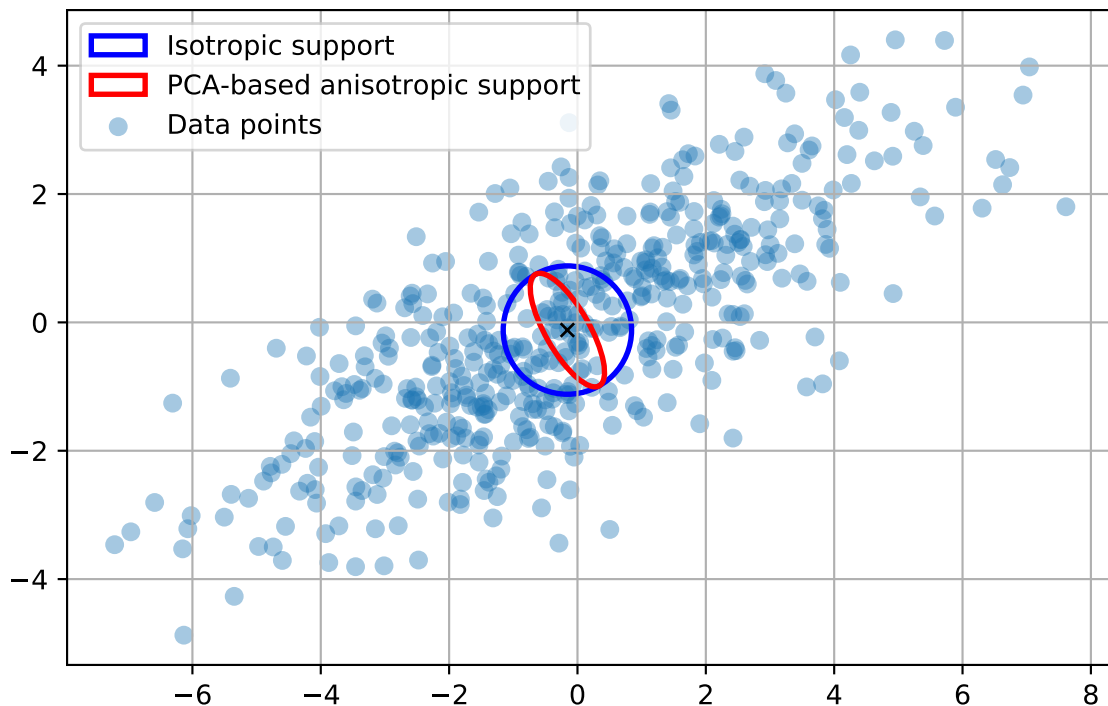


Figure 1: Comparison of isotropic and PCA-based anisotropic kernel supports.

2.4 EBF Multilevel

Building on the preceding ideas, we extend the algorithm presented in [Section 2.1](#), resulting in the enhanced version described in [Algorithm 2](#). Notably, the PCA procedure introduced in [Section 2.3](#) may, in principle, be applied locally, for instance, by computing the principal directions in spatial neighborhoods around each center. This allows for the construction of anisotropic kernels whose shape and orientation vary across the domain. However, such local adaptation may induce asymmetries in the resulting kernel matrices, and in the worst case may compromise their positive definiteness or numerical invertibility, cf. [\[13\]](#). That is why we opt to perform the PCA on the whole set of sites X_j for every level and obtain one scaling matrix A_j per level j .

The difference of [Algorithm 2](#) to the standard multilevel [Algorithm 1](#) lies in the first two lines of the for-loop, where we have to compute the scaling matrix for each level. Clearly, this

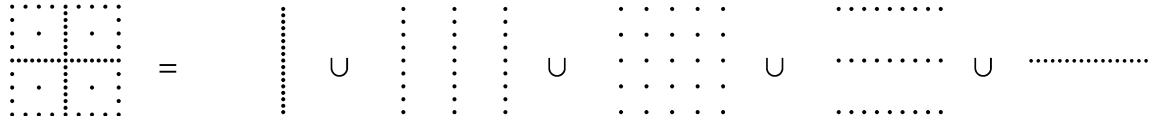


Figure 2: Example of decomposition of sparse grid \mathcal{X} for $L = 3$ in two dimensions.

can also be done in an offline phase, as soon as the sets of sites X_1, \dots, X_L are fixed. Finally, we use the elliptic basis function to compute the interpolant instead of the scaled RBF.

We also emphasize that we are not limited to PCA to obtain the scaling matrices A_j . If, for example, information of the sets of sites is known, we recommend using this.

Data: Right-hand side f , number of levels L

Result: Approximate solution f_L

Set $f_0 = 0, e_0 = f$;

for $j = 1, 2, \dots, L$ **do**

Determine the matrices Q, Λ by performing PCA on X_j ;

Compute the scaling matrix $A_j = QDQ^T$ as in (4);

Compute $s_j = \sum_{i=1}^{N_j} \alpha_{j,i} \Phi_{A_j}(\cdot - \mathbf{x}_{j,i})$ by solving the interpolation problem

$s_j|_{X_j} = e_{j-1}|_{X_j}$;

Set $f_j = f_{j-1} + s_j$;

Set $e_j = e_{j-1} - s_j$;

end

Algorithm 2: Elliptic Basis Function Multilevel approximation

3 Numerics

3.1 Application in High-dimensional Approximation

An obvious application of elliptic basis function is in sparse grid methods, already discussed in several publications, see, e.g., [11, 28].

Sparse grids appear naturally in the context of high-dimensional approximation, and, in their easiest form, combine families of sites $X_i \subseteq [0, 1], i \in \mathbb{N}$, according to

$$\mathcal{X} := \bigcup_{|\mathbf{i}| \leq L} X_{i_1} \times \dots \times X_{i_n} \subseteq [0, 1]^n,$$

where $\mathbf{i} \in \mathbb{N}^n$ and $L \in \mathbb{N}$ is a freely adjustable threshold. We illustrate the construction of a sparse grid in Fig. 2 for threshold $L = 3$ in two dimensions with uniform grids $X_i, 1 \leq i \leq 3$.

By construction \mathcal{X} is the union of different axis-aligned, anisotropic grids $X_{i_1} \times \dots \times X_{i_n}$, where the one dimensional grids are usually known. This allows us to construct elliptic kernels on each of these unions, without the need to use PCA. Numerical examples demonstrating the advantage of using elliptic kernels in this context can be found in the above mentioned literature, i.e. in [11, 28]. We thus rather refer the reader to these papers instead of repeating their numerical examples.

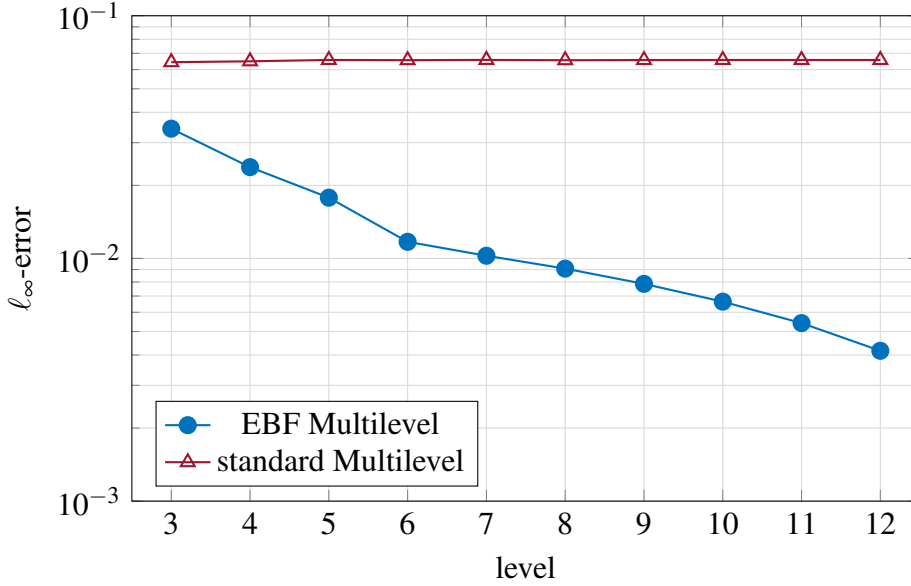


Figure 3: ℓ_∞ -error evaluated on a grid with 1500×1000 points.

3.2 EBF Multilevel with PCA

To test the proposed multilevel method using elliptic basis functions, we consider the anisotropic domain $\Omega = [-5, 5] \times [-0.5, 0.5]$ and as, target function, $f : \Omega \rightarrow \mathbb{R}$ given as $f(x, y) = \sin(10 \cdot x)$, i.e., a function that oscillates rapidly in x -direction and is constant in y -direction.

For each level j , we generate a point set consisting of a structured grid with $100 \cdot j$ points in x -direction and only 10 points in the y -direction. To ensure a more realistic, non-nested multilevel setting, we perturb every grid point independently using uniformly distributed noise in the interval $[-10^{-12}, 10^{-12}]$, generated via a Mersenne Twister engine in C++.

For the multilevel approximation, we choose the C^0 -Wendland kernel, [31],

$$\phi(r) = (1 - r)_+^2.$$

We compute the discrete maximum error on a fine grid with 1500×1000 points.

To benchmark EML, we also compute the standard multilevel method. In the isotropic case, one determines the support radius δ_j by setting $\delta_j = \nu h_j$, and if h_j is not known, one estimates it as $cN_j^{-1/n}$, where N_j is the number of sites in X_j and $c > 0$ an appropriately chosen constant. However, if the family (X_j) is not quasi-uniform, as is the case in this example, this is expected to use either too many or too few data sites within the support of the basis function, as the estimate cannot incorporate the anisotropic growth of the number of sites in x -direction. Nonetheless, as we do not have a better choice, we used this strategy in the computations. The rest of the implementation of the standard multilevel method follows [Algorithm 1](#).

We plot the respective errors in [Fig. 3](#). As we can see, the standard multilevel fails in this setting and the approximation error stagnates, while the EBF multilevel method seems to converge.

In [Table 1](#) we provide some statics on the EML in form of the average number of non-zero entries per row in the interpolation matrices $M_j = (\Phi_{A_j}(\mathbf{x}_{j,i} - \mathbf{x}_{j,k}))$, $5 \leq j \leq 12$, and the number of iterations a non-preconditioned CG solver needs to achieve a precision of 10^{-10} . We see that the number of non-zero entries per row grows, however stays under 23% of N_j . Furthermore, we see that the number of CG-iterations stays relatively small, which indicates

well-conditioned interpolation matrices M_j with a condition number that only mildly depends on the level.

Table 1: Number of average non-zero entries (#nz) per row and the number of needed CG-iterations per level (#iter) to compute EML.

level	5	6	7	8	9	10	11	12
#nz	1362	1605	1831	2056	2282	2525	2751	2976
#iter	1869	2134	2412	2666	2977	3242	3565	3836

Until now, the setting was axis-aligned, and although we used a PCA to derive the correct principle components $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, we also tested the method on a problem that is not axis-aligned. To this end, we rotated Ω and the sets of data sites first by 45° and also by a random angle, only determined during run-time. The target function then became $f(x,y) = \sin(10 \cdot (x \cos(\alpha) + y(\sin(\alpha))))$ with $\alpha = 45^\circ$ or α chosen randomly. In Fig. 4, we show a comparison of the approximation errors for these three cases. We can see that the errors behave qualitatively the same, which means that the PCA correctly determines the rotation needed to determine the elliptic support.

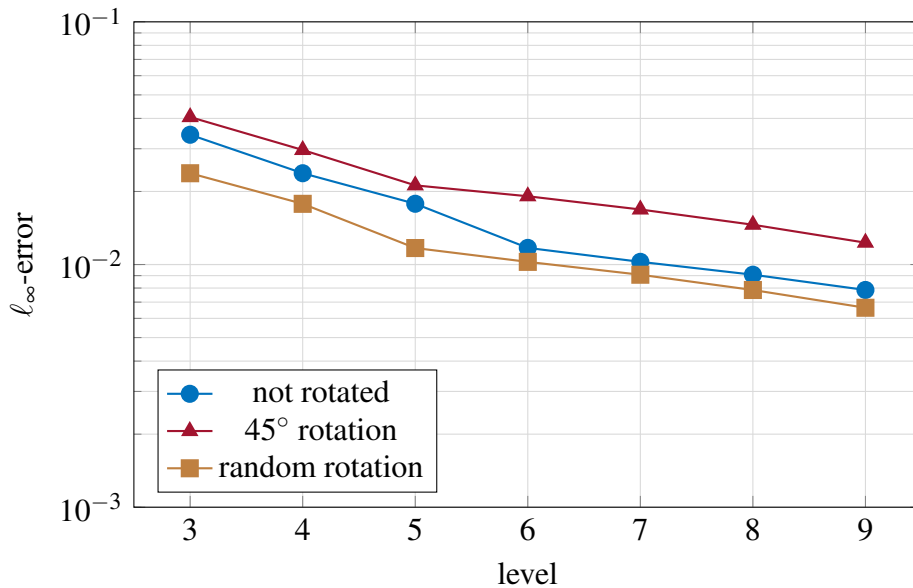


Figure 4: Comparison of EML approximation errors with different rotations.

4 Conclusion and Outlook

In this paper, we studied elliptic basis functions for the interpolation of functions measured at scattered data sites, extending by this the existing literature of such anisotropically scaled basis functions. We discussed these functions in a multilevel residual correction scheme, where the scaling matrix for the EBF is determined level-dependent by performing a PCA on the sets of sites X_j for each level j . Additionally, we derived some theoretical results on the native space of EBFs and the stability of the interpolation process. Furthermore, we demonstrated the power of

the EBF multilevel method for an anisotropic example, where the standard multilevel method fails.

Although there are some error results on EBFs this has to be studied in more detail in the future. In particular, we want to derive a convergence result of EBF multilevel similar to the one known for the standard case.

5 Code structure and usage

The accompanying C++ code provides a minimal working example for the anisotropic multilevel kernel-based approximation method proposed in this paper. The implementation is intentionally modular and focuses on clarity and reproducibility rather than performance optimizations.

Main program. The entry point of the code is a minimal main function which calls the routine `rotated_domain()`. In this function, all problem-specific parameters are defined, including the spatial dimension, the number of levels, and the overlap parameter used in the kernel supports. A random rotation angle is generated using the C++ standard library random number facilities in order to define a rotated test function. The target function is passed as a `std::function<double(const std::vector<double>&)>` to the solver.

Scattered data sites are generated via a `SitesFactory`, which produces point sets for each refinement level. Based on these inputs, an instance of the class `AnisotropicMultilevel` is constructed and the approximation is computed by calling its `solve` method.

Anisotropic multilevel solver. The core of the implementation is the class `AnisotropicMultilevel`, which encapsulates the complete multilevel algorithm. It manages the hierarchical structure of the approximation, the kernel evaluations, and the solution of the linear systems arising at each level.

The method `solve` implements the multilevel procedure described in the paper. Starting from the coarsest level, kernel interpolants are computed iteratively and the residual is updated after each level. The approximation can be evaluated at arbitrary points using the method `at`, or restricted to a given level using `atUpToLevel`.

Internally, each level stores its own kernel, a KD-tree for neighborhood search, and the corresponding kernel matrix. The method `setUpLevels` initializes these data structures, while `updateResiduals` updates the residual values passed to the next level. The number of levels and the spatial dimension are inferred from the input data.

Supporting components. Several auxiliary components are provided to keep the implementation modular. A kernel factory is used to construct Wendland kernels with appropriate smoothness. Nearest-neighbor and range queries are handled by a dedicated KD-tree implementation. Anisotropic scaling matrices are computed using a PCA-based approach relying on the Eigen library. Linear systems are solved using a custom sparse matrix class and a conjugate gradient method without preconditioning.

Code availability. A minimal working example focusing on the core routines of the proposed method is available at <https://github.com/rtomic/EBF>.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] F. Aioli and M. Donini, *Learning Anisotropic RBF Kernels*, in *Artificial Neural Networks and Machine Learning – ICANN 2014*, Springer Cham (2014), vol. 8681 of *Lecture notes in computer science*, 515–522, URL http://dx.doi.org/10.1007/978-3-319-11179-7_65
- [2] S. Avesani, R. Kempf, M. Multerer, and H. Wendland, *Multiscale Scattered Data Analysis in Samplet Coordinates*, *SIAM Journal on Scientific Computing*, vol. 47, no. 5 (2025) A3038–A3063, URL <http://dx.doi.org/10.1137/24M1696305>
- [3] R. Beatson, O. Davydov, and J. Levesley, *Error bounds for anisotropic RBF interpolation*, *Journal of Approximation Theory*, vol. 162, no. 3 (2010) 512–527, URL <http://dx.doi.org/https://doi.org/10.1016/j.jat.2009.08.004>
- [4] M. Bozzini, L. Lenarduzzi, M. Rossini, and R. Schaback, *Interpolation with variably scaled kernels*, *IMA J. Numer. Anal.*, vol. 35, no. 1 (2015) 199–219, URL <https://doi.org/10.1093/imanum/drt071>
- [5] G. Cheng and V. Shcherbakov, *Anisotropic radial basis function methods for continental size ice sheet simulations*, *J. Comput. Phys.*, vol. 372 (2018) 161–177, URL <https://doi.org/10.1016/j.jcp.2018.06.020>
- [6] P. Farrell and H. Wendland, *RBF Multiscale Collocation for Second Order Elliptic Boundary Value Problems*, *SIAM Journal on Numerical Analysis*, vol. 51, no. 4 (2013) 2403–2425, URL <http://dx.doi.org/10.1137/120898383>
- [7] G. E. Fasshauer, *Meshfree Approximation Methods with MATLAB (with CD-rom)*, vol. 6 of *Interdisciplinary mathematical sciences*, World Scientific Publishing, Singapore (2007), URL <https://doi.org/10.1142/6437>
- [8] M. S. Floater and A. Iske, *Multistep scattered data interpolation using compactly supported radial basis functions*, *Journal of Computational and Applied Mathematics*, vol. 73, no. 1 (1996) 65–78, URL [https://doi.org/10.1016/0377-0427\(96\)00035-0](https://doi.org/10.1016/0377-0427(96)00035-0)
- [9] C. Franke and R. Schaback, *Convergence order estimates of meshless collocation methods using radial basis functions*, *Adv. Comput. Math.*, vol. 8 (1998) 381–399, URL <https://doi.org/10.1023/A:1018916902176>
- [10] C. Franke and R. Schaback, *Solving partial differential equations by collocation using radial basis functions*, *Appl. Math. Comput.*, vol. 93 (1998) 73–82, URL [https://doi.org/10.1016/S0096-3003\(97\)10104-7](https://doi.org/10.1016/S0096-3003(97)10104-7)
- [11] E. H. Georgoulis, J. Levesley, and F. Subhan, *Multilevel Sparse Kernel-Based Interpolation*, *SIAM Journal on Scientific Computing*, vol. 35, no. 2 (2013) A815–A831, URL <http://dx.doi.org/10.1137/110859610>

- [12] T. Hofmann, B. Schölkopf, and A. J. Smola, *Kernel methods in machine learning*, Ann. Stat., vol. 36, no. 3 (2008) 1171–1220, URL <http://dx.doi.org/10.1214/009053607000000677>
- [13] Y. C. Hon and R. Schaback, *On unsymmetric collocation by radial basis functions*, J. Appl. Math. Comp., vol. 119 (2001) 177–186, URL [https://doi.org/10.1016/S0096-3003\(99\)00255-6](https://doi.org/10.1016/S0096-3003(99)00255-6)
- [14] E. J. Kansa, *Multiquadrics - A scattered data approximation scheme with applications to computational fluid-dynamics - II: Solutions to parabolic, hyperbolic and elliptic partial differential equations*, Comput. Math. Appl., vol. 19 (1990) 147–161, URL [https://doi.org/10.1016/0898-1221\(90\)90271-K](https://doi.org/10.1016/0898-1221(90)90271-K)
- [15] E. J. Kansa, *Multiquadrics - A scattered data approximation scheme with applications to computational fluid-dynamics I. Surface approximations and partial derivative estimates*, Comput. Math. Appl., vol. 19 (1990) 127–145, URL [https://doi.org/10.1016/0898-1221\(90\)90270-T](https://doi.org/10.1016/0898-1221(90)90270-T)
- [16] R. Kempf and H. Wendland, *High-dimensional approximation with kernel-based multi-level methods on sparse grids*, Numer. Math. (Heidelb.), vol. 154, no. 3-4 (2023) 485–519, URL <https://doi.org/10.1007/s00211-023-01363-x>
- [17] J. Künemund, F. J. Narcowich, J. D. Ward, and H. Wendland, *A high-order meshless Galerkin method for semilinear parabolic equations on spheres*, Numer. Math. (Heidelb.), vol. 142, no. 2 (2019) 383–419, URL <https://doi.org/10.1007/s00211-018-01021-7>
- [18] E. Larsson and U. Sundin, *An investigation of global radial basis function collocation methods applied to Helmholtz problems*, arXiv (2020), URL <http://dx.doi.org/10.48550/arXiv.2001.11090>
- [19] Q. Le Gia, I. Sloan, and H. Wendland, *Multiscale approximation for functions in arbitrary Sobolev spaces by scaled radial basis functions on the unit sphere*, Applied and Computational Harmonic Analysis, vol. 32, no. 3 (2012) 401–412, URL <http://dx.doi.org/10.1016/j.acha.2011.07.007>
- [20] Q. T. Le Gia, I. H. Sloan, and H. Wendland, *Multiscale RBF collocation for solving PDEs on spheres*, Numer. Math. (Heidelb.), vol. 121, no. 1 (2012) 99–125, URL <http://dx.doi.org/10.1007/s00211-011-0428-6>
- [21] B. Matern, *Spatial Variation*, Springer New York, NY, vol. 36 of *Lecture notes in statistics*, 2 ed. (1986), URL <http://dx.doi.org/10.1007/978-1-4615-7892-5>
- [22] R. Schaback, *On the Efficiency of Interpolation by Radial Basis Functions*, in *Proceedings of Chamonix 1996*, 309–318, URL https://www.researchgate.net/publication/2691797_On_the_Efficiency_of_Interpolation_by_Radial_Basis_Functions
- [23] R. Schaback, *Error estimates and condition numbers for radial basis function interpolation*, Adv. Comput. Math., vol. 3, no. 3 (1995) 251–264, URL <https://doi.org/10.1007/BF02432002>

- [24] R. Schaback and H. Wendland, *Approximation by Positive Definite Kernels*, in *Advanced Problems in Constructive Approximation*, edited by M. Buhmann and D. Mache, Birkhäuser, Basel (2002), vol. 142 of *ISNM International Series of Numerical Mathematics*, 203–222, URL https://doi.org/10.1007/978-3-0348-7600-1_15
- [25] R. Schaback and H. Wendland, *Kernel techniques: From machine learning to meshless methods*, *Acta Numerica*, vol. 15 (2006) 543–639, URL <http://dx.doi.org/10.1017/S0962492906270016>
- [26] B. Schoelkopf and A. J. Smola, *Learning with kernels*, Adaptive Computation and Machine Learning Series, MIT Press (2001), ISBN 9780262256933, URL <http://dx.doi.org/10.7551/mitpress/4175.001.0001>
- [27] I. Steinwart and A. Christmann, *Support Vector Machines*, Information science and statistics, Springer New York, NY (2008), ISBN 978-0-387-77242-4, URL <http://dx.doi.org/10.1007/978-0-387-77242-4>
- [28] F. Subhan, *Multilevel Sparse Kernel-Based Interpolation*, PhD thesis, University of Leicester (2011), URL <https://hdl.handle.net/2381/9894>
- [29] A. Townsend and H. Wendland, *Multiscale analysis in Sobolev spaces on bounded domains with zero boundary values*, *IMA J. Numer. Anal.*, vol. 33, no. 3 (2013) 1095–1114, URL <https://doi.org/10.1093/imanum/drs036>
- [30] H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, *Adv. Comput. Math.*, vol. 4, no. 1 (1995) 389–396, URL <https://doi.org/10.1007/BF02123482>
- [31] H. Wendland, *Scattered Data Approximation*, Cambridge University Press, Cambridge Monographs on Applied and Computational Mathematics (2004), ISBN 9780511617539, URL <http://dx.doi.org/10.1017/CBO9780511617539>
- [32] H. Wendland, *Multiscale analysis in Sobolev spaces on bounded domains*, *Numerische Mathematik*, vol. 116, no. 3 (2010) 493–517, URL <http://dx.doi.org/10.1007/s00211-010-0313-8>
- [33] H. Wendland and J. Künemund, *Solving partial differential equations on (evolving) surfaces with radial basis functions*, *Adv. Comput. Math.*, vol. 46, no. 4, URL <https://doi.org/10.1007/s10444-020-09803-0>