Meshfree Methods

G. E. Fasshauer*

Abstract

Meshfree methods are the topic of recent research in many areas of computational science and approximation theory. These methods come in various flavors, most of which can be explained either by what is known in the literature as radial basis functions (RBFs), or in terms of the moving least squares (MLS) method. Over the past several years meshfree approximation methods have found their way into many different application areas ranging from artificial intelligence, computer graphics, image processing and optimization to the numerical solution of all kinds of (partial) differential equations problems. Applications in computational nanotechnology are still somewhat rare, but do exist in the literature. In this chapter we will focus on the mathematical foundation of meshfree methods, and the discussion of various computational techniques presently available for a successful implementation of meshfree methods. At the end of this review we mention some initial applications of meshfree methods to problems in computational nanotechnology, and hope that this introduction will serve as a motivation for others to apply meshfree methods to many other challenging problems in computational nanotechnology.

1 Introduction

1.1 History and Outline

Originally, the motivation for two of the most common basic meshfree approximation methods (radial basis functions and moving least squares methods) came from applications in geodesy, geophysics, mapping, or meteorology. Later, applications were found in many areas such as in the numerical solution of PDEs, artificial intelligence, learning theory, neural networks, signal processing, sampling theory, statistics (kriging), finance, and optimization. It should be pointed out that (meshfree) local regression methods have been used (independently) in statistics for more than 100 years (see, e.g., [37] and references therein).

"Standard" multivariate approximation methods (splines or finite elements) require an underlying mesh (e.g., a triangulation) for the definition of basis functions or elements. This is usually rather difficult to accomplish in space dimensions > 2.

The following provides a very brief history of meshfree methods in approximation theory and lists some of the landmark papers. In the late 1960s D. Shepard [185] suggested the use of what are today known as *Shepard functions* in an application to surface

*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA, fass@amadeus.math.iit.edu

modelling. Another application of meshfree methods, this time in geodesy, was suggested in the early 1970s by Rolland Hardy (see, e.g., [83, 84, 85]). The approaches suggested by Hardy are known as the *multiquadric* (MQ) and *inverse multiquadric* (IMQ) methods. Around the same time Jean Duchon, a mathematician at the Université Joseph Fourier in Grenoble, France, formulated a variational approach that led to *thin plate splines* (TPSs), or more generally *polyharmonic splines* (see [44, 45, 46, 47]). The closely related *surface splines* were studied by Jean Meinguet at Université Catholique de Louvain in Louvain, Belgium (see [130, 131, 132, 133]), while the idea of Shepard's functions was generalized by Peter Lancaster and Kes Šalkauskas of the University of Calgary, Canada. Their method is now known as the moving least squares method [108, 109]. A very important paper was the comparison of all scattered data interpolation methods available in the early 1980s by Richard Franke of the Naval Postgraduate School in Monterey, California. In his comparison [74] he concluded that multiquadrics and thin plate splines were the best methods available at that time. Franke also conjectured that the interpolation matrix that arises in the use of MQs is invertible. This comparison in favor of MQs and TPSs and the related conjecture is what really jumpstarted the research on radial basis functions. The first two major contributions in this area were on the one hand the unpublished manuscript [123] by Wally Madych of the University of Connecticut and S. A. Nelson of Iowa State University. Using a variational approach, they verified (among many other things) Franke's conjecture and provided a complete framework for multivariate interpolation. Roughly at the same time Charles Micchelli (who was at IBM Yorktown Heights) also proved Franke's conjecture. His approach is based on conditionally positive definite functions (see [135]). Of the hundreds of papers that have appeared since the mid 1980s we would like to single out only one other paper: the 1995 paper resulting from a Masters thesis by Holger Wendland of the University of Göttingen in Germany [199]. Wendland was the first to present a class of compactly supported radial basis functions, and thus able to provide a simple approach to computationally efficient meshfree radial basis function methods.

In the following we will quote many of these results in more detail. However, since this is a review, no proofs are provided. For details the reader is referred to the original literature or the recent (and thus far only) book on radial basis functions [27] by Buhmann.

In the remainder of this introductory section we will explain the main concepts of *scattered data interpolation*, *positive definite functions* and *radial functions*. The later sections will go into more details. In section 2 we discuss positive definite and completely monotone functions, while in section 3 we add the idea of polynomial reproduction to the interpolation problem. This will lead to conditionally positive definite functions. The fourth section introduces several different families of compactly supported radial functions. In section 5 we go over the variational approach and discuss some of the error bounds available for radial basis approximation. In the same section we also explain the connection of radial basis function interpolation to the theory of optimal recovery. This connection reveals that radial basis function interpolation can be considered an optimal interpolation scheme. Section 6 deals with least squares approximation (by radial basis functions), while section 7 explains the details of moving least squares approximation. In section 8 we go over a number of issues related to

the practical implementation of meshfree methods such as conditioning problems and preconditioning methods, alternate basis representations, multilevel algorithms, fast evaluation algorithms, and domain decomposition methods. Finally, in section 9 we provide an overview of some approaches to solving partial differential equations, and end the review with a few early contributions of meshfree methods to the computational nanotechnology literature.

1.2 Motivation: Scattered Data Interpolation

We will use the scattered data fitting problem as our motivation for meshfree methods. This application is one of the fundamental problems in approximation theory and data modelling in general. Our desire to have a well-posed problem formulation will naturally lead to the concepts of positive definite matrices, and strictly positive definite functions. These functions, in turn, provide a direct entry into meshfree methods. While (scattered data) interpolation can be applied directly to solve many applications problems, this approximation method also provides a foundation to many numerical PDE solvers.

1.2.1 Scattered Data Interpolation.

In many scientific disciplines one faces the following problem. We have a set of data (measurements, and locations at which these measurements were obtained), and we want to find a rule which allows us to deduce information about the process we are studying also at locations different from those at which we obtained our measurements. Thus, we are trying to find a function which is a "good" fit to the given data. There are many ways to decide what we mean by "good", and the only criterion we will consider now is that we want our approximation (which we will denote by $\mathcal{P}f$) to exactly match the given measurements at the corresponding locations. This approach is called *interpolation*, and if the locations at which the measurements are taken do not lie on a uniform or regular grid, then the process is called *scattered data interpolation*. More precisely, we are considering the following

Problem 1.1 *Given data* $(\boldsymbol{x}_j, y_j)$, $j = 1, \ldots, N$ *with* $\boldsymbol{x}_j \in \mathbb{R}^s$, $y_j \in \mathbb{R}$ *find a (continuous) function* $\mathcal{P}f$ *such that* $\mathcal{P}f(\boldsymbol{x}_j) = y_j$, $j = 1, \ldots, N$.

Here the $\boldsymbol{x}_j$ are the measurement locations (or *data sites*), and the y_j are the corresponding measurements (or *data values*). We will often assume that these values are obtained by sampling a data function f at the data sites, i.e., $y_j = f(\boldsymbol{x}_j)$, $j = 1, \ldots, N$. The fact that we allow $\boldsymbol{x}_j$ to lie in s-dimensional space $\mathbb{R}^s$ means that the formulation of Problem 1.1 allows us to cover many different types of problems. If $s = 1$ the data could be a series of measurements taken over a certain time period, thus the "data sites" $\boldsymbol{x}_j$ would correspond to certain time instances. For $s = 2$ we can think of the data being obtained over a planar region, and so $\boldsymbol{x}_j$ corresponds to the two coordinates in the plane. For instance, we might want to produce a map which shows the rainfall in the state we live in based on the data collected at weather stations located throughout the state. For $s = 3$ we might think of a similar situation in space. One possibility is that we could be interested in the temperature distribution inside some

solid body. Higher-dimensional examples might not be that intuitive, but a multitude of them exist, e.g., in finance, economics or statistics, but also in artificial intelligence or machine learning.

A convenient and common approach to solving the scattered data problem is to make the assumption that the function $\mathcal{P}f$ is a linear combination of certain *basis functions* B_k, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k B_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s\,. \tag{1}$$

Solving the interpolation problem under this assumption leads to a system of linear equations of the form

$$A\boldsymbol{c} = \boldsymbol{y},$$

where the entries of the *interpolation matrix* A are given by $A_{jk} = B_k(\boldsymbol{x}_j)$, $j,k = 1,\ldots,N$, $\boldsymbol{c} = [c_1,\ldots,c_N]^T$, and $\boldsymbol{y} = [y_1,\ldots,y_N]^T$.

Problem 1.1 will be well-posed, i.e., a solution to the problem will exist and be unique, if and only if the matrix A is non-singular.

This kind of problem is completely understood in the univariate setting (i.e., $s = 1$) where it is well known that one can interpolate to arbitrary data at N distinct data sites using a polynomial of degree $N-1$. For the multivariate setting, however, there is the following negative result due to Mairhuber and Curtis from 1956 [125]:

Theorem 1.2 *If $\Omega \subset \mathbb{R}^s$, $s \geq 2$, contains an interior point, then there exist no Haar spaces of continuous functions except for one-dimensional ones.*

Here, a Haar space is a space of functions that guarantees invertibility of the interpolation matrix $(B_k(\boldsymbol{x}_j))_{j,k=1}^N$. As mentioned above, univariate polynomials of degree $N-1$ form an N-dimensional Haar space for data given at $x_1,\ldots,x_N \in \mathbb{R}$. The Mairhuber-Curtis Theorem implies that in the multivariate setting we can no longer expect this to be the case. E.g., it is not possible to perform unique interpolation with (multivariate) polynomials of degree N to data given at arbitrary locations in $\mathbb{R}^2$. The Mairhuber-Curtis Theorem tells us that if we want to have a well-posed multivariate scattered data interpolation problem, then the basis needs to depend on the data locations.

In order to obtain such data dependent approximation spaces we now consider positive definite matrices and functions.

1.2.2 Positive Definite Matrices and Functions

While a *positive definite matrix* is a standard concept from linear algebra we provide its precise definition in order to contrast it with that of a *positive definite function* (probably less well-known to the average reader).

Definition 1.3 *A real symmetric matrix A is called* positive semi-definite *if its associated quadratic form is non-negative, i.e.,*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k A_{jk} \geq 0 \tag{2}$$

for $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$. *If the only vector* $\boldsymbol{c}$ *that turns (2) into an equality is the zero vector, then* A *is called* positive definite.

An important property of positive definite matrices is that all their eigenvalues are positive, and therefore a positive definite matrix is non-singular (but certainly not vice versa).

If we therefore had basis functions B_k in the expansion (1) above which generate a positive definite interpolation matrix, we would always have a well-posed interpolation problem. To this end we introduce the concept of a *positive definite function* from classical analysis.

Definition 1.4 *A real-valued continuous function* Φ *is* positive definite on $\mathbb{R}^s$ *if and only if it is even and*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) \geq 0 \tag{3}$$

for any N *pairwise different points* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$, *and* $\boldsymbol{c} = [c_1, \ldots, c_N]^T \in \mathbb{R}^N$. *The function* Φ *is* strictly positive definite on $\mathbb{R}^s$ *if the only vector* $\boldsymbol{c}$ *that turns (3) into an equality is the zero vector.*

Positive definite functions were first considered in classical analysis early in the 20th century. In the 1920s Mathias [126] seems to have been the first to define and study positive definite functions. An overview of the development of positive definite functions up to the mid 1970s can be found in [191]. There seems to have been no need to study *strictly* positive functions until Micchelli [135] made the connection between scattered data interpolation and positive definite functions. Unfortunately, in the course of history it has turned out that a positive definite function is associated with a positive *semi*-definite matrix. It should be pointed out that when reading recent articles (especially in the radial basis function literature) dealing with (strictly) positive definite functions one has to be aware of the fact that some authors have tried to "correct" history, and now refer to strictly positive definite functions as positive definite functions.

Usually, positive definite functions are defined as complex-valued functions, and complex coefficients $\boldsymbol{c}$ are used. In particular, the celebrated *Bochner's Theorem* (see the next section) provides an exact characterization of complex-valued positive definite functions. In all practical circumstances, however, we will be concerned with real-valued functions only.

Example: Fix a point $\boldsymbol{y}$ in $\mathbb{R}^s$. Then the function $\Phi(\boldsymbol{x}) = e^{\boldsymbol{x}\cdot\boldsymbol{y}}$ is positive definite on $\mathbb{R}^s$ since the quadratic form in Definition 1.4 becomes

$$\begin{aligned}\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) &= \sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k e^{(\boldsymbol{x}_j - \boldsymbol{x}_k)\cdot\boldsymbol{y}} \\ &= \sum_{j=1}^{N} c_j e^{\boldsymbol{x}_j\cdot\boldsymbol{y}} \sum_{k=1}^{N} c_k e^{\boldsymbol{x}_k\cdot\boldsymbol{y}}\end{aligned}$$

$$= \left[\sum_{j=1}^{N} c_j e^{\boldsymbol{x}_j \cdot \boldsymbol{y}}\right]^2 \geq 0.$$

Definition 1.4 and the discussion preceding it suggest that we should use strictly positive definite functions as basis functions in (1), i.e., $B_k(\boldsymbol{x}) = \Phi(\boldsymbol{x} - \boldsymbol{x}_k)$, or

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k \Phi(\boldsymbol{x} - \boldsymbol{x}_k), \qquad \boldsymbol{x} \in \mathbb{R}^s\,. \tag{4}$$

The function $\mathcal{P}f$ of (4) will yield an interpolant that is *translation invariant*, i.e., the interpolant to translated data is the same as the translated interpolant to the original data.

Finally, Definition 1.4 can be generalized to the notion of strictly positive definite kernels of the form $\Phi(\boldsymbol{x}, \boldsymbol{y})$. We will interchangeably make use of (strictly) positive definite functions and (strictly) positive definite kernels later on.

1.2.3 Radial Functions

In many applications it is desirable to have invariance not only under translation, but also under rotation and reflection. This leads to positive definite functions which are also radial. Radial functions have the nice property that they are invariant under all Euclidean transformations (i.e., translations, rotations, and reflections). This is an immediate consequence of the fact that Euclidean transformations are characterized by orthogonal transformation matrices and are therefore norm-invariant. We therefore define

Definition 1.5 *A function* $\Phi : \mathbb{R}^s \to \mathbb{R}$ *is called* radial *provided there exists a* univariate *function* $\varphi : [0, \infty) \to \mathbb{R}$ *such that*

$$\Phi(\boldsymbol{x}) = \varphi(r), \quad \textit{where} \quad r = \|\boldsymbol{x}\|,$$

and $\|\cdot\|$ *is some norm on* $\mathbb{R}^s$ *– usually the Euclidean norm.*

Definition 1.5 says that for a radial function Φ

$$\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| \quad \Longrightarrow \quad \Phi(\boldsymbol{x}_1) = \Phi(\boldsymbol{x}_2), \qquad \boldsymbol{x}_1,\ \boldsymbol{x}_2 \in \mathbb{R}^d\,.$$

However, what makes radial functions most useful for applications is the fact that the interpolation problem becomes insensitive to the dimension s of the space in which the data sites lie. Instead of having to deal with a multivariate function Φ (whose complexity will increase with increasing space dimension s) we can work with the same univariate function φ for all choices of s.

We call the univariate function φ a *(strictly) positive definite radial function on* $\mathbb{R}^s$ if and only if the associated multivariate function Φ is (strictly) positive definite on $\mathbb{R}^s$ in the sense of Definition 1.4 and radial in the sense of Definition 1.5.

2 Positive Definite and Completely Monotone Functions

Below we will first summarize facts about positive definite functions, and closely related completely monotone functions. Most of these facts are integral characterizations and were established in the 1930s by Bochner and Schoenberg. In the second part of this section we will mention the more recent extensions to strictly positive definite and strictly completely monotone functions.

Integral characterizations are an essential ingredient in the theoretical analysis of radial basis functions. Before we get into the details of those integral representations we summarize some formulas for various integral transforms to be used later.

The Fourier transform conventions we will adhere to are laid out in

Definition 2.1 *The* Fourier transform *of $f \in L_1(\mathbb{R}^s)$ is given by*

$$\hat{f}(\boldsymbol{\omega}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} f(\boldsymbol{x}) e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{x}, \qquad \boldsymbol{\omega} \in \mathbb{R}^s, \tag{5}$$

and its inverse Fourier transform *is given by*

$$\check{f}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} f(\boldsymbol{\omega}) e^{i\boldsymbol{x}\cdot\boldsymbol{\omega}} d\boldsymbol{\omega}, \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

Remark: This definition of the Fourier transform can be found in Rudin [163]. Another, just as common, definition uses

$$\hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^s} f(\boldsymbol{x}) e^{-2\pi i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{x}, \tag{6}$$

and can be found in Stein and Weiss [190].

Similarly, we can define the Fourier transform of a finite (signed) measure μ on $\mathbb{R}^s$ by

$$\hat{\mu}(\boldsymbol{\omega}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\mu(\boldsymbol{x}), \qquad \boldsymbol{\omega} \in \mathbb{R}^s\,.$$

Since we will be interested in positive definite radial functions, we note that the Fourier transform of a radial function is again radial. Indeed,

Theorem 2.2 *Let $\Phi \in L_1(\mathbb{R}^s)$ be continuous and radial, i.e., $\Phi(\boldsymbol{x}) = \varphi(\|\boldsymbol{x}\|)$. Then its Fourier transform $\hat{\Phi}$ is also radial, i.e., $\hat{\Phi}(\boldsymbol{\omega}) = \mathcal{F}_s\varphi(\|\boldsymbol{\omega}\|)$ with*

$$\mathcal{F}_s\varphi(r) = \frac{1}{\sqrt{r^{s-2}}} \int_0^\infty \varphi(t) t^{\frac{s}{2}} J_{(s-2)/2}(rt) dt,$$

where $J_{(s-2)/2}$ is the classical Bessel function of the first kind of order $(s-2)/2$.

Remark: The integral transform appearing in Theorem 2.2 is also referred to as a *B*essel transform.

A third integral transform to play an important role in the following is the *Laplace transform*. We have

Definition 2.3 *The* Laplace transform *of a piecewise continuous function f that satisfies $|f(t)| \leq Me^{at}$ for some constants a and M is given by*

$$\mathcal{L}f(s) = \int_0^\infty f(t)e^{-st}dt, \qquad s > a.$$

Similarly, the Laplace transform of a Borel measure μ on $[0, \infty)$ is given by

$$\mathcal{L}\mu(s) = \int_0^\infty e^{-st}d\mu(t).$$

The Laplace transform is continuous at the origin if and only if μ is finite.

2.1 Bochner's Theorem and (Strictly) Positive Definite Functions

One of the most celebrated results on positive definite functions is their characterization in terms of Fourier transforms established by Bochner in 1932 [18] (for $s = 1$) and 1933 [19] (for general s) .

Theorem 2.4 *(Bochner's Theorem) A (complex-valued) function $\Phi \in C(\mathbb{R}^s)$ is positive definite on $\mathbb{R}^s$ if and only if it is the Fourier transform of a finite non-negative Borel measure μ on $\mathbb{R}^s$, i.e.,*

$$\Phi(\boldsymbol{x}) = \hat{\mu}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} e^{-i\boldsymbol{x}\cdot\boldsymbol{y}}d\mu(\boldsymbol{y}), \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

In order to accomplish our goal of guaranteeing a well-posed interpolation problem, we have to extend (if possible) Bochner's characterization to *strictly* positive definite functions.

A sufficient condition for a function to be strictly positive definite on $\mathbb{R}^s$ is

Theorem 2.5 *Let μ be a non-negative finite Borel measure on $\mathbb{R}^s$ whose carrier is not a set of Lebesgue measure zero. Then the Fourier transform of μ is strictly positive definite on $\mathbb{R}^s$.*

Here the *carrier* of a (non-negative) Borel measure defined on some topological space X is given by

$$X \setminus \bigcup\{O :\ O \text{ is open and } \mu(O) = 0\}.$$

The following corollary gives us a way to *construct* strictly positive definite functions.

Corollary 2.6 *Let f be a continuous non-negative function in $L_1(\mathbb{R}^s)$ which is not identically zero. Then the Fourier transform of f is strictly positive definite on $\mathbb{R}^s$.*

Example: The *Gaussian*

$$\Phi(\boldsymbol{x}) = e^{-\alpha\|\boldsymbol{x}\|^2}, \quad \alpha > 0, \tag{7}$$

is strictly positive definite on $\mathbb{R}^s$ for any s. This is essentially due to the fact that the Fourier transform of a Gaussian is again a Gaussian. In particular, for $\alpha = \frac{1}{2}$ we have $\hat{\Phi} = \Phi$ which can be verified by direct calculation. An easier argument (using completely monotone functions) will become available shortly.

Finally, a criterion to check whether a given function is strictly positive definite is given in [205].

Theorem 2.7 *Let Φ be a continuous function in $L_1(\mathbb{R}^s)$. Φ is strictly positive definite if and only if Φ is bounded and its Fourier transform is non-negative and not identically equal to zero.*

Remark: Work toward an analog of Bochner's Theorem, i.e., an complete integral characterization for functions which are strictly positive definite on $\mathbb{R}^s$, is given in [31] for $s = 1$.

2.2 Positive Definite Radial Functions

We now turn our attention to positive definite radial functions. Theorem 2.2 can be used to prove the following characterization due to Schoenberg (see [181], p.816).

Theorem 2.8 *A continuous function $\varphi : [0, \infty) \to \mathbb{R}$ is positive definite and radial on $\mathbb{R}^s$ if and only if it is the Bessel transform of a finite non-negative Borel measure μ on $[0, \infty)$, i.e.,*

$$\varphi(r) = \int_0^\infty \Omega_s(rt) d\mu(t),$$

where

$$\Omega_s(r) = \begin{cases} \cos r & \text{for } s = 1, \\ \Gamma\left(\frac{s}{2}\right)\left(\frac{2}{r}\right)^{(s-2)/2} J_{(s-2)/2}(r) & \text{for } s \geq 2, \end{cases}$$

and $J_{(s-2)/2}$ is the classical Bessel function of the first kind of order $(s-2)/2$.

Since any function which is positive definite and radial on $\mathbb{R}^{s_1}$ is also positive definite and radial on $\mathbb{R}^{s_2}$ as long as $s_2 \leq s_1$, those functions which are positive definite and radial on $\mathbb{R}^s$ *for all* s are of particular interest. This latter class of functions was also characterized by Schoenberg ([181], pp. 817–821.). We saw above that the Gaussians provide an example of such a function.

Theorem 2.9 *A continuous function $\varphi : [0, \infty) \to \mathbb{R}$ is positive definite and radial on $\mathbb{R}^s$ for all s if and only if it is of the form*

$$\varphi(r) = \int_0^\infty e^{-r^2 t^2} d\mu(t),$$

where μ is a finite non-negative Borel measure on $[0, \infty)$.

We end this section with examples of functions that are strictly positive definite and radial on $\mathbb{R}^s$ with restrictions on the space dimension s. Moreover, the following functions differ from the previous ones in that they have *compact support*.
Examples:

1. The *truncated power function*

$$\varphi_\ell(r) = (1-r)_+^\ell \tag{8}$$

is strictly positive definite and radial on $\mathbb{R}^s$ provided ℓ satisfies $\ell \geq \lfloor \frac{s}{2} \rfloor + 1$. For details see [205]. Here we have used the cutoff function $(\cdot)_+$ which is defined by

$$(x)_+ = \begin{cases} x, & \text{for } x \geq 0, \\ 0, & \text{for } x < 0. \end{cases}$$

2. Let $f \in C[0,\infty)$ be non-negative and not identically equal to zero, and define the function φ by

$$\varphi(r) = \int_0^\infty (1-rt)_+^{k-1} f(t)dt. \tag{9}$$

Then φ is strictly positive definite and radial on $\mathbb{R}^s$ provided $k \geq \lfloor \frac{s}{2} \rfloor + 2$. This can been verified by considering the quadratic form

$$\sum_{j=1}^N \sum_{k=1}^N c_j c_k \varphi(\|\boldsymbol{x}_j - \boldsymbol{x}_k\|) = \int_0^\infty \sum_{j=1}^N \sum_{k=1}^N c_j c_k \varphi_{k-1}(t\|\boldsymbol{x}_j - \boldsymbol{x}_k\|) f(t)dt$$

which is non-negative since φ_{k-1} is strictly positive definite by the first example, and f is non-negative. Since f is also assumed to be not identically equal to zero, the only way for the quadratic form to equal zero is if $\boldsymbol{c} = \boldsymbol{0}$.
Note that (9) amounts to another integral transform of f with the compactly supported truncated power function as integration kernel.

An interesting consequence of the Schoenberg characterization of positive definite radial functions on $\mathbb{R}^s$ for all s is that *there are no compactly supported univariate continuous functions that are positive definite and radial on* $\mathbb{R}^s$ *for all* s.

2.3 Completely Monotone Functions

We now introduce a class of functions which is very closely related to positive definite radial functions and leads to a simple characterization of such functions.

Definition 2.10 *A function* $\varphi : [0,\infty) \to \mathbb{R}$ *which is in* $C[0,\infty) \cap C^\infty(0,\infty)$ *and which satisfies*

$$(-1)^\ell \varphi^{(\ell)}(r) \geq 0, \qquad r > 0,\ \ell = 0,1,2,\ldots,$$

is called completely monotone on $[0,\infty)$.

The following theorem which dates back to at least the 1940s [208] gives an integral characterization of completely monotone functions.

Theorem 2.11 *(Hausdorff-Bernstein-Widder Theorem) A function $\varphi : [0,\infty) \to \mathbb{R}$ is completely monotone on $[0,\infty)$ if and only if it is the Laplace transform of a finite non-negative Borel measure μ on $[0,\infty)$, i.e., φ is of the form*

$$\varphi(r) = \mathcal{L}\mu(r) = \int_0^\infty e^{-rt} d\mu(t).$$

In 1938 Schoenberg first linked positive definite radial functions to completely monotone functions:

Theorem 2.12 *A function φ is completely monotone on $[0,\infty)$ if and only if $\Phi = \varphi(\|\cdot\|^2)$ is positive definite and radial on $\mathbb{R}^s$ for all s.*

Remark: Note that the function Φ is now defined via the *square* of the norm. This is different from our earlier definition of radial functions (see Definition 1.5).

Moreover, the following *interpolation theorem* was already proved by Schoenberg in 1938 ([181], p. 823).

Theorem 2.13 *If the function $\varphi : [0,\infty) \to \mathbb{R}$ is completely monotone but not constant, then $\varphi(\|\cdot\|^2)$ is strictly positive definite and radial on $\mathbb{R}^s$ for any s.*

Example: The following functions are completely monotone and not constant. Therefore, they lead to strictly positive definite radial functions on any $\mathbb{R}^s$, and can be used as basic functions to generate bases for (4).

1. The functions $\varphi(r) = (r+\alpha^2)^{-\beta}$, $\alpha,\ \beta > 0$, are completely monotone and not constant since

   $$(-1)^\ell \varphi^{(\ell)}(r) = (-1)^{2\ell}\beta(\beta+1)\cdots(\beta+\ell-1)(r+\alpha^2)^{-\beta-\ell} \geq 0, \qquad \ell = 0,1,2,\ldots.$$

   Thus

   $$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^N c_j \left(\|\boldsymbol{x}-\boldsymbol{x}_j\|^2 + \alpha^2\right)^{-\beta}, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

   can be used to solve the scattered data interpolation problem. The associated interpolation matrix is guaranteed to be positive definite. These functions are known as *inverse multiquadrics*.

2. The functions $\varphi(r) = e^{-\alpha r}$, $\alpha > 0$, are completely monotone and not constant since

   $$(-1)^\ell \varphi^{(\ell)}(r) = \alpha^\ell e^{-\alpha r} \geq 0, \qquad \ell = 0,1,2,\ldots.$$

   Thus

   $$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^N c_j e^{-\alpha\|\boldsymbol{x}-\boldsymbol{x}_j\|^2}, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

   corresponds to interpolation with Gaussian radial basis functions.

As a final remark in this section we mention that we are a long way from having a complete characterization of (radial) functions for which the scattered data interpolation problem has a unique solution. As we will see later, such a characterization will involve also functions which are not strictly positive definite. For example, we will mention a result of Micchelli's according to which *conditionally* positive definite functions of order one can be used for the scattered data interpolation problem. Furthermore, all of the results dealt with so far involve radial basis functions which are centered at the given data sites. There are only limited results addressing the situation in which the centers for the basis functions and the data sites may differ.

3 Scattered Data Interpolation with Polynomial Precision and Conditionally Positive Definite Functions

3.1 Scattered Data Interpolation with Polynomial Precision

Sometimes the assumption on the form (1) of the solution to the scattered data interpolation Problem 1.1 is extended by adding certain polynomials to the expansion, i.e., $\mathcal{P}f$ is now assumed to be of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{k=1}^{N} c_k B_k(\boldsymbol{x}) + \sum_{l=1}^{M} d_l p_l(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{10}$$

where $p_1, \ldots, p_M$ form a basis for the $M = \binom{s+m-1}{m-1}$-dimensional linear space Π^s_{m-1} of polynomials of total degree less than or equal to $m-1$ in s variables.

Since enforcing the interpolation conditions $\mathcal{P}f(\boldsymbol{x}_i) = f(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, leads to a system of N linear equations in the $N+M$ unknowns c_k and d_l one usually adds the M additional conditions

$$\sum_{k=1}^{N} c_k p_l(\boldsymbol{x}_k) = 0, \qquad l = 1, \ldots, M,$$

to ensure a unique solution.

Example: For $m = s = 2$ we add the space of bivariate linear polynomials, i.e., $\Pi^2_1 = \text{span}\{1, x, y\}$. Using the notation $\boldsymbol{x} = (x, y)$ we get the expansion

$$\mathcal{P}f(x, y) = \sum_{k=1}^{N} c_k B_k(x, y) + d_1 + d_2 x + d_3 y, \qquad \boldsymbol{x} = (x, y) \in \mathbb{R}^2,$$

which we use to solve

$$\mathcal{P}f(x_i, y_i) = f(x_i, y_i), \qquad i = 1, \ldots, N,$$

together with the three additional conditions

$$\sum_{k=1}^{N} c_k \quad = \quad 0,$$

$$\begin{aligned} \sum_{k=1}^{N} c_k x_k &= 0, \\ \sum_{k=1}^{N} c_k y_k &= 0. \end{aligned}$$

Remark: While the use of polynomials is somewhat arbitrary (any other set of M linearly independent functions could be used), it is obvious that the addition of polynomials of total degree at most $m-1$ guarantees polynomial precision, i.e., if the data come from a polynomial of total degree less than or equal to $m-1$ they are fitted by that polynomial.

In general, solving the interpolation problem based on the extended expansion (10) now amounts to solving a system of linear equations of the form

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}, \tag{11}$$

where the pieces are given by $A_{jk} = B_k(\boldsymbol{x}_j)$, $j,k = 1,\ldots,N$, $P_{jl} = p_l(\boldsymbol{x}_j)$, $j = 1,\ldots,N$, $l = 1,\ldots,M$, $\boldsymbol{c} = [c_1,\ldots,c_N]^T$, $\boldsymbol{d} = [d_1,\ldots,d_M]^T$, $\boldsymbol{y} = [y_1,\ldots,y_N]^T$, and $\boldsymbol{0}$ is a zero vector of length M.

It is again possible to formulate a theorem concerning the well-posedness of this interpolation problem.

3.2 Conditionally Positive Definite Functions

In analogy to the earlier discussion of interpolation with positive definite functions we will now introduce conditionally positive definite and strictly conditionally positive definite functions of order m. We will not present a linear algebra analog here since in that context only orders $m = 0$ and $m = 1$ are relevant.

As for positive definite functions earlier, we can restrict ourselves to real-valued, even functions Φ and real coefficients. A detailed discussion is presented in [205].

Definition 3.1 *A real-valued continuous even function Φ is called* conditionally positive definite of order m on $\mathbb{R}^s$ *if*

$$\sum_{j=1}^{N}\sum_{k=1}^{N} c_j c_k \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) \geq 0 \tag{12}$$

for any N points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N \in \mathbb{R}^s$, and $\boldsymbol{c} = [c_1,\ldots,c_N]^T \in \mathbb{R}^N$ satisfying

$$\sum_{j=1}^{N} c_j \boldsymbol{x}_j^{\boldsymbol{\alpha}} = 0, \qquad |\boldsymbol{\alpha}| < m, \quad \boldsymbol{\alpha} \in \mathbb{N}_0^s\,.$$

The function Φ is called strictly conditionally positive definite of order m on $\mathbb{R}^s$ *if the points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N \in \mathbb{R}^s$ are distinct, and $\boldsymbol{c} \neq \boldsymbol{0}$ implies strict inequality in (12).*

Here we have used the usual *multi-integer notation*, i.e.,

$$\boldsymbol{\alpha} \in \mathbb{N}_0^s, \quad |\boldsymbol{\alpha}| = \sum_{i=1}^{s} \alpha_i, \quad \text{and} \quad \boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_s^{\alpha_s}.$$

An immediate observation is that a function which is conditionally positive definite of order m on $\mathbb{R}^s$ also is conditionally positive definite of any higher order. In particular, this definition is more general than that for positive definite functions since the case $m = 0$ yields that class of functions, i.e., (strictly) conditionally positive definite functions of order zero are (strictly) positive definite, and therefore a (strictly) positive definite function is always (strictly) conditionally positive definite of any order.

The matrix A with entries $A_{jk} = \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)$ corresponding to a real and even strictly conditionally positive definite function of order m can also be interpreted as being positive definite on the space of vectors $\boldsymbol{c}$ such that

$$\sum_{j=1}^{N} c_j \boldsymbol{x}^{\boldsymbol{\alpha}} = 0, \qquad |\boldsymbol{\alpha}| < m.$$

Thus, in this sense, A is positive definite on the space of vectors $\boldsymbol{c}$ "perpendicular" to polynomials of degree at most $m - 1$.

Using the Courant-Fischer Theorem from linear algebra Micchelli [135] showed that interpolation with strictly conditionally positive definite functions of order one is possible even without adding a polynomial term.

Theorem 3.2 *Suppose Φ is strictly conditionally positive definite of order one and that $\Phi(\mathbf{0}) \leq 0$. Then for any distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^s$ the matrix A with entries $A_{jk} = \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)$ has $N - 1$ positive and 1 negative eigenvalue, and is therefore non-singular.*

As we will see below, this theorem covers Franke's conjecture about *multiquadrics* $\Phi(x) = -(\|\boldsymbol{x}\|^2 + \alpha^2)^\beta$, $\alpha \geq 0$, $0 < \beta < 1$, mentioned in the introduction.

Before we formulate the theorem about the uniqueness of the solution to the interpolation problem based on expansion (10), we define a property which forms a very mild restriction on the location of the data sites.

Definition 3.3 *We call a set of points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^s$* m-unisolvent *if the only polynomial of total degree at most m interpolating zero data on $\mathcal{X}$ is the zero polynomial.*

This definition comes from polynomial interpolation, in which case it guarantees a unique solution for interpolation to given data at a subset of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ by a polynomial of degree m. A sufficient condition (to be found in [35], Ch. 9) on the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ to form an m-unisolvent set in $\mathbb{R}^2$ is

Theorem 3.4 *Suppose $\{L_0, \ldots, L_m\}$ is a set of $m+1$ distinct lines in $\mathbb{R}^2$, and that $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M\}$ is a set of $M = (m+1)(m+2)/2$ distinct points such that the first point lies on L_0, the next two points lie on L_1 but not on L_0, and so on, so that the last $m+1$ points lie on L_m but not on any of the previous lines $L_0, \ldots, L_{m-1}$. Then there exists a unique interpolation polynomial of total degree at most m to arbitrary data given at the points in $\mathcal{U}$. Furthermore, if the data sites $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ contain $\mathcal{U}$ as a subset then they form an m-unisolvent set on $\mathbb{R}^2$.*

This theorem can be generalized to $\mathbb{R}^s$ by using hyperplanes in $\mathbb{R}^s$. A theorem similar to Theorem 3.4 is already proved by Chung and Yao [36].

Example: As can easily be verified, three collinear points in $\mathbb{R}^2$ are not 1-unisolvent, since a linear interpolant, i.e., a plane through three arbitrary heights at these 3 collinear points is not uniquely determined. On the other hand, if a set of points in $\mathbb{R}^2$ contains 3 non-collinear points, then it is 1-unisolvent.

Now we are ready to state

Theorem 3.5 *If the real-valued even function Φ is strictly conditionally positive definite of order m on $\mathbb{R}^s$ and the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ form an $(m-1)$-unisolvent set, then the system of linear equations (11) is uniquely solvable.*

There also exists an analog of Bochner's Theorem, i.e., an integral characterization, for conditionally positive definite functions. However, that goes beyond the scope of this review.

3.3 Conditionally Positive Definite Radial Functions

In analogy to the discussion in Section 2 we now focus on conditionally positive definite functions which are radial on $\mathbb{R}^s$ for all s. The paper [81] by Guo, Hu and Sun contains an integral characterization for such functions. Again, this characterization is too technical to be included here.

The main result in [81] is a characterization of conditionally positive definite radial functions on $\mathbb{R}^s$ for all s in terms of completely monotone functions.

Theorem 3.6 *Let $\varphi \in C[0,\infty) \cap C^\infty(0,\infty)$. Then the function $\Phi = \varphi(\|\cdot\|^2)$ is conditionally positive definite of order m and radial on $\mathbb{R}^s$ for all s if and only if $(-1)^m \varphi^{(m)}$ is completely monotone on $(0,\infty)$.*

In order to get strict conditional positive definiteness we need to generalize Theorem 2.13, i.e., the fact that φ not be constant.

Theorem 3.7 *If φ is as in Theorem 3.6 and not a polynomial of degree at most m, then Φ is strictly conditionally positive definite of order m and radial on $\mathbb{R}^s$ for all s.*

Examples: We can easily verify the conditional positive definiteness of a number of functions used in the radial basis function literature.

1. The functions
$$\varphi(r) = (-1)^{\lceil \beta \rceil}(r+\alpha^2)^\beta, \qquad \alpha > 0,\ \beta > 0, \beta \notin \mathbb{N}$$
imply
$$\varphi^{(k)}(r) = (-1)^{\lceil \beta \rceil}\beta(\beta-1)\cdots(\beta-k+1)(r+\alpha^2)^{\beta-k}$$
so that
$$(-1)^{\lceil \beta \rceil}\varphi^{(\lceil \beta \rceil)}(r) = \beta(\beta-1)\cdots(\beta-\lceil \beta \rceil+1)(r+\alpha^2)^{\beta-\lceil \beta \rceil}$$
is completely monotone. Moreover, $m = \lceil \beta \rceil$ is the smallest possible m such that $(-1)^m\varphi^{(m)}$ is completely monotone. Therefore, the *multiquadrics*
$$\Phi(r) = (-1)^{\lceil \beta \rceil}(r^2+\alpha^2)^\beta, \qquad \alpha > 0,\ \beta > 0,$$
are strictly conditionally positive definite of order $m \geq \lceil \beta \rceil$ and radial on $\mathbb{R}^s$ for all s.

2. The functions
$$\varphi(r) = (-1)^{\lceil \beta/2 \rceil} r^{\beta/2}, \qquad \beta > 0,\ \beta \notin 2\,\mathbb{N},$$
imply
$$\varphi^{(k)}(r) = (-1)^{\lceil \beta/2 \rceil}\frac{\beta}{2}\left(\frac{\beta}{2}-1\right)\cdots\left(\frac{\beta}{2}-k+1\right)r^{\beta/2-k}$$
so that $(-1)^{\lceil \beta/2 \rceil}\varphi^{(\lceil \beta/2 \rceil)}$ is completely monotone and $m = \lceil \beta/2 \rceil$ is the smallest possible m such that $(-1)^m\varphi^{(m)}$ is completely monotone. Therefore, the *powers*
$$\Phi(r) = (-1)^{\lceil \beta/2 \rceil} r^\beta, \qquad \beta > 0,\ \beta \notin 2\,\mathbb{N},$$
are strictly conditionally positive definite of order $m \geq \lceil \beta/2 \rceil$ and radial on $\mathbb{R}^s$ for all s.

3. The *thin plate splines*
$$\Phi(\|\boldsymbol{x}\|) = (-1)^{k+1}\|\boldsymbol{x}\|^{2k}\log\|\boldsymbol{x}\|, \qquad k \in \mathbb{N},$$
are strictly conditionally positive definite of order $m \geq k+1$ and radial on $\mathbb{R}^s$ for all s. To see this we observe that
$$2\Phi(\|\boldsymbol{x}\|) = (-1)^{k+1}\|\boldsymbol{x}\|^{2k}\log(\|\boldsymbol{x}\|^2).$$
Therefore, we let
$$\varphi(r) = (-1)^{k+1}r^k\log r, \qquad k \in \mathbb{N},$$
and get
$$\varphi^{(\ell)}(r) = (-1)^{k+1}k(k-1)\cdots(k-\ell+1)r^{k-\ell}\log r + p_\ell(r), \qquad 1 \leq \ell \leq k,$$
with p_ℓ a polynomial of degree $k-\ell$. Therefore,
$$\varphi^{(k)}(r) = (-1)^{k+1}k!\log r + C$$
and
$$\varphi^{(k+1)}(r) = (-1)^{k+1}\frac{k!}{r},$$
which is completely monotone on $(0,\infty)$.

Just as we mentioned earlier that compactly supported radial function cannot be strictly positive definite on $\mathbb{R}^s$ for all s, it is important to note that *there are no truly conditionally positive definite functions with compact support.*

4 Compactly Supported Radial Basis Functions

As just mentioned, compactly supported functions Φ that are truly strictly conditionally positive definite of order $m > 0$ do not exist. The compact support automatically ensures that Φ is strictly positive definite. Another observation was that compactly supported radial functions can be strictly positive definite on $\mathbb{R}^s$ only for a fixed maximal s-value. It is not possible for a function to be strictly positive definite and radial on $\mathbb{R}^s$ for all s and also have a compact support. Therefore we focus our attention on the characterization and construction of functions that are compactly supported, strictly positive definite and radial on $\mathbb{R}^s$ for some fixed s.

According to results stated earlier (Bochner's Theorem and generalizations thereof), a function is strictly positive definite and radial on $\mathbb{R}^s$ if its s-variate Fourier transform is non-negative. Theorem 2.2 gives the Fourier transform of $\Phi = \varphi(\|\cdot\|)$ as

$$\hat{\Phi}(\boldsymbol{x}) = \mathcal{F}_s\varphi(r) = r^{-(s-2)/2}\int_0^\infty \varphi(t)t^{s/2}J_{(s-2)/2}(rt)dt.$$

4.1 Operators for Radial Functions and Dimension Walks

Schaback and Wu [180] defined an integral operator and its inverse differential operator, and discussed an entire calculus for how these operators act on radial functions. These operators facilitate the construction of compactly supported radial functions.

Definition 4.1 *1. Let φ be such that $t \mapsto t\varphi(t) \in L_1[0,\infty)$, then we define*

$$(\mathcal{I}\varphi)(r) = \int_r^\infty t\varphi(t)dt, \qquad r \geq 0.$$

2. For even $\varphi \in C^2(\mathbb{R})$ we define

$$(\mathcal{D}\varphi)(r) = -\frac{1}{r}\varphi'(r), \qquad r \geq 0.$$

In both cases the resulting functions are to be interpreted as even functions using even extension.

The most important properties of these operators are (see, e.g., [180] or [199]):

Theorem 4.2 *1. Both $\mathcal{D}$ and $\mathcal{I}$ preserve compact support, i.e., if φ has compact support, then so do $\mathcal{D}\varphi$ and $\mathcal{I}\varphi$.*

2. If $\varphi \in C^(\mathbb{R})$ and $t \mapsto t\phi(t) \in L_1[0,\infty)$, then $\mathcal{D}\mathcal{I}\varphi = \varphi$.

3. If $\varphi \in C^2(\mathbb{R})$ is even and $\varphi' \in L_1[0,\infty)$, then $\mathcal{I}\mathcal{D}\varphi = \varphi$.

4. *If* $t \mapsto t^{s-1}\varphi(t) \in L_1[0,\infty)$ *and* $s \geq 3$, *then* $\mathcal{F}_s(\varphi) = \mathcal{F}_{s-2}(\mathcal{I}\varphi)$.

5. *If* $\varphi \in C^2(\mathbb{R})$ *is even and* $t \mapsto t^s\varphi'(t) \in L_1[0,\infty)$, *then* $\mathcal{F}_s(\varphi) = \mathcal{F}_{s+2}(\mathcal{D}\varphi)$.

The operators $\mathcal{I}$ and $\mathcal{D}$ allow us to express s-variate Fourier transforms as $(s-2)$- or $(s+2)$-variate Fourier transforms, respectively. In particular, a direct consequence of the above properties and the characterization of strictly positive definite radial functions (Theorem 2.8) is

Theorem 4.3 1. *Suppose* $\varphi \in C(\mathbb{R})$. *If* $t \mapsto t^{s-1}\varphi(t) \in L_1[0,\infty)$ *and* $s \geq 3$, *then* φ *is strictly positive definite and radial on* $\mathbb{R}^s$ *if and only if* $\mathcal{I}\varphi$ *is strictly positive definite and radial on* $\mathbb{R}^{s-2}$.

2. *If* $\varphi \in C^2(\mathbb{R})$ *is even and* $t \mapsto t^s\varphi'(t) \in L_1[0,\infty)$, *then* φ *is strictly positive definite and radial on* $\mathbb{R}^s$ *if and only if* $\mathcal{D}\varphi$ *is strictly positive definite and radial on* $\mathbb{R}^{s+2}$.

This allows us to construct new strictly positive definite radial functions from given ones by a "dimension-walk" technique that steps through multivariate Euclidean space in even increments.

4.2 Wendland's Compactly Supported Functions

In [199] Wendland constructed a popular family of compactly supported radial functions by starting with the truncated power function (which we know to be strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 2\ell - 1$), and then walking through dimensions by repeatedly applying the operator I.

Definition 4.4 *With* $\varphi_\ell(r) = (1-r)_+^\ell$ *we define*

$$\varphi_{s,k} = \mathcal{I}^k \varphi_{\lfloor s/2 \rfloor + k + 1}.$$

It turns out that the functions $\varphi_{s,k}$ are all supported on $[0,1]$ and have a polynomial representation there. More precisely,

Theorem 4.5 *The functions* $\varphi_{s,k}$ *are strictly positive definite and radial on* $\mathbb{R}^s$ *and are of the form*

$$\varphi_{s,k}(r) = \begin{cases} p_{s,k}(r), & r \in [0,1], \\ 0, & r > 1, \end{cases}$$

with a univariate polynomial $p_{s,k}$ *of degree* $\lfloor s/2 \rfloor + 3k + 1$. *Moreover,* $\varphi_{s,k} \in C^{2k}(\mathbb{R})$ *are unique up to a constant factor, and the polynomial degree is minimal for given space dimension* s *and smoothness* $2k$.

Wendland gave recursive formulas for the functions $\varphi_{s,k}$ for all s, k. We instead list the explicit formulas of [57]

Theorem 4.6 *The functions $\varphi_{s,k}$, $k = 0, 1, 2, 3$, have the form*

$$\begin{aligned}
\varphi_{s,0}(r) &= (1-r)_+^{\ell}, \\
\varphi_{s,1}(r) &\doteq (1-r)_+^{\ell+1}\left[(\ell+1)r+1\right], \\
\varphi_{s,2}(r) &\doteq (1-r)_+^{\ell+2}\left[(\ell^2+4\ell+3)r^2+(3\ell+6)r+3\right], \\
\varphi_{s,3}(r) &\doteq (1-r)_+^{\ell+3}\left[(\ell^3+9\ell^2+23\ell+15)r^3+(6\ell^2+36\ell+45)r^2+(15\ell+45)r+15\right],
\end{aligned}$$

where $\ell = \lfloor s/2 \rfloor + k + 1$, and the symbol $\doteq$ denotes equality up to a multiplicative positive constant.

Examples: For $s = 3$ we get some of the most commonly used functions as

$$\begin{aligned}
\varphi_{3,0}(r) &= (1-r)_+^2, && \in C^0 \cap SPD(\mathbb{R}^3) \\
\varphi_{3,1}(r) &\doteq (1-r)_+^4\,(4r+1), && \in C^2 \cap SPD(\mathbb{R}^3) \\
\varphi_{3,2}(r) &\doteq (1-r)_+^6\left(35r^2+18r+3\right), && \in C^4 \cap SPD(\mathbb{R}^3) \\
\varphi_{3,3}(r) &\doteq (1-r)_+^8\left(32r^3+25r^2+8r+1\right), && \in C^6 \cap SPD(\mathbb{R}^3).
\end{aligned}$$

4.3 Wu's Compactly Supported Functions

In [212] Wu presents another way to construct strictly positive definite radial functions with compact support. He starts with the function

$$\psi(r) = (1-r^2)_+^{\ell}, \qquad \ell \in \mathbb{N},$$

which is strictly positive definite and radial. Wu then constructs another function that is strictly positive definite and radial on $\mathbb{R}$ by convolution, i.e.,

$$\begin{aligned}
\psi_\ell(r) &= (\psi * \psi)(2r) \\
&= \int_{-\infty}^{\infty} (1-t^2)_+^{\ell}(1-(2r-t)^2)_+^{\ell}\,dt \\
&= \int_{-1}^{1} (1-t^2)^{\ell}(1-(2r-t)^2)^{\ell}\,dt.
\end{aligned}$$

This function is strictly positive definite since its Fourier transform is essentially the square of the Fourier transform of ψ. Just like the Wendland functions, this function is a polynomial on its support. In fact, the degree of the polynomial is $4\ell + 1$, and $\psi_\ell \in C^{2\ell}(\mathbb{R})$.

Now, a family of strictly positive definite radial functions is constructed by a dimension walk using the $\mathcal{D}$ operator, i.e.,

$$\psi_{k,\ell} = \mathcal{D}^k \psi_\ell.$$

The functions $\psi_{k,\ell}$ are strictly positive definite and radial in $\mathbb{R}^s$ for $s \le 2k+1$, are polynomials of degree $4\ell - 2k + 1$ on their support and in $C^{2(\ell-k)}$ in the interior of the support. On the boundary the smoothness increases to $C^{2\ell-k}$.

Example: For $\ell = 3$ we can compute the three functions

$$\psi_{k,3}(r) = \mathcal{D}^k \psi_3(r) = \mathcal{D}^k((1-\cdot^2)_+^3 * (1-\cdot^2)_+^3)(2r), \qquad k = 0, 1, 2, 3.$$

This results in

$$\begin{aligned}
\psi_{0,3}(r) &\doteq \left(5-39r^2+143r^4-429r^6+429r^7-143r^9+39r^{11}-5r^{13}\right)_+ \\
&\doteq (1-r)_+^7(5+35r+101r^2+147r^3+101r^4+35r^5+5r^6) \qquad \in C^6\cap SPD(\mathbb{R}) \\
\psi_{1,3}(r) &\doteq \left(6-44r^2+198r^4-231r^5+99r^7-33r^9+5r^{11}\right)_+ \\
&\doteq (1-r)_+^6(6+36r+82r^2+72r^3+30r^4+5r^5) \qquad \in C^4\cap SPD(\mathbb{R}^3) \\
\psi_{2,3}(r) &\doteq \left(8-72r^2+105r^3-63r^5+27r^7-5r^9\right)_+ \\
&\doteq (1-r)_+^5(8+40r+48r^2+25r^3+5r^4) \qquad \in C^2\cap SPD(\mathbb{R}^5) \\
\psi_{3,3}(r) &\doteq \left(16-35r+35r^3-21r^5+5r^7\right)_+ \\
&\doteq (1-r)_+^4(16+29r+20r^2+5r^3) \qquad \in C^0\cap SPD(\mathbb{R}^7).
\end{aligned}$$

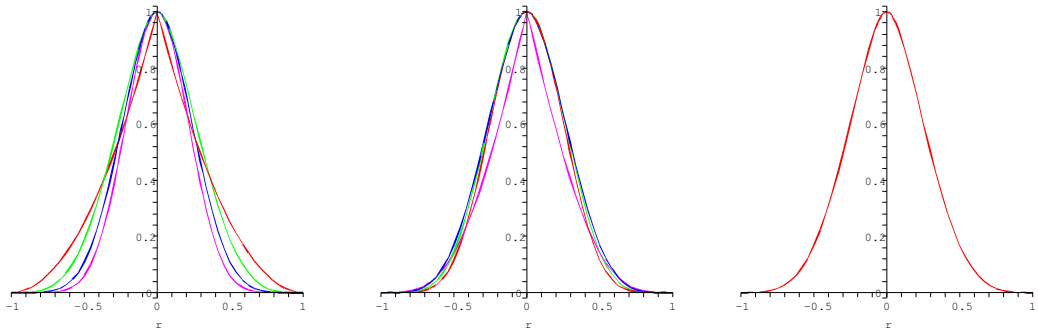

Figure 1: Plot of Wendland's functions (left), Wu's functions (center), and Buhmann's function (right) listed as examples.

Remarks:

1. For a prescribed smoothness the polynomial degree of Wendland's functions is lower than that of Wu's functions. For example, both Wendland's function $\varphi_{3,2}$ and Wu's function $\psi_{1,3}$ are C^4 smooth and strictly positive definite and radial in $\mathbb{R}^3$. However, the polynomial degree of Wendland's function is 8, whereas that of Wu's function is 11.

2. While both families of strictly positive definite compactly supported functions are constructed via dimension walk, Wendland uses integration (and thus obtains a family of increasingly smoother functions), whereas Wu needs to start with a function of sufficient smoothness, and then obtains successively less smooth functions (via differentiation).

4.4 Buhmann's Compactly Supported Functions

A third family of compactly supported strictly positive definite radial functions that has appeared in the literature is due to Buhmann (see [25]). Buhmann's functions contain a logarithmic term in addition to a polynomial. His functions have the general form

$$\phi(r)=\int_0^\infty (1-r^2/t)_+^\lambda t^\alpha (1-t^\delta)_+^\rho dt.$$

Here $0 < \delta \leq \frac{1}{2}$, $\rho \geq 1$, and in order to obtain functions that are strictly positive definite and radial on $\mathbb{R}^s$ for $s \leq 3$ the constraints for the remaining parameters are $\lambda \geq 0$, and $-1 < \alpha \leq \frac{\lambda-1}{2}$.

Example: An example with $\alpha = \delta = \frac{1}{2}$, $\rho = 1$ and $\lambda = 2$ is listed in [26]:

$$\phi(r) \doteq 12r^4 \log r - 21r^4 + 32r^3 - 12r^2 + 1, \qquad 0 \leq r \leq 1, \in C^2 \cap SPD(\mathbb{R}^3).$$

Remarks:

1. While Buhmann [26] claims that his construction encompasses both Wendland's and Wu's functions, Wendland [205] gives an even more general theorem that shows that integration of a positive function $f \in L_1[0, \infty)$ against a strictly positive definite (compactly supported) kernel K results in a (compactly supported) strictly positive definite function, i.e.,

$$\varphi(r) = \int_0^\infty K(t, r) f(t) dt$$

is strictly positive definite. Buhmann's construction then corresponds to choosing $f(t) = t^\alpha (1 - t^\delta)_+^\rho$ and $K(t, r) = (1 - r^2/t)_+^\lambda$.

5 Error Bounds and the Variational Approach

In order to estimate the approximation properties of the functions studied thus far we will now consider the variational approach to scattered data interpolation. This approach was used first for radial basis function interpolation by Madych and Nelson [123], and later adopted by many others (see, e.g., [117, 118], [155], [167], [200, 201], [213]). We will see that for every strictly positive definite radial function there is an associated Hilbert space in which the radial basis function interpolant provides the best approximation to a given function. This optimality of interpolants in Hilbert space is the subject of the theory of optimal recovery described in the late 1950s by Golomb and Weinberger in their paper [79]. The following discussion follows mostly the presentation in [205].

5.1 Reproducing Kernel Hilbert Spaces

We begin with

Definition 5.1 *Let $\mathcal{H}$ be a real Hilbert space of functions $f : \Omega \to \mathbb{R}$. A function $K : \Omega \times \Omega \to \mathbb{R}$ is called* reproducing kernel for $\mathcal{H}$ *if*

1. *$K(\boldsymbol{x}, \cdot) \in \mathcal{H}$ for all $\boldsymbol{x} \in \Omega$,*
2. *$f(\boldsymbol{x}) = \langle f, K(\cdot, \boldsymbol{x}) \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and all $\boldsymbol{x} \in \Omega$.*

It is known that the reproducing kernel of a Hilbert space is unique, and that existence of a reproducing kernel is equivalent to the fact that the point evaluation functionals $\delta_{\boldsymbol{x}}$ are bounded linear functionals, i.e., there exists a positive constant $M = M_{\boldsymbol{x}}$ such that

$$|\delta_{\boldsymbol{x}} f| = |f(\boldsymbol{x})| \leq M \|f\|_{\mathcal{H}}$$

for all $f \in \mathcal{H}$. This latter fact is due to the Riesz Representation Theorem.

Other properties of reproducing kernels are

Theorem 5.2 *Suppose $\mathcal{H}$ is a Hilbert space of functions $f : \Omega \to \mathbb{R}$ with reproducing kernel K and $\mathcal{H}^*$ its dual space, i.e., the space of linear functionals on $\mathcal{H}$. Then we have*

1. $K(\boldsymbol{x}, \boldsymbol{y}) = \langle K(\boldsymbol{x}, \cdot), K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}}$ *for* $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.
2. $K(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{y}, \boldsymbol{x})$ *for* $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.
3. *Convergence in Hilbert space norm implies pointwise convergence.*

Moreover, the reproducing kernel K is known to be positive definite and K is strictly positive definite if and only if the point evaluation functionals are linearly independent in $\mathcal{H}^*$. In the following we use a slight generalization of the notion of a positive definite function to a positive definite kernel. Essentially, we replace $\Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)$ in Definition 1.4 by $K(\boldsymbol{x}_j, \boldsymbol{x}_k)$.

While it is good to know that every reproducing kernel is positive definite our interest, however, lies in the other direction. Since we are starting with strictly positive definite functions, we need to show how to construct an associated reproducing kernel Hilbert space.

5.2 Native Spaces for Strictly Positive Definite Functions

First, we note that Definition 5.1 tells us that $\mathcal{H}$ contains all functions of the form

$$f = \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot)$$

provided $\boldsymbol{x}_j \in \Omega$. Theorem 5.2 implies that

$$\begin{aligned} \|f\|_{\mathcal{H}}^2 &= \langle f, f \rangle_{\mathcal{H}} = \langle \sum_{j=1}^{N} c_j K(\boldsymbol{x}_j, \cdot), \sum_{k=1}^{N} c_k K(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k \langle K(\boldsymbol{x}_j, \cdot), K(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{H}} \\ &= \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k K(\boldsymbol{x}_j, \boldsymbol{x}_k). \end{aligned}$$

Therefore, we *define* the space

$$H_K(\Omega) = \operatorname{span}\{K(\cdot, \boldsymbol{y}) : \boldsymbol{y} \in \Omega\}$$

with an associated bilinear form

$$\langle \sum_{j=1}^{N} c_j K(\cdot, \boldsymbol{x}_j), \sum_{k=1}^{N} d_k K(\cdot, \boldsymbol{y}_k) \rangle_K = \sum_{j=1}^{N} \sum_{k=1}^{N} c_j d_k K(\boldsymbol{x}_j, \boldsymbol{y}_k).$$

Theorem 5.3 *If $K : \Omega \times \Omega \to \mathbb{R}$ is a symmetric strictly positive definite kernel, then the bilinear form $\langle \cdot, \cdot \rangle_K$ defines an inner product on $H_K(\Omega)$. Furthermore, $H_K(\Omega)$ is a pre-Hilbert space with reproducing kernel K.*

The *native space* $\mathcal{N}_K(\Omega)$ of K is now defined to be the completion of $H_K(\Omega)$ with respect to the K-norm $\| \cdot \|_K$ so that $\|f\|_K = \|f\|_{\mathcal{N}_K(\Omega)}$ for all $f \in H_K(\Omega)$. There are a number of technical details concerned with this construction which are discussed in the papers [172, 173] by Schaback or [205] by Wendland.

In the special case when we are dealing with strictly positive definite (translation invariant) functions $\Phi(\boldsymbol{x} - \boldsymbol{y}) = K(\boldsymbol{x}, \boldsymbol{y})$ and when $\Omega = \mathbb{R}^s$ we get a characterization of native spaces in terms of Fourier transforms.

Theorem 5.4 *Suppose $\Phi \in C(\mathbb{R}^s) \cap L_1(\mathbb{R}^s)$ is a real-valued strictly positive definite function. Define*

$$\mathcal{G} = \{f \in L_2(\mathbb{R}^s) \cap C(\mathbb{R}^s) : \frac{\hat{f}}{\sqrt{\hat{\Phi}}} \in L_2(\mathbb{R}^s)\}$$

and equip this space with the bilinear form

$$\langle f, g \rangle_{\mathcal{G}} = \frac{1}{\sqrt{(2\pi)^s}} \langle \frac{\hat{f}}{\sqrt{\hat{\Phi}}}, \frac{\hat{g}}{\sqrt{\hat{\Phi}}} \rangle_{L_2(\mathbb{R}^s)} = \frac{1}{\sqrt{(2\pi)^s}} \int_{\mathbb{R}^s} \frac{\hat{f}(\boldsymbol{\omega})\overline{\hat{g}(\boldsymbol{\omega})}}{\hat{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega}.$$

Then $\mathcal{G}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$ and reproducing kernel $\Phi(\cdot - \cdot)$. Hence, $\mathcal{G}$ is the native space of Φ on $\mathbb{R}^s$, i.e., $\mathcal{G} = \mathcal{N}_\Phi(\mathbb{R}^s)$ and both inner product coincide. In particular, every $f \in \mathcal{N}_\Phi(\mathbb{R}^s)$ can be recovered from its Fourier transform $\hat{f} \in L_1(\mathbb{R}^s) \cap L_2(\mathbb{R}^s)$.

Remarks:

1. This theorem shows that the native spaces can be viewed as a generalization of the standard Sobolev spaces. Indeed, for $m > s/2$ the Sobolev space W_2^m can be defined as

   $$W_2^m(\mathbb{R}^s) = \{f \in L_2(\mathbb{R}^s) \cap C(\mathbb{R}^s) : \hat{f}(\cdot)(1 + \| \cdot \|_2^2)^{m/2} \in L_2(\mathbb{R}^s)\}.$$

   Therefore, any strictly positive definite function Φ whose Fourier transform decays only algebraically has a Sobolev space as its native space. In particular, the compactly supported Wendland functions $\Phi_{s,k} = \varphi_{s,k}(\| \cdot \|_2)$ of Section 4 can be shown to have native spaces $\mathcal{N}_{\Phi_{s,k}}(\mathbb{R}^s) = W_2^{s/2+k+1/2}(\mathbb{R}^s)$ (where the restriction $s \geq 3$ is required for the special case $k = 0$).

2. Native spaces for strictly conditionally positive definite functions can also be constructed. However, since this is more technical, we limit the discussion here to strictly positive definite functions, and refer the interested reader to the papers [172, 173] or [205].

3. The native spaces of the powers and thin plate (or surface) splines of Examples 2 and 3 of Section 3.3 can be shown to be the so-called Beppo-Levi spaces of order k

$$BL_k(\mathbb{R}^s) = \{f \in C(\mathbb{R}^s) : D^{\boldsymbol{\alpha}} f \in L_2(\mathbb{R}^s) \text{ for all } |\boldsymbol{\alpha}| = k,\ \boldsymbol{\alpha} \in \mathbb{N}^s\},$$

where $D^{\boldsymbol{\alpha}}$ denotes a *generalized derivative* of order $\boldsymbol{\alpha}$. In fact, the intersection of all Beppo-Levi spaces $BL_k(\mathbb{R}^s)$ of order $k \le m$ yields the Sobolev space $W_2^m(\mathbb{R}^s)$. For more details see [205]. These spaces were already studied in the early papers by Duchon [44, 45, 46, 47].

4. The native spaces for Gaussians and (inverse) multiquadrics are rather small. For example, according to Theorem 5.4, for Gaussians the Fourier transform of $f \in \mathcal{N}_\Phi(\Omega)$ must decay faster than the Fourier transform of the Gaussian (which is itself a Gaussian). It is known that, however, even though the native space of Gaussians is small, it does contain the so-called *band-limited functions*, i.e., functions whose Fourier transform is compactly supported. These functions play an important role in *sampling theory* where Shannon's famous Sampling Theorem [184] states that any band-limited function can be completely recovered from its discrete samples provided the function is sampled at a sampling rate at least twice its bandwidth. The content of this theorem was already known to Whitaker [206] in 1915.

   Theorem 5.5 *Suppose $f \in C(\mathbb{R}^s) \cap L_1(\mathbb{R}^s)$ such that its Fourier transform vanishes outside the cube $Q = \left[-\frac{1}{2}, \frac{1}{2}\right]^s$. Then f can be uniquely reconstructed from its values on $\mathbb{Z}^s$, i.e.,*

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^s} f(\boldsymbol{\xi}) \mathrm{sinc}(\boldsymbol{x} - \boldsymbol{\xi}), \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

   Here the sinc function is defined for any $\boldsymbol{x} = (x_1, \ldots, x_s) \in \mathbb{R}^s$ as $\mathrm{sinc}\, \boldsymbol{x} = \prod_{i=1}^s \frac{\sin(\pi x_i)}{\pi x_i}$. For more details on Shannon's Sampling Theorem see, e.g., Chapter 29 in the book [33] by Cheney and Light or the paper [194] by Unser.

5.3 The Power Function and Error Estimates

Our goal in this section is to provide error estimates for scattered data interpolation with strictly (conditionally) positive definite functions. In their final form these estimates will need to involve some kind of measure of the data distribution. The measure that is usually used is the so-called *fill distance*

$$h = h_{\mathcal{X},\Omega} = \sup_{\boldsymbol{x} \in \Omega} \min_{\boldsymbol{x}_j \in \mathcal{X}} \|\boldsymbol{x} - \boldsymbol{x}_j\|_2$$

which indicates how well the data fill out the domain Ω. The fill distance denotes the radius of the largest possible empty ball that can be placed among the data locations. We will be interested in whether the error

$$\|f - \mathcal{P}_h f\|_\infty$$

tends to zero as $h \to 0$, and if so, how fast. Here $\{\mathcal{P}_h\}_h$ presents a sequence of interpolation (or, more generally, projection) operators that vary with the fill distance h. For example, $\mathcal{P}_h$ could denote interpolation to data given at $(2^n+1)^s$, $n = 1, 2, \ldots$, equally spaced points in the unit cube in $\mathbb{R}^s$ (with $h = 2^{-n}$). Of course, the definition of the fill distance allows for scattered data as well.

Since we want to use the machinery of reproducing kernel Hilbert spaces we will concentrate on error estimates for functions $f \in \mathcal{N}_\Phi$. At the end of this section we will also mention some more general estimates.

The term that is often used to measure the speed of convergence to zero is *approximation order*. We say that the approximation operator $\mathcal{P}_h$ has L_p-approximation order k if

$$\|f - \mathcal{P}_h f\|_p = \mathcal{O}(h^k) \quad \text{for } h \to 0.$$

Moreover, if we can also show that $\|f - \mathcal{P}_h f\|_p \neq o(h^k)$, then $\mathcal{P}_h$ has *exact* L_p-approximation order k. We will concentrate mostly on the case $p = \infty$, but approximation order in other norms can also be studied.

In order to keep the following discussion as transparent as possible we will restrict ourselves to strictly positive definite functions. With (considerably) more technical details the following can also be formulated for strictly conditionally positive definite functions (see [205] for details).

The key idea for the following discussion is to express the interpolant in Lagrange form, i.e., using *cardinal basis functions*. This idea is due to Schaback and Wu [213]. In the previous sections we established that, for any strictly positive definite function Φ, the linear system

$$A\boldsymbol{c} = \boldsymbol{y}$$

with $A_{ij} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$, $i, j = 1, \ldots, N$, $\boldsymbol{c} = [c_1, \ldots, c_N]^T$, and $\boldsymbol{y} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ has a unique solution. In the following we will consider the more general situation where Φ is a strictly positive definite kernel, i.e., the entries of A are given by $A_{ij} = \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

In order to obtain the cardinal basis functions u_j^*, $j = 1, \ldots, N$, with the property $u_j^*(\boldsymbol{x}_i) = \delta_{ij}$ we consider the linear system

$$A\boldsymbol{u}^*(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}),$$

where the matrix A is as above (and therefore invertible), $\boldsymbol{u}^* = [u_1^*, \ldots, u_N^*]^T$, and $\boldsymbol{b} = [\Phi(\cdot, \boldsymbol{x}_1), \ldots, \Phi(\cdot, \boldsymbol{x}_N)]^T$. Thus,

Theorem 5.6 *Suppose Φ is a strictly positive definite kernel on $\mathbb{R}^s$. Then, for any distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, there exist functions $u_j^* \in \operatorname{span}\{\Phi(\cdot, \boldsymbol{x}_j), j = 1, \ldots, N\}$ such that $u_j^*(\boldsymbol{x}_i) = \delta_{ij}$.*

Therefore, we can write $\mathcal{P}f$ in the cardinal form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s.$$

Another important ingredient in our estimates is the so-called *power function*. To this end, we consider a domain $\Omega \subseteq \mathbb{R}^s$. Then for any strictly positive definite kernel

$\Phi \in C(\Omega \times \Omega)$, any set of distinct points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$, and any vector $\boldsymbol{u} \in \mathbb{R}^N$, we define the quadratic form

$$\begin{aligned} Q(\boldsymbol{u}) &= \Phi(\boldsymbol{x},\boldsymbol{x}) - 2\sum_{j=1}^{N} u_j \Phi(\boldsymbol{x},\boldsymbol{x}_j) + \sum_{i=1}^{N}\sum_{j=1}^{N} u_i u_j \Phi(\boldsymbol{x}_i,\boldsymbol{x}_j) \\ &= \langle \Phi(\cdot,\boldsymbol{x}), \Phi(\cdot,\boldsymbol{x})\rangle_{\mathcal{N}_\Phi(\Omega)} - 2\sum_{j=1}^{N} u_j \langle \Phi(\cdot,\boldsymbol{x}), \Phi(\cdot,\boldsymbol{x}_j)\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{N} u_i u_j \langle \Phi(\cdot,\boldsymbol{x}_i), \Phi(\cdot,\boldsymbol{x}_j)\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle \Phi(\cdot,\boldsymbol{x}) - \sum_{j=1}^{N} u_j \Phi(\cdot,\boldsymbol{x}_j), \Phi(\cdot,\boldsymbol{x}) - \sum_{j=1}^{N} u_j \Phi(\cdot,\boldsymbol{x}_j)\rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \left\| \Phi(\cdot,\boldsymbol{x}) - \sum_{j=1}^{N} u_j \Phi(\cdot,\boldsymbol{x}_j) \right\|^2_{\mathcal{N}_\Phi(\Omega)} . \end{aligned} \tag{13}$$

Here we have used the definition of the native space norm from the previous section.

Then

Definition 5.7 *Suppose $\Omega \subseteq \mathbb{R}^s$ and $\Phi \in C(\Omega \times \Omega)$ is strictly positive definite on $\mathbb{R}^s$. For any distinct points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$ the* power function *is defined by*

$$[P_{\Phi,\mathcal{X}}(\boldsymbol{x})]^2 = Q(\boldsymbol{u}^*(\boldsymbol{x})),$$

where $\boldsymbol{u}^*$ is the vector of cardinal functions from Theorem 5.6.

The name *power function* was chosen by Schaback based on its connection to the power function of a statistical decision function [197]. In the paper [213] by Wu and Schaback the power function was referred to as *kriging function*. This terminology comes from geostatistics (see, e.g., [144]).

Now we can give a first generic error estimate. Contrary to our principles applied throughout the rest of this survey we will provide a proof of this theorem.

Theorem 5.8 *Let $\Omega \subseteq \mathbb{R}^s$, $\Phi \in C(\Omega \times \Omega)$ be strictly positive definite on $\mathbb{R}^s$, and suppose that the points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ are distinct. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X}$ by $\mathcal{P}f$. Then for every $\boldsymbol{x} \in \Omega$*

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \le P_{\Phi,\mathcal{X}}(\boldsymbol{x}) \|f\|_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: We express the interpolant in its cardinal form and apply the reproducing property of Φ. This gives us

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x})$$

$$
\begin{aligned}
&= \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \langle f, \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \\
&= \langle f, \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)}.
\end{aligned}
$$

For f the reproducing property of Φ yields

$$
f(\boldsymbol{x}) = \langle f, \Phi(\cdot, \boldsymbol{x}) \rangle_{\mathcal{N}_\Phi(\Omega)}.
$$

Now we combine these two formulas and apply the Cauchy-Schwarz inequality

$$
\begin{aligned}
|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| &= \left| \langle f, \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{N}_\Phi(\Omega)} \right| \\
&\leq \|f\|_{\mathcal{N}_\Phi(\Omega)} \left\| \Phi(\cdot, \boldsymbol{x}) - \sum_{j=1}^{N} u_j^*(\boldsymbol{x}) \Phi(\cdot, \boldsymbol{x}_j) \right\|_{\mathcal{N}_\Phi(\Omega)} \\
&= \|f\|_{\mathcal{N}_\Phi(\Omega)} P_{\Phi,\mathcal{X}}(\boldsymbol{x}),
\end{aligned}
$$

where we have applied (13) and the definition of the power function. □

Remark: One of the main benefits of Theorem 5.8 is that we are now able to estimate the interpolation error by considering two independent phenomena:

- the smoothness of the data (measured in terms of the native space norm of f – which is independent of the data locations),
- and the contribution due to the use of the basic function Φ and the distribution of the data (measured in terms of the power function – independent of the actual data values).

This is analogous to the standard error estimate for polynomial interpolation cited in most numerical analysis texts.

The next step is to refine this error bound by expressing the influence of the data locations in terms of the fill distance. And then, of course, the bound needs to be specialized to various choices of basic functions Φ.

The strategy to obtaining most error bounds in numerical analysis is to take advantage of the polynomial precision of a method (at least locally), and then to apply a Taylor expansion. We will not go into any of the details here, and instead state the resulting theorem that includes the more general case of strictly conditionally positive definite functions and also covers the error for approximating the derivatives of f by derivatives of $\mathcal{P}f$ (see [205] for details).

Theorem 5.9 *Suppose $\Omega \subseteq \mathbb{R}^s$ is open and bounded and satisfies an interior cone condition. Suppose $\Phi \in C^{2k}(\Omega \times \Omega)$ is symmetric and strictly conditionally positive definite of order m on $\mathbb{R}^s$. Denote the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on the $(m-1)$-unisolvent set $\mathcal{X}$ by $\mathcal{P}f$. Fix $\boldsymbol{\alpha} \in \mathbb{N}_0^s$ with $|\boldsymbol{\alpha}| \leq k$. Then there exist positive constant h_0 and C (independent of $\boldsymbol{x}$, f and Φ) such that*

$$
|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P}f(\boldsymbol{x})| \leq C C_\Phi(\boldsymbol{x})^{1/2} h_{\mathcal{X},\Omega}^{k-|\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)},
$$

provided $h_{\mathcal{X},\Omega} \le h_0$. *Here*

$$C_\Phi(\boldsymbol{x}) = \max_{\substack{\boldsymbol{\beta},\boldsymbol{\gamma}\in\mathbb{N}_0^s \\ |\boldsymbol{\beta}|+|\boldsymbol{\gamma}|=2k}} \max_{\boldsymbol{w},\boldsymbol{z}\in\Omega\cap B(\boldsymbol{x},c_2 h_{\mathcal{X},\Omega})} |D_1^{\boldsymbol{\beta}} D_2^{\boldsymbol{\gamma}} \Phi(\boldsymbol{w},\boldsymbol{z})|.$$

Remarks:

1. For infinitely smooth strictly positive definite functions such as the Gaussians and the inverse multiquadrics we see that the approximation order k is arbitrarily high.

2. For strictly positive definite functions with limited smoothness such as the Wendland functions $\varphi_{s,k}$ the approximation order is limited by the smoothness of the basic function.

3. The estimate in Theorem 5.9 is still generic, since it does not account for the particular basic function Φ being used for the interpolation.

4. We point out that the factor C_Φ may still depend on $h_{\mathcal{X},\Omega}$. For most basic functions it will be possible to use C_Φ to "squeeze out" additional powers of h. This is the reason for splitting the constant in front of the h-power into a generic C and a C_Φ.

The additional refinement of the error estimate of Theorem 5.9 for specific functions Φ is rather technical (for details see, e.g., [205]). We only list the final bounds for various functions Φ.

Application of Theorem 5.9 to infinitely smooth functions such as Gaussians or (inverse) multiquadrics immediately yields arbitrarily high algebraic convergence rates, i.e., for every $\ell \in \mathbb{N}$ and $|\boldsymbol{\alpha}| \le \ell$ we have

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P} f(\boldsymbol{x})| \le C_\ell h^{\ell-|\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)}.$$

whenever $f \in \mathcal{N}_\Phi(\Omega)$. Considerable amount of work has gone into investigating the dependence of the constant C_ℓ on ℓ. It is possible to show that for Gaussians $\Phi(\boldsymbol{x}) = e^{-\alpha\|\boldsymbol{x}\|^2}$, $\alpha > 0$, we get for some positive constant c that

$$\|f - \mathcal{P} f\|_{L_\infty(\Omega)} \le e^{\frac{-c|\log h_{\mathcal{X},\Omega}|}{h_{\mathcal{X},\Omega}}} \|f\|_{\mathcal{N}_\Phi(\Omega)} \tag{14}$$

provided $h_{\mathcal{X},\Omega}$ is sufficiently small and $f \in \mathcal{N}_\Phi(\Omega)$. The corresponding result for (inverse) multiquadrics $\Phi(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 + \alpha^2)^\beta$, $\alpha > 0$, $\beta < 0$, or $\beta > 0$ and $\beta \notin \mathbb{N}$, is

$$\|f - \mathcal{P} f\|_{L_\infty(\Omega)} \le e^{\frac{-c}{h_{\mathcal{X},\Omega}}} |f|_{\mathcal{N}_\Phi(\Omega)} \tag{15}$$

For functions with finite smoothness (such as powers, thin plate splines, and Wendland's compactly supported functions) it is possible to bound the constant $C_\Phi(\boldsymbol{x})$ and thereby to improve the order predicted by Theorem 5.9 by some additional powers of h. This results in the following error estimates.

For the powers $\Phi(\boldsymbol{x}) = (-1)^{\lceil \beta/2 \rceil} \|\boldsymbol{x}\|^\beta$, $\beta > 0$, $\beta \notin 2\,\mathbb{N}$, we get

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P} f(\boldsymbol{x})| \le C h^{\frac{\beta}{2}-|\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)}. \tag{16}$$

provided $|\boldsymbol{\alpha}| \leq \frac{\lceil \beta \rceil - 1}{2}$ and $f \in \mathcal{N}_\Phi(\Omega)$.

For thin plate splines $\Phi(\boldsymbol{x}) = (-1)^{k+1} \|\boldsymbol{x}\|^{2k} \log \|\boldsymbol{x}\|$, we get

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P} f(\boldsymbol{x})| \leq C h^{k-|\boldsymbol{\alpha}|} |f|_{\mathcal{N}_\Phi(\Omega)}. \tag{17}$$

provided $|\boldsymbol{\alpha}| \leq k-1$ and $f \in \mathcal{N}_\Phi(\Omega)$.

For Wendland's compactly supported functions $\Phi_{s,k}(\boldsymbol{x}) = \varphi_{s,k}(\|\boldsymbol{x}\|)$ this first refinement leads to

$$|D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} \mathcal{P} f(\boldsymbol{x})| \leq C h^{k+\frac{1}{2}-|\boldsymbol{\alpha}|} \|f\|_{\mathcal{N}_\Phi(\Omega)}. \tag{18}$$

provided $|\boldsymbol{\alpha}| \leq k$ and $f \in \mathcal{N}_\Phi(\Omega)$.

Remark: The convergence result for the compactly supported functions assumes that the support radius is kept fixed, and that only the domain Ω is filled out by adding more points to $\mathcal{X}$, and thus decreasing the fill distance $h_{\mathcal{X},\Omega}$. However, this means that for small fill distances (with fixed support radius) the system matrices of the interpolation problem become more and more dense – and thus the advantage of the compact support is lost. This point of view is referred to in the literature as the *non-stationary approach*. We are guaranteed convergence, but at the cost of increased computational complexity. Another possibility is presented by the *stationary approach*, for which we scale the support radius proportional to the fill distance. In this case the sparsity of the interpolation matrix remains fixed, however, convergence is lost. We will revisit this phenomenon later.

The powers and thin plate splines can be interpreted as a generalization of univariate natural splines. Therefore, one can see that the approximation order estimates obtained via the native space approach are not optimal. For example, for interpolation with thin plate splines $\Phi(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 \log \|\boldsymbol{x}\|$ one would expect order $\mathcal{O}(h^2)$, but the above estimate yields only $\mathcal{O}(h)$.

One can improve the estimates for functions with finite smoothness (i.e., powers, thin plate splines, Wendland's functions) by either (or both) of the following two ideas:

- by requiring the data function f to be even smoother than what the native space prescribes, i.e., by building certain boundary conditions into the native space;
- by using weaker norms to measure the error.

The idea to localize the data by adding boundary conditions was introduced in a paper by Light and Wayne [118]. This "trick" allows us to double the approximation order. The second idea can already be found in the early work by Duchon [46]. After applying both improvements the final approximation order estimate for interpolation with the compactly supported functions $\Phi_{s,k}$ is (see [201])

$$\|f - \mathcal{P} f\|_{L_2(\Omega)} \leq C h^{2k+1+s} \|f\|_{W_2^{2k+1+s}(\mathbb{R}^s)}, \tag{19}$$

where f is assumed to lie in the subspace $W_2^{2k+1+s}(\mathbb{R}^s)$ of $\mathcal{N}_\Phi(\mathbb{R}^s)$. For powers and thin plate splines one obtains L_2-error estimates of order $\mathcal{O}(h^{\beta+s})$ and $\mathcal{O}(h^{2k+s})$, respectively. These estimates are optimal, i.e., exact approximation orders, as shown by Bejancu [16].

Work on improved error bounds is also due to others such as Bejancu, Johnson, Powell, Ron, Schaback, and Yoon. In particular, recent work by Yoon provides L_p error estimates for so-called *shifted surface splines* for functions f is standard Sobolev spaces. These functions include all of the (inverse) multiquadrics, powers and thin plate splines. They are of the form

$$\Phi(\boldsymbol{x}) = \begin{cases} (-1)^{\lceil \beta - s/2 \rceil} (\|\boldsymbol{x}\|^2 + \alpha^2)^{\beta - s/2}, & s \text{ odd}, \\ (-1)^{\beta - s/2 + 1} (\|\boldsymbol{x}\|^2 + \alpha^2)^{\beta - s/2} \log(\|\boldsymbol{x}\|^2 + \alpha^2)^{1/2}, & s \text{ even}, \end{cases}$$

where $\beta \in \mathbb{N}$, $\beta > s/2$.

Yoon [214] has the following theorem that is formulated in the stationary setting.

Theorem 5.10 *Let Φ be a shifted surface spline with parameter α proportional to the fill distance $h_{\mathcal{X},\Omega}$. Then there exists a positive constant C (independent of $\mathcal{X}$) such that for every $f \in W_2^m(\Omega) \cap W_\infty^m(\Omega)$ we have*

$$\|f - \mathcal{P}f\|_{L_p(\Omega)} \leq C h^{\gamma_p} |f|_{W_2^m(\mathbf{R}^s)}, \qquad 1 \leq p \leq \infty,$$

with

$$\gamma_p = \min\{m, m - s/2 + s/p\}.$$

Furthermore, if $f \in W_2^k(\Omega) \cap W_\infty^k(\Omega)$ with $\max\{0, s/2 - s/p\} < k < m$, then

$$\|f - \mathcal{P}f\|_{L_p(\Omega)} = o(h^{\gamma_p - m + k}).$$

Remarks:

1. Using the localization idea mentioned above Yoon's estimates can be "doubled" to $\mathcal{O}(h^{m+\gamma_p})$.

2. Yoon's estimates address the question of how well the infinitely smooth (inverse) multiquadrics approximate functions that are less smooth than those in their native space. For example, Theorem 5.10 states that approximation to functions in $W_2^2(\Omega)$, $\Omega \subseteq \mathbb{R}^3$, by multiquadrics $\Phi(\boldsymbol{x}) = \sqrt{\|\boldsymbol{x}\|^2 + \alpha^2}$ is of the order $\mathcal{O}(h^2)$. However, it needs to be emphasized that this refers to stationary approximation, i.e., α is scaled proportional to the fill distance, whereas the spectral order given in (15) corresponds to the non-stationary case with fixed α. Similar numerical evidence was also provided much earlier by Schaback [167].

3. Moreover, the second part of Yoon's result is a step toward *exact* approximation orders.

5.4 The Connection to Optimal Recovery

In the paper [79] by Michael Golomb and Hans Weinberger the following general problem is studied: Given the values $f_1 = \lambda_1(f), \ldots, f_N = \lambda_N(f) \in \mathbb{R}$, where $\{\lambda_1, \ldots, \lambda_N\}$ is a linearly independent set of linear functionals (called *information functionals* yielding the *information* about f), how does one "best" approximate the value $\lambda(f)$ where λ is a given linear functional and f is unknown? The value $\lambda(f)$ is also referred to as a *feature* of f.

Remarks:

1. This is a very general problem formulation that allows not only for interpolation of function values, but also for other types of data (such as values of derivatives, integrals of f, moments of f, etc.), as well as other types of approximation.

2. Optimal recovery was also studied in detail by Micchelli, Rivlin and Winograd [136, 137, 138, 139].

In a Hilbert space setting the solution to this "optimal recovery problem" is shown to be the *minimum-norm interpolant*. More precisely, given $f_1 = \lambda_1(f), \ldots, f_N = \lambda_N(f) \in \mathbb{R}$ with $\{\lambda_1, \ldots, \lambda_N\} \subseteq \mathcal{H}^*$, the minimum-norm interpolant is that function $s^* \in \mathcal{H}$ that satisfies

$$\lambda_j(s^*) = f_j, \qquad j = 1, \ldots, N,$$

and for which

$$\|s^*\|_{\mathcal{H}} = \min_{\substack{s \in \mathcal{H} \\ \lambda_j(s) = f_j, j = 1, \ldots, N}} \|s\|_{\mathcal{H}}.$$

It turns out that the radial basis function interpolant satisfies these criteria if $\mathcal{H}$ is taken as the associated native space $\mathcal{N}_\Phi(\Omega)$. Since we are working in a Hilbert space, the following two orthogonality lemmas are easy to prove. They provide the essence of the two "optimality theorems" stated below.

Lemma 5.11 *Assume Φ is a symmetric strictly positive definite kernel on $\mathbb{R}^s$ and let $\mathcal{P}f = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j)$ be the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$. Then*

$$\langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$$

for all interpolants $s \in \mathcal{N}_\Phi(\mathcal{X})$, i.e., with $s(\boldsymbol{x}_j) = f(\boldsymbol{x}_j)$, $j = 1, \ldots, N$.

Lemma 5.12 *Assume Φ is a strictly positive definite kernel on $\mathbb{R}^s$ and let $\mathcal{P}f$ be the interpolant to $f \in \mathcal{N}_\Phi(\Omega)$ on $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$. Then*

$$\langle f - \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$$

for all $s \in H_\Phi(\mathcal{X}) = \{s = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j), \boldsymbol{x}_j \in \mathcal{X}\}$.

The following "energy splitting" theorem is an immediate consequence of Lemma 5.12. It says that the native space energy of f can be split into the energy of the interpolant $\mathcal{P}f$ and that of the residual $f - \mathcal{P}f$.

Corollary 5.13 *The orthogonality property of Lemma 5.12 implies the* energy split

$$\|f\|^2_{\mathcal{N}_\Phi(\Omega)} = \|f - \mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: The statement follows from

$$\begin{aligned} \|f\|^2_{\mathcal{N}_\Phi(\Omega)} &= \|f - \mathcal{P}f + \mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle (f - \mathcal{P}f) + \mathcal{P}f, (f - \mathcal{P}f) + \mathcal{P}f \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \|f - \mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} + 2\langle f - \mathcal{P}f, \mathcal{P}f \rangle_{\mathcal{N}_\Phi(\Omega)} + \|\mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} \end{aligned}$$

and the fact that $\langle f - \mathcal{P}f, \mathcal{P}f \rangle_{\mathcal{N}_\Phi(\Omega)} = 0$ by Lemma 5.12. □

Remark: The above energy split is the fundamental idea behind a number of Krylov-type iterative algorithms for approximately solving the interpolation problem when very large data sets are involved (see, e.g., the papers [66] and [67] by Faul and Powell or [178] by Schaback and Wendland mentioned in Section 8).

The following theorem shows the first optimality property of strictly conditionally positive definite kernels. It is taken from [205].

Theorem 5.14 *Suppose $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$ and that $\mathcal{X}$ is P-unisolvent. If the values $f_1, \ldots, f_N$ are given, then the interpolant $\mathcal{P}f$ is the minimum-norm interpolant to $\{f_j\}_{j=1}^N$, i.e.,*

$$|\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} = \min_{\substack{s \in \mathcal{N}_\Phi(\Omega) \\ s(\boldsymbol{x}_j) = f_j, j = 1, \ldots, N}} |s|_{\mathcal{N}_\Phi(\Omega)}.$$

Proof: We consider only the strictly positive definite case. Consider an arbitrary interpolant $s \in \mathcal{N}_\Phi(\Omega)$ to $f_1, \ldots, f_N$. Since $\mathcal{P}f \in \mathcal{N}_\Phi(\Omega)$ we can apply Lemma 5.11 and get

$$\langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} = 0.$$

Now

$$\begin{aligned} |\mathcal{P}f|^2_{\mathcal{N}_\Phi(\Omega)} &= \langle \mathcal{P}f, \mathcal{P}f - s + s \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle \mathcal{P}f, \mathcal{P}f - s \rangle_{\mathcal{N}_\Phi(\Omega)} + \langle \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &= \langle \mathcal{P}f, s \rangle_{\mathcal{N}_\Phi(\Omega)} \\ &\leq |\mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} |s|_{\mathcal{N}_\Phi(\Omega)} \end{aligned}$$

so that the statement follows. □

Remarks:

1. The space P mentioned in Theorem 5.14 is usually taken as a space of multivariate polynomials.

2. For thin plate splines $\phi(r) = r^2 \log r$, $r = \|\boldsymbol{x}\|_2$ with $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$, the corresponding semi-norm in the Beppo-Levi space $BL_2(\mathbb{R}^2)$ is

$$|f|^2_{BL_2(\mathbb{R}^2)} = \int_{\mathbb{R}^2} \left| \frac{\partial^2 f}{\partial x^2}(\boldsymbol{x}) \right|^2 + 2 \left| \frac{\partial^2 f}{\partial x \partial y}(\boldsymbol{x}) \right|^2 + \left| \frac{\partial^2 f}{\partial y^2}(\boldsymbol{x}) \right|^2 d\boldsymbol{x},$$

   which is the bending energy of a thin plate, and thus explains the name of these functions.

Another nice property of the radial basis function interpolant is that it is at the same time the best Hilbert-space approximation to the given data, and thus not just any projection of f but the *orthogonal projection*. More precisely,

Theorem 5.15 *Let*

$$H_\Phi(\mathcal{X}) = \{s = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j) + p \mid p \in P \text{ and } \sum_{j=1}^{N} c_j q(\boldsymbol{x}_j) = 0 \text{ for all } q \in P \text{ and } \boldsymbol{x}_j \in \mathcal{X}\},$$

where $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$ and $\mathcal{X}$ is P-unisolvent. If only the values $f_1 = f(\boldsymbol{x}_1), \ldots, f_N = f(\boldsymbol{x}_N)$ are given, then the interpolant $\mathcal{P}f$ is the best approximation to f from $H_\Phi(\mathcal{X})$ in $\mathcal{N}_\Phi(\Omega)$, i.e.,

$$|f - \mathcal{P}f|_{\mathcal{N}_\Phi(\Omega)} \leq |f - s|_{\mathcal{N}_\Phi(\Omega)}$$

for all $s \in H_\Phi(\mathcal{X})$.

Remarks:

1. The connection between radial basis function interpolation and the optimal recovery theory by Golomb and Weinberger was pointed out by various people (e.g., Schaback [167, 170], or Light and Wayne [118]).

2. These optimality properties of radial basis function interpolants play an important role in applications such as in the design of support vector machines in artificial intelligence or the numerical solutions of partial differential equations.

3. The optimality results above imply that one could also start with some Hilbert space $\mathcal{H}$ with norm $\|\cdot\|_\mathcal{H}$ and ask to find the minimum norm interpolant (i.e., Hilbert space best approximation) to some given data. In this way any given space defines a set of *optimal basis functions*, generated by the reproducing kernel of $\mathcal{H}$. This is how Duchon approached the subject in his papers [44, 45, 46, 47]. More recently, Kybic, Blu and Unser [106, 107] take this point of view and explain very nicely from a sampling theory point of view how the thin plate splines can be interpreted a fundamental solutions of the differential operator defining the semi-norm in the Beppo-Levi space $BL_2(\mathbb{R}^2)$, and thus radial basis functions can be viewed as *Green's functions*.

A third optimality result is in the context of quasi-interpolation, i.e.,

Theorem 5.16 *Suppose $\Phi \in C(\Omega \times \Omega)$ is a strictly conditionally positive definite kernel with respect to the finite-dimensional space $P \subseteq C(\Omega)$. Suppose $\mathcal{X}$ is P-unisolvent and $\boldsymbol{x} \in \Omega$ is fixed. Let $u_j^*(\boldsymbol{x})$, $j = 1, \ldots, N$, be the cardinal basis functions for interpolation with Φ. Then*

$$\left| f(\boldsymbol{x}) - \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j^*(\boldsymbol{x}) \right| \leq \left| f(\boldsymbol{x}) - \sum_{j=1}^{N} f(\boldsymbol{x}_j) u_j \right|$$

for all choices of $u_1, \ldots, u_N \in \mathbb{R}$ with $\sum_{j=1}^{N} u_j p(\boldsymbol{x}_j) = p(\boldsymbol{x})$ for any $p \in P$.

Theorem 5.16 is proved in [205]. It says in particular that the minimum norm interpolant $\mathcal{P}f$ is also more accurate (in the pointwise sense) than any linear combination of the given data values.

6 Least Squares Approximation

As we saw in Section 5 we can interpret radial basis function interpolation as a constrained optimization problem. We now take this point of view again, but start with a more general formulation. Let's assume we are seeking a function $\mathcal{P}f$ of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{x}_j), \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

such that the quadratic form

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} \tag{20}$$

with $\boldsymbol{c} = [c_1, \ldots, c_M]^T$ and some symmetric positive definite matrix Q is minimized subject to the linear constraints

$$A\boldsymbol{c} = \boldsymbol{f} \tag{21}$$

where A is an $N \times M$ matrix, and the right-hand side $\boldsymbol{f} = [f_1, \ldots, f_N]^T$ is given. Such a constrained quadratic minimization problem can be converted to a system of linear equations by introducing *Lagrange multipliers*, i.e., we consider finding the minimum of

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} - \boldsymbol{\lambda}^T \left[A\boldsymbol{c} - \boldsymbol{f}\right] \tag{22}$$

with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$. Since Q is a positive definite matrix, it is well known that the functional to be minimized is convex, and thus has a unique minimum. Therefore, the standard necessary condition for such a minimum (which is obtained by differentiating with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda}$ and finding the zeros of those derivatives) is also sufficient. This leads to

$$\begin{aligned} Q\boldsymbol{c} - A^T\boldsymbol{\lambda} &= \boldsymbol{0} \\ A\boldsymbol{c} - \boldsymbol{f} &= \boldsymbol{0} \end{aligned}$$

or, in matrix form,

$$\begin{bmatrix} Q & -A^T \\ A & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{f} \end{bmatrix}.$$

By applying Gaussian elimination to this block matrix (Q is invertible since it is assumed to be positive definite) we get

$$\boldsymbol{\lambda} = \left(AQ^{-1}A^T\right)^{-1} \boldsymbol{f} \tag{23}$$

$$\boldsymbol{c} = Q^{-1}A^T \left(AQ^{-1}A^T\right)^{-1} \boldsymbol{f}. \tag{24}$$

In particular, if the quadratic form represents the native space norm of the interpolant $\mathcal{P}f = \sum_{j=1}^{M} c_j \Phi(\cdot, \boldsymbol{x}_j)$, i.e.,

$$\|\mathcal{P}f\|^2_{\mathcal{N}_\Phi(\Omega)} = \sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{c}^T Q \boldsymbol{c}$$

with $Q_{ij} = \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{c} = [c_1, \ldots, c_M]^T$, and the linear side conditions are the interpolation conditions

$$A\boldsymbol{c} = \boldsymbol{f} \qquad \Longleftrightarrow \qquad \mathcal{P}f(\boldsymbol{x}_i) = f_i, \quad i = 1, \ldots, M,$$

with $A = Q$ (symmetric) and the same $\boldsymbol{c}$ as above and data vector $\boldsymbol{f} = [f_1, \ldots, f_M]^T$, then we see that the Lagrange multipliers (23) become

$$\boldsymbol{\lambda} = A^{-T}\boldsymbol{f} = A^{-1}\boldsymbol{f}$$

and the coefficients are given by

$$\boldsymbol{c} = \boldsymbol{\lambda}$$

via (24). Therefore, as we saw earlier, the minimum norm interpolant is obtained by solving the interpolation equations alone.

Since we took the more general point of view that $\mathcal{P}$ is generated by M basis functions, and N linear constraints are specified, the above formulation also covers both over- and under-determined least squares fitting where the quadratic form $\boldsymbol{c}^T Q \boldsymbol{c}$ represents an added *smoothing* (or *regularization*) term. This term is not required to obtain a unique solution of the system $A\boldsymbol{c} = \boldsymbol{f}$ in the over-determined case ($M \leq N$), but in the under-determined case such a constraint is needed (cf. the solution of under-determined linear systems via singular value decomposition in the numerical linear algebra literature (e.g., [193])).

Usually the regularized least squares approximation problem is formulated as minimization of

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} + \omega \sum_{j=1}^{N} \left(\mathcal{P}f(\boldsymbol{x}_j) - f_j\right)^2 . \tag{25}$$

The quadratic form controls the smoothness of the fitting function and the least squares term measures the closeness to the data. The parameter ω controls the tradeoff between these two terms. The formulation (25) is used in *regularization theory* (see, e.g., [52, 76]). The same formulation is also used in *penalized least squares* fitting (see, e.g., [78]), the literature on smoothing splines [160, 182], and in papers by Wahba on thin plate splines (e.g., [195, 196]). In fact, the idea of smoothing a data fitting process by this kind of formulation seems to go back to at least Whittaker [207] in 1923. In practice a penalized least squares formulation is especially useful if the data f_i cannot be completely trusted, i.e., it is contaminated by noise. In this case, a (penalized) least squares fit is advisable. The problem of minimizing (25) is known as *ridge regression* in the statistics literature.

The equivalence with our formulation (22) above follows from

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} + \omega \sum_{j=1}^{N} \left(\mathcal{P}f(\boldsymbol{x}_j) - f_j\right)^2 \;\; = \;\; \frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} + \omega [A\boldsymbol{c} - \boldsymbol{f}]^T [A\boldsymbol{c} - \boldsymbol{f}]$$

$$= \frac{1}{2}\boldsymbol{c}^T Q\boldsymbol{c} - \boldsymbol{\lambda}^T[A\boldsymbol{c} - \boldsymbol{f}],$$

where

$$\boldsymbol{\lambda} = -\omega[A\boldsymbol{c} - \boldsymbol{f}].$$

We are now interested in the more general setting where we still sample the given function f on the set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, but now introduce a second set $\Xi = \{\boldsymbol{\xi}_i\}_{i=1}^M$ at which we center the basis functions. Usually we will have $M \leq N$, and the case $M = N$ with $\Xi = \mathcal{X}$ recovers the traditional interpolation setting discussed thus far. Therefore,

$$\mathcal{Q}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j), \quad \boldsymbol{x} \in \mathbb{R}^s, \tag{26}$$

and the coefficients c_j can be found by minimizing $\|Qf - f\|_2^2$, where the ℓ_2-norm

$$\|f\|_2^2 = \sum_{i=1}^{N} [f(\boldsymbol{x}_i)]^2$$

is induced by the discrete inner product

$$\langle f, g \rangle = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i). \tag{27}$$

This approximation problem has a unique solution if the (rectangular) *collocation matrix* A with entries

$$A_{jk} = \Phi(\boldsymbol{x}_j, \boldsymbol{\xi}_k), \quad j = 1, \ldots, N, \ k = 1, \ldots, M,$$

has full rank.

Remarks:

1. If the centers in Ξ are chosen to form a subset of the data locations $\mathcal{X}$ then A has full rank provided the radial basis function is selected according to our previous sections on interpolation. This is true, since in this case A will have an $M \times M$ square submatrix which is non-singular (by virtue of being an *interpolation matrix*).

2. The over-determined linear system $A\boldsymbol{c} = \boldsymbol{y}$ which arises in the solution of the least squares problem can be solved using standard algorithms from numerical linear algebra such as QR or singular value decomposition.

In the following section we give a brief account of theoretical results known for the general problem in which the centers and data sites differ.

6.1 Theoretical Results

The results mentioned here are due to Sivakumar and Ward [187], and Quak, Sivakumar and Ward [159]. The first paper deals with discrete least squares, the second with continuous least squares approximation. In both papers the authors do not discuss the collocation matrix A above, but rather base their results on the non-singularity of the coefficient matrix obtained from a system of normal equations. In the discrete setting they use the inner product (27) which induces the ℓ_2 norm, and then discuss non-singularity of the *Gramian* which occurs in the following system of *normal equations*

$$G\boldsymbol{c} = \boldsymbol{w}, \tag{28}$$

where the entries of G are the ℓ_2 inner products of the radial basis functions, i.e.,

$$G_{jk} = \langle \Phi(\cdot, \boldsymbol{\xi}_j), \Phi(\cdot, \boldsymbol{\xi}_k) \rangle = \sum_{i=1}^{N} \Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j)\Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_k), \quad j, k = 1, \ldots, M,$$

and the right-hand side vector $\boldsymbol{w}$ in (28) is given by

$$\boldsymbol{w}_j = \langle \Phi(\cdot, \boldsymbol{\xi}_j), \boldsymbol{f} \rangle = \sum_{i=1}^{N} \Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j) f(\boldsymbol{x}_i), \quad j = 1, \ldots, M.$$

Remarks:

1. Note that in the interpolation case with $M = N$ and $\Xi = \mathcal{X}$ we have

   $$\langle \Phi(\cdot, \boldsymbol{x}_j), \Phi(\cdot, \boldsymbol{x}_k) \rangle = \langle \Phi(\cdot, \boldsymbol{x}_j), \Phi(\cdot, \boldsymbol{x}_k) \rangle_{\mathcal{N}_\Phi(\Omega)} = \Phi(\boldsymbol{x}_j, \boldsymbol{x}_k)$$

   so that G is just the interpolation matrix A.

2. Of course, this also presents an interpretation of the interpolation matrix A as the system matrix for the normal equations in the case of best approximation with respect to the native space norm – a fact already mentioned earlier in the section on optimal recovery.

In both papers, [187] as well as [159], even the formulation of the main theorems is very technical. We therefore just try to give a feel for their results.

Essentially, the authors show that the Gramian for certain radial basis functions (norm, (inverse) multiquadrics, and Gaussians) is non-singular if the centers $\boldsymbol{\xi}_k$, $k = 1, \ldots, M$, are sufficiently well distributed and the data points $\boldsymbol{x}_j$, $j = 1, \ldots, N$, are fairly evenly clustered about the centers with the diameter of the clusters being relatively small compared to the separation distance of the data points. Figure 2 illustrates the clustering idea.

As a by-product of this argumentation the authors obtain a proof for the non-singularity of *interpolation* matrices for the case in which the centers of the basis functions are chosen different from the data sites, namely as small perturbations thereof.

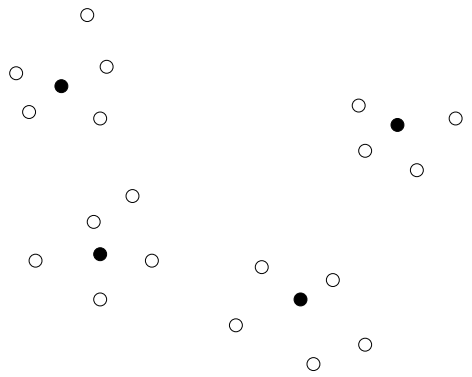

Figure 2: Clusters of data points ∘ around well separated centers •.

7 Moving Least Squares Approximation

An alternative to radial basis function interpolation and approximation is the so-called *moving least squares* method. As we will see below, in this method the approximation $\mathcal{P}f$ to f is obtained by solving many (small) linear systems, instead of via solution of a single – but large – linear system as we did in the previous sections.

To establish a connection with the previous sections we start with the Backus-Gilbert formulation of the moving least squares method since this corresponds to a linearly constrained quadratic minimization problem.

7.1 Moving Least Squares Approximation: The Backus-Gilbert Approach

The connection between the standard moving least squares formulation (to be explained in the next section) and Backus-Gilbert theory was pointed out by Bos and Šalkauskas in [21]. Mathematically, in the Backus-Gilbert approach one considers a quasi-interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}), \tag{29}$$

where $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ represents the given data. From Theorem 5.16 we know that the quasi-interpolant that minimizes the point-wise error is given if the *generating functions* Ψ_i are cardinal functions, i.e., $\Psi_i(\boldsymbol{x}_j) = \delta_{ij}$, $i, j = 1, \ldots, N$.

In the moving least squares method one does not attempt to minimize the pointwise error, but instead seeks to find the values of the generating functions $\Psi_i(\boldsymbol{x}) = \Psi(\boldsymbol{x}, \boldsymbol{x}_i)$ by minimizing

$$\frac{1}{2}\sum_{i=1}^{N} \Psi_i^2(\boldsymbol{x})\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_i)} \tag{30}$$

subject to the polynomial reproduction constraints

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s, \tag{31}$$

where Π_d^s is the space of s-variate polynomials of total degree at most d which has dimension $m = \binom{s+d}{d}$.

Remarks:

1. In the above formulation there is no explicit emphasis on nearness of fit as this is implicitly obtained by the quasi-interpolation "ansatz" and the closeness of the generating functions to the pointwise optimal delta functions. This is achieved by the above problem formulation if the $W(\cdot, \boldsymbol{x}_i)$ are weight functions that decrease with distance from the origin. Many of the radial functions used earlier are candidates for the weight functions. However, strict positive definiteness is not required, so that, e.g., (radial or tensor product) B-splines can also be used. The polynomial reproduction constraint is added so that the quasi-interpolant will achieve a desired approximation order. This will become clear in Section 7.4 below.

2. The smoothness functional (30) used here is also motivated by practical applications. In the Backus-Gilbert theory which was developed in the context of geophysics (see [6]) it is desired that the generating functions Ψ_i are as close as possible to the ideal cardinal functions (i.e., delta functions). Therefore, one needs to minimize their "spread". The polynomial reproduction constraints correspond to *discrete moment conditions* for the function $\Psi = \Psi(\boldsymbol{x}, \cdot)$.

If we think of $\boldsymbol{x}$ as a fixed (evaluation) point, then we have another constrained quadratic minimization problem of the form discussed in previous sections. The unknowns are collected in the "coefficient vector" $\Psi(\boldsymbol{x}) = [\Psi(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, \Psi(\boldsymbol{x}, \boldsymbol{x}_N)]^T$. The smoothness functional (30)

$$\frac{1}{2}\Psi(\boldsymbol{x})^T Q(\boldsymbol{x})\Psi(\boldsymbol{x})$$

is given via the diagonal matrix

$$Q(\boldsymbol{x}) = \operatorname{diag}\left(\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_1)}, \ldots, \frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_N)}\right), \tag{32}$$

where $W(\cdot, \boldsymbol{x}_i)$ are positive *weight functions* (and thus for any $\boldsymbol{x}$ the matrix $Q(\boldsymbol{x})$ is positive definite).

The linear polynomial reproduction constraint (31) can be written in matrix form as

$$A\Psi(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}),$$

where A is the $m \times N$ matrix with entries $A_{ji} = p_j(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, $j = 1, \ldots, m$, and $\boldsymbol{p} = [p_1, \ldots, p_m]^T$ is a vector that contains a basis for the space Π_d^s of polynomials of degree d.

According to our earlier work we use Lagrange multipliers and then know that (cf. (23) and (24))

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \left(AQ^{-1}(\boldsymbol{x})A^T\right)^{-1}\boldsymbol{p}(\boldsymbol{x}) \tag{33}$$

$$\Psi(\boldsymbol{x}) = Q^{-1}(\boldsymbol{x})A^T\boldsymbol{\lambda}(\boldsymbol{x}). \tag{34}$$

Equation (33) implies that the Lagrange multipliers are obtained as the solution of a Gram system

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}),$$

where the entries of G are the weighted ℓ_2 inner products

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i) p_k(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j,k = 1, \ldots, m. \tag{35}$$

The special feature here is that the weight varies with the evaluation point $\boldsymbol{x}$.

Two short comments are called for. First, the Gram matrix is symmetric and positive definite since the polynomial basis is linearly independent and the weights are positive. Second, in practice, the polynomials will be represented in shifted form, i.e., centered at the point of evaluation $\boldsymbol{x}$, so that only $p_1(\boldsymbol{x}) \equiv 1 \neq 0$.

Equation (34) can be written componentwise, i.e., the generating functions in (29) are given by

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i) \sum_{j=1}^{m} \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

Therefore, once the values of the Lagrange multipliers $\lambda_j(\boldsymbol{x})$, $j = 1, \ldots, N$, have been determined we have explicit formulas for the values of the generating functions. In general, however, finding the Lagrange multipliers involves solving a (small) linear system that changes as soon as $\boldsymbol{x}$ changes.

7.2 Standard Interpretation of MLS Approximation

The standard interpretation of moving least squares method first appeared in the approximation theory literature in a paper by Lancaster and Šalkauskas [108] who also pointed out the connection to earlier (more specialized) work by Shepard [185] and McLain [129]. We now consider the following approximation problem. Assume we are given data values $f(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, on some set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^s$ of distinct data sites, where f is some (smooth) function, as well as an approximation space $\mathcal{U} = \text{span}\{u_1, \ldots, u_m\}$ (with $m < N$), along with the same weighted ℓ_2 inner product

$$\langle f, g \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i) W_i(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s \text{ fixed}, \tag{36}$$

as introduced above in (35). Again, the positive weights W_i, $i = 1, \ldots, N$, depend on the evaluation point $\boldsymbol{x}$. We will interpret the weight functions in a way similar to radial basis functions, i.e., $W_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)$, with the points $\boldsymbol{x}_i$ coming from the set $\mathcal{X}$.

We now wish to find the best approximation from $\mathcal{U}$ to f at the point $\boldsymbol{x}$ with respect to the norm induced by (36). This means we will obtain the approximation (at the point $\boldsymbol{x}$) as

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) u_j(\boldsymbol{x}), \tag{37}$$

where the coefficients $c_j(\boldsymbol{x})$ are such that

$$\sum_{i=1}^{N} \left[\mathcal{P}f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)\right]^2 W_i(\boldsymbol{x}) \tag{38}$$

is minimized. Due to the definition of the inner product (36) whose weight function "moves" with the evaluation point $\boldsymbol{x}$, the coefficients c_j in (37) depend also on $\boldsymbol{x}$. As a consequence one has to solve the normal equations

$$\sum_{j=1}^{m} c_j(\boldsymbol{x}) \langle u_j, u_k \rangle_{W(\boldsymbol{x})} = \langle f, u_k \rangle_{W(\boldsymbol{x})}, \qquad k = 1, \ldots, m, \tag{39}$$

anew each time the evaluation point $\boldsymbol{x}$ is changed. In matrix notation (39) becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_u(\boldsymbol{x}), \tag{40}$$

with the positive definite Gram matrix $G(\boldsymbol{x}) = \left(\langle u_j, u_k \rangle_{W(\boldsymbol{x})}\right)_{j,k=1}^{m}$, coefficient vector $\boldsymbol{c}(\boldsymbol{x})$ and right-hand side vector $\boldsymbol{f}_u(\boldsymbol{x})$ as in (39) all depending on $\boldsymbol{x}$.

In the moving least squares method one usually takes $\mathcal{U}$ to be a space of (multivariate) polynomials, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) p_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{41}$$

where the $\{p_1, \ldots, p_m\}$ is a basis for the space Π_d^s of s-variate polynomials of degree d.

The Gram system (40) now becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}), \tag{42}$$

where the matrix $G(\boldsymbol{x})$ has entries

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i) p_k(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j, k = 1, \ldots, m, \tag{43}$$

and the right-hand side vector consists of the projections of the data f onto the basis functions, i.e.,

$$\boldsymbol{f}_p(\boldsymbol{x}) = \left[\langle f, p_1 \rangle_{W(\boldsymbol{x})}, \ldots, \langle f, p_m \rangle_{W(\boldsymbol{x})}\right]^T.$$

Remarks:

1. The fact that the coefficients depend on the evaluation point $\boldsymbol{x}$, and thus for every evaluation of $\mathcal{P}f$ a Gram system (with different matrix $G(\boldsymbol{x})$) needs to be solved, initially scared people away from the moving least squares approach. However, one can either choose compactly supported weight functions so that only a few terms are "active" in the sum in (43), or even completely avoid the solution of linear systems (see, e.g., [60]).

2. We point out that since we are working with a polynomial basis, the matrix G can also be interpreted as a moment matrix for the weight W. This interpretation is used in the engineering literature (see, e.g., [116]), and also plays an essential role when connecting moving least squares approximation to the more efficient concept of *approximate approximation* [127]. For a discussion of approximate moving least squares approximation see [61, 62, 63, 64].

The connection to the constrained quadratic minimization problems discussed earlier can be seen as follows. We are now minimizing (for fixed $\boldsymbol{x}$)

$$\frac{1}{2}\boldsymbol{c}^T(\boldsymbol{x})G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{\mu}^T(\boldsymbol{x})\left[G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) - AQ^{-1}(\boldsymbol{x})\boldsymbol{f}\right], \tag{44}$$

where $G(\boldsymbol{x})$ is the Gram matrix (35), $Q(\boldsymbol{x})$ the diagonal matrix of weight functions (32) and A the matrix of polynomials used earlier. The term $AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$ corresponds to the right-hand side vector $\boldsymbol{f}_p(\boldsymbol{x})$ of (42). The solution of the linear system resulting from the minimization problem (44) gives us

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{x}) &= \left(G(\boldsymbol{x})G^{-1}(\boldsymbol{x})G^T(\boldsymbol{x})\right)^{-1} AQ^{-1}(\boldsymbol{x})\boldsymbol{f} = G^{-T}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}\\ \boldsymbol{c}(\boldsymbol{x}) &= G^{-1}(\boldsymbol{x})G^T(\boldsymbol{x})\boldsymbol{\mu}(\boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{x})\end{aligned}$$

so that – as in the case of radial basis function interpolation – by solving only the Gram system $G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x})$ we automatically minimize the functional

$$\begin{aligned}\boldsymbol{c}^T(\boldsymbol{x})G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) &= \sum_{j=1}^{m}\sum_{k=1}^{m} c_j(\boldsymbol{x})c_k(\boldsymbol{x})G_{jk}(\boldsymbol{x})\\ &= \sum_{j=1}^{m}\sum_{k=1}^{m} c_j(\boldsymbol{x})c_k(\boldsymbol{x})\langle p_j, p_k\rangle_{W(\boldsymbol{x})}\end{aligned}$$

which should be interpreted as the native space norm of the approximant $\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x})p_j(\boldsymbol{x})$.

Since one can show that the two approaches are equivalent the moving least squares approximant has all of the following properties:

- It reproduces any polynomial of degree at most d in s variables exactly.
- It produces the best locally weighted least squares fit.
- Viewed as a quasi-interpolant, the generating functions Ψ_i are as close as possible to the optimal cardinal basis functions in the sense that (30) is minimized.
- Since polynomials are infinitely smooth, either of the representations of $\mathcal{P}f$ shows that its smoothness is determined by the smoothness of the weight function(s) $W_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)$.

In particular, the standard moving least squares method will reproduce the polynomial basis functions $p_1, \ldots, p_m$ even though this is not explicitly enforced by the minimization (solution of the normal equations). Moreover, the more general "ansatz" with approximation space $\mathcal{U}$ allows us to build moving least squares approximations that also reproduce any other function that is included in $\mathcal{U}$. This can be very beneficial for the solution of partial differential equations with known singularities (see, e.g., the papers [5] by Babuška and Melenk, and [17] by Belytschko and co-authors).

7.3 An Example: Shepard's Method

The moving least squares method in the case $m = 1$ with $p_1(\boldsymbol{x}) \equiv 1$ is known to yield Shepard's method [185]. In the statistics literature Shepard's method is known as a *kernel method* (see, e.g., the papers from the 1950s and 60s [162, 153, 145, 198]). Using our notation we have

$$\mathcal{P}f(\boldsymbol{x}) = c_1(\boldsymbol{x}).$$

The Gram "matrix" consists of only one element

$$G(\boldsymbol{x}) = \langle p_1, p_1 \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)$$

so that $G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x})$ implies

$$c_1(\boldsymbol{x}) = \frac{\displaystyle\sum_{i=1}^{N} f(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i)}{\displaystyle\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

The values of the Lagrange multiplier can be found from $G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x})$ so that

$$\lambda_1(\boldsymbol{x}) = \frac{1}{\displaystyle\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

Finally, the generating functions are defined as

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)\lambda_1(\boldsymbol{x})p_1(\boldsymbol{x}_i) = \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\displaystyle\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

This gives rise to the well-known quasi-interpolation formula for Shepard's method

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i) \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\displaystyle\sum_{k=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_k)}.$$

Of course this is the same as the basis expansion $c_1(\boldsymbol{x})$.

It is interesting to note that the polynomial basis function and the Lagrange multiplier are orthogonal on the data, i.e., $\langle \lambda_1, p_1 \rangle_{W(\boldsymbol{x})} = 1$. Indeed

$$\begin{aligned} \langle \lambda_1, p_1 \rangle_{W(\boldsymbol{x})} &= \sum_{i=1}^{m} \lambda_1(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{k=1}^{N} W(\boldsymbol{x}_i, \boldsymbol{x}_k)}, \end{aligned}$$

and this equals 1 if we restrict $\boldsymbol{x}$ to be an element of the set $\mathcal{X}$. This fact can be shown to hold for higher-order MLS methods also.

7.4 Approximation Order of Moving Least Squares

Since the moving least squares approximants can be written as quasi-interpolants, we can use standard techniques to derive their point-wise error estimates. The standard argument proceeds as follows. Let f be a given (smooth) function that generates the data, i.e., $f_1 = f(\boldsymbol{x}_1), \ldots, f_N = f(\boldsymbol{x}_N)$, and let p be an arbitrary polynomial. Moreover, assume that the moving least squares approximant is given in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x})$$

with the generating functions Ψ_i satisfying the polynomial reproduction property

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s,$$

as described at the beginning of this section. Then, due to the polynomial reproduction property of the generating functions,

$$\begin{aligned} |f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| &\leq |f(\boldsymbol{x}) - p(\boldsymbol{x})| + |p(\boldsymbol{x}) - \sum_{i=1}^{N} f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x})| \\ &= |f(\boldsymbol{x}) - p(\boldsymbol{x})| + |\sum_{i=1}^{N} p(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x}) - \sum_{i=1}^{N} f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x})| \\ &\leq |f(\boldsymbol{x}) - p(\boldsymbol{x})| + \sum_{i=1}^{N} |p(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)| |\Psi_i(\boldsymbol{x})| \\ &\leq \|f - p\|_\infty \left[1 + \sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})| \right]. \end{aligned} \tag{45}$$

We see that in order to refine the error estimate we now have to answer two questions:

- How well do polynomials approximate f? This will be done with standard Taylor expansions.
- Are the generating functions bounded? The expression $\sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})|$ is known as the *Lebesgue function*, and finding a bound for the Lebesgue function is the main task that remains.

By taking the polynomial p above to be the Taylor polynomial for f at $\boldsymbol{x}$ of total degree d, the remainder term immediately yields an estimate of the form

$$\begin{aligned} \|f - p\|_\infty &\leq C_1 h^{d+1} \max_{\boldsymbol{x}\in\Omega} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|, \qquad |\boldsymbol{\alpha}| = d+1, \\ &= C_1 h^{d+1} |f|_{d+1}, \end{aligned} \tag{46}$$

where we have used the abbreviation

$$|f|_{d+1} = \max_{\boldsymbol{x}\in\Omega} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|, \qquad |\boldsymbol{\alpha}| = d+1.$$

Thus, if we can establish a uniform bound for the Lebesgue function, then (45) and (46) will result in

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq C h^{d+1} |f|_{d+1}$$

which shows that moving least squares approximation with polynomial reproduction of degree d has approximation order $\mathcal{O}(h^{d+1})$.

For Shepard's method, i.e., moving least squares approximation with constant reproduction (i.e., $m = 1$ or $d = 0$), we saw above that the generating functions are of the form

$$\Psi_i(\boldsymbol{x}) = \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{j=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_j)}$$

and therefore the Lebesgue function admits the uniform bound

$$\sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})| = 1,$$

This shows that Shepard's method has approximation order $\mathcal{O}(h)$.

Bounding the Lebesgue function in the general case is more involved and is the subject of the papers [113] by Levin and [204] by Wendland. This results in approximation order $\mathcal{O}(h^{d+1})$ for a moving least squares method that reproduces polynomials of degree d. In both papers the authors assumed that the weight function is compactly supported, and that the support size is scaled proportional to the fill distance. However, similar estimates should be possible if the weight function only decays fast enough (see, e.g., the survey by de Boor [20]).

Aside from this consideration, the choice of weight function W does not play a role in determining the approximation order of the moving least squares method. As noted earlier, it only determines the smoothness of $\mathcal{P}f$. For example, in the paper [37] from

the statistics literature on local regression the authors state that often "the choice [of weight function] is not too critical", and the use of the so-called *tri-cube*

$$W(\boldsymbol{x}, \boldsymbol{x}_i) = (1 - \|\boldsymbol{x} - \boldsymbol{x}_i\|^3)_+^3, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

is suggested. Of course, many other weight functions such as (radial) B-splines or any of the (bell-shaped) radial basis functions studied earlier can also be used. If the weight function is compactly supported, then the generating functions Ψ_i will be so, too. This leads to computationally efficient methods since the Gram matrix $G(\boldsymbol{x})$ will be sparse.

An interesting question is also the size of the support of the different local weight functions. Obviously, a fixed support size for all weight functions is possible. However, this will cause serious problems as soon as the data are not uniformly distributed. Therefore, in the arguments in [113] and [204] the assumption is made that the data are at least quasi-uniformly distributed. Another choice for the support size of the individual weight functions is based on the number of nearest neighbors, i.e., the support size is chosen so that each of the local weight functions contains the same number of centers in its support. A third possibility is suggested by Schaback [174]. He proposes to use another moving least squares approximation based on (equally spaced) auxiliary points to determine a smooth function δ so that at each evaluation point $\boldsymbol{x}$ the radius of the support of the weight function is given by $\delta(\boldsymbol{x})$. However, convergence estimates for these latter two choices do not exist.

Sobolev error estimates are provided for moving least squares approximation with compactly supported weight functions in [1]. The rates obtained in that paper are not in terms of the fill distance but instead in terms of the support size R of the weight function. Moreover, it is assumed that for general s and $m = \binom{s+d}{d}$ the local Lagrange functions are bounded. As mentioned above, this is the hard part, and such bounds are only provided in the case $s = 2$ with $d = 1$ and $d = 2$ in [1]. However, if combined with the general bounds for the Lebesgue function provided by Wendland the paper [1] yields the following estimates:

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \le CR^{d+1}|f|_{d+1}$$

but also

$$|\nabla(f - \mathcal{P}f)(\boldsymbol{x})| \le CR^{d}|f|_{d+1}.$$

In the weaker L_2-norm we have

$$\|f - \mathcal{P}f\|_{L_2(B_j \cap \Omega)} \le CR^{d+1}|f|_{W_2^{d+1}(B_j \cap \Omega)}$$

and

$$\|\nabla(f - \mathcal{P}f)\|_{L_2(B_j \cap \Omega)} \le CR^{d}|f|_{W_2^{d+1}(B_j \cap \Omega)},$$

where the balls B_j provide a finite cover of the domain Ω, i.e., $\Omega \subseteq \bigcup_j B_j$, and the number of overlapping balls is bounded.

Remarks:

1. In the statistics literature the moving least squares idea is known as local (polynomial) regression. There is a book by Fan and Gijbels [53] and a review article by Cleveland and Loader [37] according to which the basic ideas of local regression can be traced back at least to work of Gram [80], Woolhouse [210], and De Forest [38, 39] from the 1870s and 1880s.

2. In particular, in the statistics literature one learns that the use of least squares approximation is justified when the data $f_1, \ldots, f_N$ are normally distributed, whereas, if the noise in the data is not Gaussian, then other criteria should be used. See, e.g., the survey article [37] for more details.

3. Early error estimates for some special cases were provided by Farwig in [54, 55].

8 Some Issues Related to Practical Implementations

In this section we will collect some information about issues that are important for the practical use of radial basis function and moving least squares methods. These issues include stability and conditioning of radial basis function interpolants, the *trade-off principle* which explains the trade-off between achievable convergence rates and numerical stability or efficiency, as well as algorithms for fast solution and evaluation of radial basis interpolants and moving least squares approximants.

8.1 Stability and Conditioning of Radial Basis Function Interpolants

A standard criterion to measure the numerical stability of an approximation method is its condition number. In particular, for radial basis function interpolation we need to look at the condition number of the interpolation matrix A with entries $A_{ij} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$. For any matrix A the ℓ_2-condition number of A is given by

$$\mathrm{cond}(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}},$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and smallest singular values of A. If we concentrate on positive definite matrices A, then we can also take the ratio

$$\frac{\lambda_{\max}}{\lambda_{\min}}$$

of largest and smallest eigenvalues as an indicator for the condition number of A.

What do we know about these eigenvalues? First, Gershgorin's Theorem says that

$$|\lambda_{\max} - A_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{N} |A_{ij}|.$$

Therefore,

$$\lambda_{\max} \leq N \max_{i,j=1,\ldots,N} |A_{ij}| = N \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}} \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j),$$

which, since Φ is strictly positive definite, becomes

$$\lambda_{\max} \leq N\Phi(\mathbf{0}).$$

Now, as long as the data are not too wildly distributed, N will grow as $h_{\mathcal{X},\Omega}^{-s}$ which is acceptable. Therefore, the main work in establishing a bound for the condition number of A lies in finding lower bounds for $\lambda_{\min}$ (or correspondingly upper bounds

for the norm of the inverse $\|A^{-1}\|_2$). This is the subject of several papers by Ball, Narcowich, Sivakumar and Ward [8, 147, 148, 149, 150] who make use of a result by Ball [7] on eigenvalues of distance matrices. Ball's result follows from the Rayleigh quotient, which gives the smallest eigenvalue of a positive definite matrix as

$$\lambda_{\min} = \inf_{\boldsymbol{c} \in \mathbb{R}^N \setminus \mathbf{0}} \frac{\boldsymbol{c}^T A \boldsymbol{c}}{\boldsymbol{c}^T \boldsymbol{c}}.$$

This leads to the following bound for the norm of the inverse of A.

Lemma 8.1 *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, be distinct points in $\mathbb{R}^s$ and let $\Phi : \mathbb{R}^s \to \mathbb{R}$ be either strictly positive definite or strictly conditionally negative definite of order one with $\Phi(\mathbf{0}) \leq 0$. Also, let A be the interpolation matrix with entries $A_{ij} = \Phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$. If the inequality*

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j A_{ij} \geq \theta \|\boldsymbol{c}\|_2^2$$

is satisfied whenever the components of $\boldsymbol{c}$ satisfy $\sum_{j=1}^{N} c_j = 0$, then

$$\|A^{-1}\|_2 \leq \theta^{-1}.$$

Note that for positive definite matrices the Rayleigh quotient implies $\theta = \lambda_{\min}$ which shows why lower bounds on the smallest eigenvalue correspond to to upper bounds on the norm of the inverse of A. In order to obtain the bound for conditionally negative matrices the Courant-Fischer Theorem needs to be employed.

Narcowich and Ward establish bounds on the norm of the inverse of A in terms of the *separation distance* of the data sites

$$q_{\mathcal{X}} = \frac{1}{2} \min_{i \neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2.$$

We can picture $q_{\mathcal{X}}$ as the radius of the largest ball that can be placed around every point in $\mathcal{X}$ such that no two balls overlap (see Figure 3).

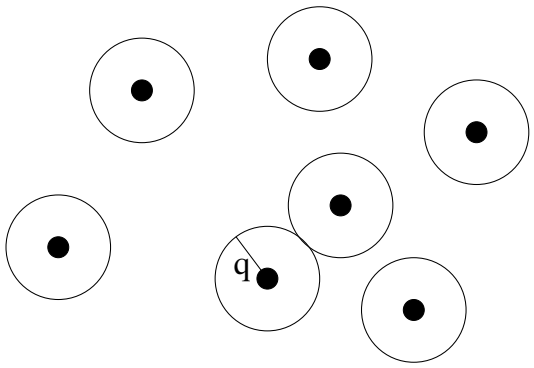

Figure 3: The separation distance $q_{\mathcal{X}}$ for a set of data sites in $\mathbb{R}^2$.

The derivation of these bounds is rather technical, and for details we refer to either the original papers by Narcowich, Ward and co-workers, the more recent paper [175]

by Schaback (who uses a slightly simpler strategy), or Wendland's manuscript [205]. We now list several bounds as derived in [205].

Examples: In the examples below the explicit constants

$$M_s = 12\left(\frac{\pi\Gamma^2(\frac{s+2}{2})}{9}\right)^{1/(s+1)} \leq 6.38s \quad \text{and} \quad C_s = \frac{1}{2\Gamma(\frac{s+2}{2})}\left(\frac{M_s}{\sqrt{8}}\right)^s$$

are used. The upper bound for M_s can be obtained using Stirling's formula (see, e.g., [205]).

1. For Gaussians $\Phi(\boldsymbol{x}) = e^{-\alpha\|\boldsymbol{x}\|^2}$ one obtains

$$\lambda_{\min} \geq C_s(2\alpha)^{-s/2}e^{-40.71s^2/(q_{\mathcal{X}}^2\alpha)}q_{\mathcal{X}}^{-s}.$$

2. For (inverse) multiquadrics $\Phi(\boldsymbol{x}) = \left(\|\boldsymbol{x}\|^2 + \alpha^2\right)^\beta$, $\beta \in \mathbb{R} \setminus \mathbb{N}_0$ one obtains

$$\lambda_{\min} \geq C(\alpha, \beta, s)q_{\mathcal{X}}^{\beta-\frac{s}{2}+\frac{1}{2}}e^{-2\alpha M_s/q_{\mathcal{X}}}$$

   with another explicitly known constant $C(\alpha, \beta, s)$.

3. For thin plate splines $\Phi(\boldsymbol{x}) = (-1)^{k+1}\|\boldsymbol{x}\|^{2k}\log\|\boldsymbol{x}\|$, $k \in \mathbb{N}$, one obtains

$$\lambda_{\min} \geq C_s c_k(2M_s)^{-s-2k}q_{\mathcal{X}}^{2k}$$

   with another explicitly known constant c_k.

4. For the powers $\Phi(\boldsymbol{x}) = (-1)^{\lceil\beta/2\rceil}\|\boldsymbol{x}\|^\beta$, $\beta > 0$, $\beta \notin 2\,\mathbb{N}$, one obtains

$$\lambda_{\min} \geq C_s c_\beta(2M_s)^{-s-\beta}q_{\mathcal{X}}^{\beta}$$

   with another explicitly known constant c_β.

5. For the compactly supported functions $\Phi_{s,k}(\boldsymbol{x}) = \varphi_{s,k}(\|\boldsymbol{x}\|)$ of Section 4 one obtains

$$\lambda_{\min} \geq C(s, k)q_{\mathcal{X}}^{2k+1}$$

   with a constant $C(s, k)$ depending on s and k.

By providing matching *lower* bounds for $\|A^{-1}\|_2$ Schaback [168] showed that the upper bounds on the norm of the inverse obtained earlier by Narcowich, Ward and others are near optimal.

For the infinitely smooth functions of Examples 1 and 2 we see that, for a fixed shape parameter α, the lower bound for $\lambda_{\min}$ goes exponentially to zero, and therefore the condition number of the interpolation matrix A grows exponentially, as the separation distance $q_{\mathcal{X}}$ decreases. This shows that, if one adds more interpolation points in order to improve the accuracy of the interpolant (within the same domain Ω), then the problem becomes increasingly ill-conditioned. Of course one would always expect this to happen, but here the ill-conditioning grows primarily due to the decrease in the

separation distance $q_{\mathcal{X}}$, and not to the increase in the number N of data points. We will come back to this observation when we discuss a possible change of basis in Section 8.4.

On the other hand, if one keeps the number of points (or at least the separation distance) fixed and instead increases (reduces) the value of α for Gaussians (multiquadrics), then the condition number of A is improved. This corresponds to the *stationary* approximation setting (which we did not discuss in detail earlier). In this case it is possible to show that the upper bound for the error estimate increases, i.e., the accuracy of the interpolant deteriorates. Conversely, one can attempt to improve the accuracy of a radial basis function interpolant by decreasing (increasing) α for Gaussians (multiquadrics). However, this is only possible at the cost of numerical instability (ill-conditioning of A). This is to be expected since for small (large) values of α the Gaussians (multiquadrics) more and more resemble a constant function, and therefore the rows (as well as columns) of the matrix A become more and more alike, so that the matrix becomes almost singular – even for well separated data sites. In the literature this phenomenon has been referred to as *trade-off* or (*uncertainty*) *principle* (see, e.g., the paper [169] by Schaback).

This trade-off has led a number of people to search for an "optimal" value of the shape parameter, i.e., a value that yields maximal accuracy, while still maintaining numerical stability. For example, in his original work on (inverse) multiquadric interpolation in $\mathbb{R}^2$ Hardy [83] suggested using $\alpha = 0.815d$, where $d = \frac{1}{N}\sum_{i=1}^{N} d_i$, and d_i is the distance from $\boldsymbol{x}_i$ to its nearest neighbor. Later Franke [74] suggested using $\alpha = 1.25\frac{D}{\sqrt{N}}$, where D is the diameter of the smallest circle containing all data points. Foley [70] based his strategy for finding a good value for α on the observation that that good value was similar for multiquadrics and inverse multiquadrics. Other studies were reported in [28] and [29]. A more recent algorithm was proposed by Rippa in [161]. He suggests a variant of cross validation known as "leave-one-out" cross validation. This method is rather popular in the statistics literature where it is also known as PRESS (Predictive REsidual Sum of Squares). In this algorithm an "optimal" value of α is selected by minimizing the least squares error for a fit based on the data for which one of the centers was "left out". A similar strategy was proposed earlier in [77] for the solution of elliptic partial differential equations via the dual reciprocity method based on multiquadric interpolation.

More recently, Fornberg and co-workers have investigated the dependence of the stability on the values of the shape parameter α in a series of papers (e.g., [40, 71, 111]). On the one hand, they suggest a way of stably computing very accurate (inverse) multiquadric and Gaussian interpolants (with extreme values of α) by using a complex Contour-Padé integration algorithm. This algorithm is rather expensive, and so far only applicable for problems involving no more than 100 centers. On the other hand, Fornberg and co-workers as well as Schaback [176] have shown that in the limiting case of the shape parameter α, i.e., with very "flat" basis functions, the infinitely smooth radial basis function interpolants approach multivariate polynomial interpolants. Therefore, Fornberg and his co-workers suggest using radial basis functions as a generalization of spectral methods (applicable also in the case of scattered data) for the numerical solution of partial differential equations. This approach was also taken recently by Sarra [166].

There is also a trade-off principle for compactly supported functions. This was

Mesh	ℓ_2-error	rate	% nonzero	time
3×3	2.367490e-01		100	0
5×5	6.572754e-02	1.849	57.8	0
9×9	1.740723e-02	1.917	23.2	0
17×17	2.362950e-03	2.881	7.47	1
33×33	2.060493e-03	0.198	2.13	1
65×65	2.012010e-03	0.034	0.06	11
129×129	2.007631e-03	0.003	0.01	158

Table 1: 2D stationary interpolation with $\varphi(r) = (1-r)_+^4(4r+1)$, 25 points in support.

explained theoretically as well as illustrated with numerical experiments by Schaback [171]. The consequences are as follows. In the case of stationary interpolation, i.e., if we scale the support size of the basis functions proportional to the fill distance $h_{\mathcal{X},\Omega}$, then the "bandwidth" of the interpolation matrix A is constant. This means we can apply numerical algorithms (e.g., conjugate gradient) that can be performed in $\mathcal{O}(N)$ computational complexity. The method is numerically stable, but there will be essentially no convergence (see Table 1). In the non-stationary case, i.e., with fixed support size, the bandwidth of A increases as $h_{\mathcal{X},\Omega}$ decreases. This results in convergence (i.e., the error decreases) as we showed in Section 5, but the interpolation matrices will become more and more dense as well as ill-conditioned. Therefore, this approach is not very efficient (see Table 2).

In Tables 1 and 2 we illustrate this behavior. We use the compactly supported function $\varphi_{3,1}(r) = (1-r)_+^4(4r+1)$ to interpolate Franke's function

$$\begin{aligned} F(x,y) &= \frac{3}{4}\left[\exp\left(-\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4}\right) + \exp\left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10}\right)\right] \\ &\quad + \frac{1}{2}\exp\left(-\frac{(9x-7)^2}{4} - (9y-3)^2\right) - \frac{1}{5}\exp\left(-(9x-4)^2 - (9y-7)^2\right) \end{aligned}$$

on a grid of equally spaced points in the unit square $[0,1]^2$. In the stationary case (Table 1) the support of the basis function is scaled to contain 25 grid points. Therefore, the "bandwidth" of the interpolation matrix A is kept constant (at 25), so that A is very sparse for finer grids. We can observe convergence for the first few iterations, but once an ℓ_2-error of approximately 2×10^{-3} is reached, there is no further improvement. This behavior is not yet fully understood. However, it is similar to what happens in the *approximate approximation* method of Maz'ya (see, e.g., [127]). The rate listed in the table is the exponent β of the observed ℓ_2-convergence rate $\mathcal{O}(h^\beta)$. The % nonzero column indicates the sparsity of the interpolation matrices, and the time is measured in seconds.

In the non-stationary case (Table 2) we used the basis function without adjusting its support size. This is the situation to which the error bounds of Section 5 apply. We have convergence – although it is not obvious what the rate might be. However, the matrices become increasingly dense. Therefore, Table 2 is missing the entry for the 129×129 case, and even though no times are provided in that table, the time for the 65×65 case is already more than 20 minutes on a standard desktop PC.

Mesh	ℓ_2-error	rate
3×3	2.407250e-01	
5×5	7.101748e-02	1.761
9×9	1.833534e-02	1.954
17×17	1.392914e-03	3.718
33×33	3.050789e-04	2.191
65×65	9.314516e-06	5.034

Table 2: 2D non-stationary interpolation with $\varphi(r) = (1-r)_+^4(4r+1)$, unit support.

8.2 Multilevel Interpolation and Approximation

In order to overcome the problems with both approaches for interpolation with compactly supported radial functions described above, Schaback suggested using a multilevel stationary scheme. This scheme was implemented first by Floater and Iske [69] and later studied by a number of other researchers (see, e.g., [32, 65, 82, 86, 96, 146, 202].

The basic idea of the multilevel interpolation algorithm is to scale the size of the support of the basis function with $h_{\mathcal{X},\Omega}$, but to interpolate to residuals on progressively refined sets of centers. This method has all of the combined benefits of the methods described earlier: it is computationally efficient (can be performed in $\mathcal{O}(N)$ operations), well-conditioned, and convergent.

An algorithm for multilevel interpolation is as follows:

Algorithm: (Multilevel interpolation)

1. Create nested point sets $\mathcal{X}_1 \subset \cdots \subset \mathcal{X}_K = \mathcal{X} \subset \mathbb{R}^s$, and initialize $\mathcal{P}f(\boldsymbol{x}) = 0$.
2. For $k = 1, 2, \ldots, K$ do
   (a) Solve $u(\boldsymbol{x}) = f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})$ on $\mathcal{X}_k$.
   (b) Update $\mathcal{P}f(\boldsymbol{x}) = \mathcal{P}f(\boldsymbol{x}) + u(\boldsymbol{x})$.

The representation of the update u at step k is of the form

$$u(\boldsymbol{x}) = \sum_{\boldsymbol{x}_j \in \mathcal{X}_k} c_j^{(k)} \varphi\left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_j\|}{\rho_k}\right)$$

with $\rho_k \simeq h_{\mathcal{X}_k,\Omega}$. This requires the solution of a linear system whose size is determined by the number of points in $\mathcal{X}_k$.

In the numerical example listed in Table 3 we again use the compactly supported function $\varphi_{3,1}(r) = (1-r)_+^4(4r+1)$ and Franke's function.

The initial scale ρ_1 was chosen so that the basis function was supported on $[-2, 2]$. Subsequent scales were successively divided by 2 – just as the fill distance of the computational grids $\mathcal{X}_k$. The rate listed in the table is the exponent β of the observed ℓ_2-convergence rate $\mathcal{O}(h^\beta)$. The % nonzero column indicates the sparsity of the interpolation matrices, and the time is measured in seconds.

So far there are only limited theoretical results concerning the convergence of this multilevel algorithm. Narcowich, Schaback and Ward [146] show that a related algorithm (in which additional boundary conditions are imposed) converges at least linearly,

Mesh	ℓ_2-error	rate	% nonzero	time
3×3	2.367490e-01		100	0
5×5	6.665899e-02	1.828	57.8	0
9×9	2.087575e-02	1.675	23.2	0
17×17	1.090837e-04	4.258	7.47	0
33×33	1.497227e-04	2.865	2.13	6
65×65	5.313053e-05	1.495	0.06	37
129×129	1.112638e-05	2.256	0.01	212

Table 3: 2D (stationary) multilevel interpolation with $\varphi(r) = (1-r)_+^4(4r+1)$.

and Hartmann analyzed the multilevel algorithm in his Ph.D. thesis [86]. He showed at least linear convergence for multilevel interpolation on a regular lattice for various radial basis functions. Similar results are obtained by Hales and Levesley [82] for polyharmonic splines, i.e., thin plate splines and powers. The main difficulty in proving the convergence of the multilevel algorithm is the fact that the approximation space changes from one level to the next. The approximation spaces are not nested (as they usually are for wavelets). This means that the native space norm changes from one level to the next. Hales and Levesley avoid this problem by scaling the (uniformly spaced) data instead of the basis functions. Then the fact that polyharmonic splines are in a certain sense harmonic (see Section 8.4) simplifies the analysis. This fact was also used by Wendland [205] to prove linear convergence for multilevel (scattered data) interpolation based on thin plate splines.

The same basic multilevel algorithm can also be used for other approximation methods. In [61] the idea was applied to moving least squares methods and approximate moving least squares methods. Tables 4 and 5 illustrate the effect of the multilevel algorithm for Shepard's method and a moving least squares approximation with linear precision, both based on the compactly supported weight function $\varphi_{3,1}(r) = (1-r)_+^4(4r+1)$. This experiment was conducted with a mollified Franke function f on the unit square $[0,1]^2$, i.e.,

$$\begin{aligned} F(x,y) &= \frac{3}{4}\left[\exp\left(-\frac{(9x-2)^2}{4} - \frac{(9y-2)^2}{4}\right) + \exp\left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10}\right)\right] \\ &\quad + \frac{1}{2}\exp\left(-\frac{(9x-7)^2}{4} - (9y-3)^2\right) - \frac{1}{5}\exp\left(-(9x-4)^2 - (9y-7)^2\right), \\ f(x,y) &= 15\exp\left(\frac{-1}{1-4(x-1/2)^2}\right)\exp\left(\frac{-1}{1-4(y-1/2)^2}\right)F(x,y) \ . \end{aligned}$$

The support scaling was as in the previous multilevel example.

One can observe that the basic Shepard's method actually performs much better than the predicted $\mathcal{O}(h)$ (see Table 4). Notice that the multilevel algorithm (illustrated in Table 5) improves the accuracy considerably at very little extra cost. It is interesting to note that this effect is much more pronounced for computations in $\mathbb{R}^2$ than in $\mathbb{R}$ (cf. [60]). The times listed in Tables 4 and 5 are due only to the evaluation on a very fine evaluation mesh since the method was coded so that no linear systems had to be solved. This means that the Lagrange multipliers for the case of linear precision were

Mesh	Shepard			linear precision		
	ℓ_2-error	rate	time	ℓ_2-error	rate	time
3×3	2.737339e-01		7	2.749670e-01		14
5×5	1.100713e-01	1.314	7	1.033060e-01	1.412	13
9×9	5.393041e-02	1.029	5	5.242492e-02	0.979	9
17×17	1.507797e-02	1.839	3	1.502361e-02	1.803	5
33×33	4.124059e-03	1.870	3	4.111092e-03	1.870	4
65×65	1.061904e-03	1.957	2	1.047348e-03	1.973	3
129×129	2.628645e-04	2.014	2	2.628645e-04	1.994	3

Table 4: 2D MLS approximation with weight $\varphi(r) = (1-r)_+^4(4r+1)$.

Mesh	Shepard			linear precision		
	ℓ_2-error	rate	time	ℓ_2-error	rate	time
3×3	2.737339e-01		7	2.749670e-01		14
5×5	1.076424e-01	1.347	7	1.013114e-01	1.440	12
9×9	3.909725e-02	1.461	5	4.308322e-02	1.234	9
17×17	7.327282e-03	2.416	3	8.549613e-03	2.333	6
33×33	9.545860e-04	2.940	2	8.937409e-04	3.258	4
65×65	1.424136e-04	2.745	2	9.896052e-05	3.175	3
129×129	3.946680e-05	1.851	2	1.361339e-05	2.872	2

Table 5: 2D multilevel MLS approximation with $\varphi(r) = (1-r)_+^4(4r+1)$.

determined explicitly by solving the 3×3 Gram system analytically (cf. (33)). The resulting generating functions 34) were directly coded into the program.

This type of multilevel (and also multiscale) moving least squares method is at the heart of the meshfree method for the nanomechanics application mentioned at the end of Section 9.

There seems to be no theoretical investigation of the convergence properties of the multilevel algorithm for moving least squares approximation.

8.3 Preconditioning

In the first part of this section we noted that the system matrices arising in scattered data interpolation with radial basis functions tend to become very ill-conditioned as the minimal separation distance $q_{\mathcal{X}}$ between the data sites $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, is reduced. Therefore it is natural to devise strategies to prevent such instabilities by either preconditioning the system, or by finding a better basis for the approximation space we are using. The former approach is standard procedure in numerical linear algebra, and in fact we can use any of the well-established methods (such as preconditioned conjugate gradient iteration) to improve the stability and convergence of the interpolation systems that arise for strictly positive definite functions. In particular, the sparse systems that arise in (multilevel) interpolation with compactly supported radial basis

functions can be efficiently solved with the preconditioned conjugate gradient method, and in fact the examples reported in the previous section were implemented using the conjugate gradient method with a diagonal (Jacobi) preconditioner.

The idea of using a more stable basis is well known from univariate polynomial and spline interpolation. The Lagrange basis functions for univariate polynomial interpolation are of course the ideal basis if we are interested in stably solving the interpolation equations since the resulting interpolation matrix is the identity matrix (which is certainly much better conditioned than, e.g., the Vandermonde matrix that we get if we use a monomial basis). Similarly, B-splines give rise to diagonally dominant, sparse system matrices which are much easier to deal with than the matrices we would get if we were to represent a spline interpolant using the alternative truncated power basis. Both of these examples are studied in great detail in standard numerical analysis texts (see, e.g., [103]) or in the literature on splines (see, e.g., [183]). We will address an analogous approach for radial basis functions in the next subsection.

8.3.1 Early Preconditioners

Ill-conditioning of the interpolation matrices was identified as a serious problem very early, and Nira Dyn along with some of her co-workers (see, e.g., [48], [49], [50], or [51]) provided some of the first preconditioning strategies tailored especially to radial basis functions.

For the following discussion we consider the general interpolation problem which includes polynomial reproduction. Therefore, we have to solve the following system of linear equations

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}, \tag{47}$$

with the individual pieces given by $A_{jk} = \varphi(\|\boldsymbol{x}_j - \boldsymbol{x}_k\|)$, $j,k = 1,\ldots,N$, $P_{j\ell} = p_\ell(\boldsymbol{x}_j)$, $j = 1,\ldots,N$, $\ell = 1,\ldots,M$, $\boldsymbol{c} = [c_1,\ldots,c_N]^T$, $\boldsymbol{d} = [d_1,\ldots,d_M]^T$, $\boldsymbol{y} = [y_1,\ldots,y_N]^T$, and $\boldsymbol{0}$ a zero vector of length M with $M = \dim\Pi^s_{m-1}$. Here, as discussed earlier, φ should be strictly conditionally positive definite of order m and radial on $\mathbb{R}^s$ and the set $\mathcal{X} = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$ should be $(m-1)$-unisolvent.

The preconditioning scheme proposed by Dyn and her co-workers is motivated by the fact that the polyharmonic splines

$$\varphi(r) = \begin{cases} r^{2k-s}\log r, & s \text{ even}, \\ r^{2k-s}, & s \text{ odd}, \end{cases}$$

$2k > s$, are fundamental solutions of the k-th iterated Laplacian in $\mathbb{R}^s$, i.e.,

$$\Delta^k\varphi(\|\boldsymbol{x}\|) = c\delta_{\boldsymbol{0}}(\boldsymbol{x}),$$

where $\delta_{\boldsymbol{0}}$ is the Dirac delta function, and c is an appropriate constant.

One now wants to discretize the Laplacian on the (irregular) mesh given by the (scattered) data sites in $\mathcal{X}$. To this end Dyn, Levin, and Rippa [51] suggest a procedure that is based on a discretization of the Laplacian on the Delaunay triangulation of the scattered centers for the interpolation problem in $\mathbb{R}^2$. However, in order to also take into account the special role of the boundary points Dyn, Levin and Rippa instead use a discretization of an iterated Green's formula which has the space Π^2_{m-1} as its null

space. The necessary partial derivatives are then approximated on the triangulation using certain sets of vertices of the triangulation (3 points for first order partials, 6 for second order).

The discretization described above yields a matrix $B = (b_{ji})_{j,i=1}^{N}$ as the preconditioning matrix. We now obtain

$$(BA)_{jk} = \sum_{i=1}^{N} b_{ji}\varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_k\|) \approx \Delta^m \varphi(\|\cdot - \boldsymbol{x}_k\|)(\boldsymbol{x}_j), \quad j,k = 1,\ldots,N, \tag{48}$$

which has the property that the entries close to the diagonal are large compared to those away from the diagonal, which decay to zero as the distance between the two points involved goes to infinity. Since the part $BP = 0$ by construction, one must now solve the system

$$\begin{aligned} BA\boldsymbol{c} &= B\boldsymbol{y} \\ P^T\boldsymbol{c} &= 0. \end{aligned}$$

Actually, the system $BA\boldsymbol{c} = B\boldsymbol{y}$ is singular, but it is shown in the paper [51] that the additional constraints $P^T\boldsymbol{c} = 0$ guarantee existence of a unique solution. Furthermore, the coefficients $\boldsymbol{d}$ in the original expansion of the interpolant s can be obtained by solving

$$P\boldsymbol{d} = \boldsymbol{y} - A\boldsymbol{c},$$

i.e., by fitting the polynomial part of the expansion to the residual $\boldsymbol{y} - A\boldsymbol{c}$.

The approach just described leads to localized basis functions ψ which are linear combinations of the original basis functions φ. More precisely,

$$\psi_j(\boldsymbol{x}) = \sum_{i=1}^{N} b_{ji}\varphi(\|\boldsymbol{x} - \boldsymbol{x}_i\|) \approx \Delta^m \varphi(\|\cdot - \boldsymbol{x}_j\|)(\boldsymbol{x}), \tag{49}$$

where the coefficients b_{ji} are those determined by the discretization above.

The localized basis functions ψ_j, $j = 1,\ldots,N$, (see (49)) can be viewed as an alternative (better conditioned) basis for the approximation space spanned by the functions $\varphi_j = \varphi(\|\cdot - \boldsymbol{x}_j\|)$.

Baxter [9, 10] discusses the use of a preconditioned conjugate gradient method for solving the interpolation problem in the case when Gaussians or multiquadrics are used on a regular grid. The resulting matrices are Toeplitz matrices, and a large body of literature exists for dealing with this special case (see, e.g., [30]).

8.3.2 Preconditioned GMRES via Approximate Cardinal Functions

More recently, Beatson, Cherrie and Mouat [11] have proposed a preconditioner for the iterative solution of radial basis function interpolation systems using the GMRES method of Saad and Schultz [165]. The GMRES method is a general purpose iterative solver that can be applied to nonsymmetric (nondefinite) systems. For fast convergence the matrix should be preconditioned such that its eigenvalues are clustered around 1 and away from the origin. Obviously, if the basis functions for the radial basis function

space were cardinal functions, then the matrix would be the identity matrix with all its eigenvalues equal to 1. Therefore, the GMRES method would converge in a single iteration. Consequently, the preconditioning strategy for the GMRES method is to obtain a preconditioning matrix B that is close to the inverse of A.

Since it is too expensive to find the true cardinal basis (this would involve at least as much work as solving the interpolation problem), the idea pursued in [11] (and suggested earlier in [12, 14]) is to find *approximate* cardinal functions similar to the functions ψ_j in the previous subsection. Now, however, there is also an emphasis on efficiency, i.e., we are interested in *local* approximate cardinal functions, if possible. Several different strategies were suggested in [11]. We will now explain the basic idea.

Given the centers $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, the j-th approximate cardinal function is given as a linear combination of the basis functions $\varphi_i = \varphi(\| \cdot - \boldsymbol{x}_i\|)$, where i runs over (some subset of) $\{1, \ldots, N\}$, i.e.,

$$\psi_j = \sum_{i=1}^{N} b_{ji}\varphi(\| \cdot - \boldsymbol{x}_i\|) + p_j, \tag{50}$$

where (for the conditionally positive definite case) p_j is a polynomial in Π^s_{m-1} and the coefficients b_{ji} satisfy the usual conditions

$$\sum_{i=1}^{N} b_{ji}p_j(\boldsymbol{x}_i) = 0 \qquad \text{for all } p_j \in \Pi^s_{m-1}. \tag{51}$$

The key feature in designing the approximate cardinal functions is to have only a few $n \ll N$ coefficients in (50) to be nonzero. In that case the functions ψ_j are found by solving small $n \times n$ linear systems, which is much more efficient than dealing with the original $N \times N$ system. For example, in [11] the authors use $n \approx 50$ for problems involving up to 10,000 centers. The resulting preconditioned system is of the same form as the earlier preconditioner (48), i.e., we now have to solve the preconditioned problem

$$(BA)\boldsymbol{c} = B\boldsymbol{y},$$

where the entries of the matrix BA are just $\psi_j(\boldsymbol{x}_k)$, $j, k = 1, \ldots, N$.

The simplest strategy for determining the coefficients b_{ji} is to select the n nearest neighbors of $\boldsymbol{x}_j$, and to find b_{ji} by solving the (local) cardinal interpolation problem

$$\psi_j(\boldsymbol{x}_i) = \delta_{ij}, \qquad i = 1, \ldots, n,$$

subject to the moment constraint (51) listed above. Here δ_{ij} is the Kronecker-delta, and the points $\boldsymbol{x}_i$ are the nearest neighbors selected above.

This basic strategy is improved by adding so-called *special points* that are distributed (very sparsely) throughout the domain.

A few numerical results for thin plate spline and multiquadric interpolation in $\mathbb{R}^2$ from [11] are listed in Table 6. The condition numbers are ℓ_2-condition numbers, and the points were randomly distributed in the unit square. The "local precond." column uses the $n = 50$ nearest neighbors to determine the approximate cardinal functions, whereas the right-most column uses the 41 nearest neighbors plus 9 special points placed uniformly in the unit square. The effect of the preconditiong on the performance of

φ	N	unprecond.	local precond.	local precond. w/special
TPS	289	4.005e06	1.464e03	5.721e00
	1089	2.753e08	6.359e05	1.818e02
	4225	2.605e09	2.381e06	1.040e06
MQ	289	1.506e08	3.185e03	2.639e02
	1089	2.154e09	8.125e05	5.234e04
	4225	3.734e10	1.390e07	4.071e04

Table 6: Condition numbers without and with preconditioning.

the GMRES algorithm was, e.g., a reduction from 103 to 8 iterations for the 289 point data set for thin plate splines, or from 145 to 11 for multiquadrics.

Remark: An extension of the ideas of Beatson, Cherrie and Mouat [11] to linear systems arising in the collocation solution of partial differential equations (see Section 9) was explored in Mouat's Ph.D. thesis [143] and also in the recent paper by Ling and Kansa [119].

8.4 Change of Basis

Another idea that can be used to obtain a "better" basis for conditionally positive definite radial basis functions is closely connected to finding the reproducing kernel of the associated native space. Since we did not elaborate on the construction of the native spaces for conditionally positive definite functions earlier, we will now present the relevant formulas (without going into the details). In particular, for polyharmonic splines we will be able to find a basis that is in a certain sense *homogeneous*, and therefore the condition number of the related interpolation matrix will depend only on the number N of data points, but *not* on their separation distance.

This approach was suggested by Beatson, Light and Billings [13], and has its roots in work by Sibson and Stone [186].

Let Φ be a strictly conditionally positive definite kernel of order m, and $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega \subset \mathbb{R}^s$ be an $(m-1)$-unisolvent set of centers. Then the reproducing kernel for the native space $\mathcal{N}_\Phi(\Omega)$ is given by

$$\begin{aligned} K(\boldsymbol{x}, \boldsymbol{y}) &= \Phi(\boldsymbol{x}, \boldsymbol{y}) - \sum_{k=1}^{M} p_k(\boldsymbol{x}) \Phi(\boldsymbol{x}_k, \boldsymbol{y}) - \sum_{\ell=1}^{M} p_\ell(\boldsymbol{y}) \Phi(\boldsymbol{x}, \boldsymbol{x}_\ell) \\ &\quad + \sum_{k=1}^{M} \sum_{\ell=1}^{M} p_k(\boldsymbol{x}) p_\ell(\boldsymbol{y}) \Phi(\boldsymbol{x}_k, \boldsymbol{x}_\ell) + \sum_{\ell=1}^{M} p_\ell(\boldsymbol{x}) p_\ell(\boldsymbol{y}), \end{aligned}$$

where the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ are an $(m-1)$-unisolvent subset of $\mathcal{X}$ and the polynomials p_k, $k = 1, \ldots, M$, form a *cardinal* basis for Π^s_{m-1} whose dimension is $M = \binom{s+m-1}{m-1}$, i.e.,

$$p_\ell(\boldsymbol{x}_k) = \delta_{k,\ell}, \qquad k, \ell = 1, \ldots, M.$$

An immediate consequence is that we can express the radial basis function interpolant to values of some function f given on $\mathcal{X}$ in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j K(\boldsymbol{x}, \boldsymbol{x}_j), \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

The coefficients c_j are determined by satisfying the interpolation conditions

$$\mathcal{P}f(\boldsymbol{x}_i) = f(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

We will see below (in Tables 7 and 8) that this basis already performs "better" than the standard basis $\{\Phi(\cdot, \boldsymbol{x}_1), \ldots, \Phi(\cdot, \boldsymbol{x}_N)\}$ if we keep the number of centers fixed, and vary only their separation distance.

To obtain the homogeneous basis referred to above we modify K by subtracting the tensor product polynomial, i.e.,

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{x}, \boldsymbol{y}) - \sum_{\ell=1}^{M} p_\ell(\boldsymbol{x}) p_\ell(\boldsymbol{y}).$$

Now, if $\boldsymbol{y}$ is one of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ in the $(m-1)$-unisolvent subset of $\mathcal{X}$ mentioned above, then

$$\begin{aligned}
\kappa(\cdot, \boldsymbol{y}) &= \Phi(\cdot, \boldsymbol{y}) - \sum_{k=1}^{M} p_k(\cdot)\Phi(\boldsymbol{x}_k, \boldsymbol{y}) - \sum_{\ell=1}^{M} p_\ell(\boldsymbol{y})\Phi(\cdot, \boldsymbol{x}_\ell) + \sum_{k=1}^{M}\sum_{\ell=1}^{M} p_k(\cdot) p_\ell(\boldsymbol{y})\Phi(\boldsymbol{x}_k, \boldsymbol{x}_\ell) \\
&= \Phi(\cdot, \boldsymbol{y}) - \sum_{k=1}^{M} p_k(\cdot)\Phi(\boldsymbol{x}_k, \boldsymbol{y}) - \Phi(\cdot, \boldsymbol{y}) + \sum_{k=1}^{M} p_k(\cdot)\Phi(\boldsymbol{x}_k, \boldsymbol{y}) = 0.
\end{aligned}$$

This means that the functions $\kappa(\cdot, \boldsymbol{x}_j)$, $j = 1, \ldots, N$, cannot be used as a basis of our approximation space. However, it turns out that the matrix C with entries $C_{i,j} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $i, j = M+1, \ldots, N$, is positive definite, and therefore we obtain the following basis

$$\{p_1, \ldots, p_M\} \cup \{\kappa(\cdot, \boldsymbol{x}_{M+1}), \ldots, \kappa(\cdot, \boldsymbol{x}_N)\},$$

and the interpolant can be represented in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} d_j p_j(\boldsymbol{x}) + \sum_{k=M+1}^{N} c_k \kappa(\boldsymbol{x}, \boldsymbol{x}_k), \qquad \boldsymbol{x} \in \mathbb{R}^s\,.$$

Since the polynomials p_k are cardinal on $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ the coefficients are determined by solving the linear system

$$\begin{bmatrix} I & 0 \\ P^T & C \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{c} \end{bmatrix} = \boldsymbol{y}, \tag{52}$$

with I an $M \times M$ identity matrix, C as above, $P_{ij} = p_j(\boldsymbol{x}_i)$, $j = 1, \ldots, M$, $i = M+1, \ldots, N$, $\boldsymbol{c} = [c_{M+1}, \ldots, c_N]^T$, $\boldsymbol{d} = [d_1, \ldots, d_M]^T$, and the right-hand side $\boldsymbol{y} =$

Spacing h	Standard matrix	Reproducing kernel	Homogeneous matrix
1/8	3.5158e03	1.8930e04	7.5838e03
1/16	3.8938e04	2.6514e05	1.1086e05
1/32	5.1363e05	4.0007e06	1.6864e06
1/64	7.6183e06	6.2029e07	2.6264e07

Table 7: Condition numbers for different thin plate spline bases on $[0,1]^2$ with increasing number of points and varying separation distance.

$[f(\boldsymbol{x}_1),\ldots,f(\boldsymbol{x}_M),f(\boldsymbol{x}_{M+1}),\ldots,f(\boldsymbol{x}_N)]^T$. The identity block (cardinality of the polynomial basis functions) implies that the coefficient vector $\boldsymbol{d}$ is given by

$$d_j = f(\boldsymbol{x}_j), \qquad j = 1,\ldots,M,$$

and therefore the system (52) can be solved as

$$C\boldsymbol{c} = \tilde{\boldsymbol{y}} - P^T\boldsymbol{d},$$

where $\tilde{\boldsymbol{y}} = [f(\boldsymbol{x}_{M+1}),\ldots,f(\boldsymbol{x}_N)]^T$ and the matrix C is symmetric and positive definite. Finally, for polyharmonic splines, the ℓ_2-condition number of the matrix C is invariant under a uniform scaling of the centers, i.e., if $C^h = (\kappa(h\boldsymbol{x}_i, h\boldsymbol{x}_j))$, then $\mathrm{cond}(C^h) = \mathrm{cond}(C)$. This is proved to varying degrees in the paper [13] by Beatson, Light and Billings, the thesis [205] by Wendland, and the paper [97] by Iske.

We close with some numerical experiments from [13]. They use thin plate splines in $\mathbb{R}^2$. In the first experiment (illustrated in Table 7) the problem is formulated on the unit square $[0,1]^2$. Here both the number of points and the separation distance vary from one row in the table to the next. The three different columns list the ℓ_2-condition numbers of the interpolation matrix for the three different approaches mentioned above, i.e., using the standard basis consisting of functions $\Phi(\cdot,\boldsymbol{x}_j)$ and monomials, using the reproducing kernels $K(\cdot,\boldsymbol{x}_j)$, and using the matrix C. The three polynomial cardinal functions are based on the three corners $(0,0)$, $(0,1)$, and $(1,0)$. With this setup all three methods perform comparably.

In the second experiment (shown in Table 8) the number of points is kept fixed at 5×5 equally spaced points. However, the domain is scaled to the square $[0,a]^2$ with scale parameter a, so that only the separation distance $q_{\mathcal{X}}$ changes from one row to the next. Now, clearly the two new methods show less dependence on the separation distance, with the homogeneous matrix C being completely insensitive as claimed earlier.

Remark: Iske takes advantage of the scale invariance of polyharmonic splines (and thin plate splines in particular) in the construction of a numerical multiscale solver for transport problems (see, e.g., [15]).

8.5 Special Numerical Algorithms

Since the use of radial basis functions for interpolation of scattered data leads to (large) linear systems that are frequently ill-conditioned it is important to devise algorithms that can

Scale parameter	Standard matrix	Reproducing kernel	Homogeneous matrix
0.001	2.4349e08	8.4635e08	5.4938e02
0.01	2.4364e06	8.4640e06	5.4938e02
0.1	2.5179e04	8.5134e04	5.4938e02
1.0	3.6458e02	1.3660e03	5.4938e02
10	1.8742e06	1.2609e03	5.4938e02
100	1.1520e11	1.1396e05	5.4938e02
1000	3.4590e15	1.1386e07	5.4938e02

Table 8: Condition numbers for different thin plate spline bases on $[0, a]^2$ with fixed number of points and varying separation distance.

1. efficiently solve the interpolation system (preferably in $\mathcal{O}(N)$ operations), and
2. efficiently evaluate a radial basis function expansion once its coefficients have been determined (preferably in a constant number of operations – independent of N).

The second goal is also important for approximation via the moving least squares method or by quasi-interpolation.

All of the work described below is very recent, and it is quite likely that much more insight can be gained, and many improvements are still possible.

Various iterative algorithms for the (approximate) evaluation of the scattered data interpolant have recently been suggested by Schaback and Wendland (see, e.g., [178, 179, 205]) as well as by Faul and Powell (see, e.g., [66, 67]). Both groups of authors base their algorithms on an iteration on residuals. We will not discuss the details of these algorithms here.

For most algorithms one needs to make sure that the residuals are evaluated efficiently. Common approaches are fast multipole expansions, fast tree codes, fast Fourier transforms, or compactly supported functions.

Since fast multipole expansions and tree codes are a standard tool in molecular dynamics simulations (and most likely discussed at other places in this handbook) we concentrate on some recent work by Kunis, Nieslony, Potts and Steidl on the fast Fourier transform at *nonuniformly spaced points*. In the recent papers [105, 152, 154] the authors suggest the use of the fast Fourier transform at nonuniformly spaced points as an efficient way to solve and evaluate radial basis function problems. The software package NFFT by the authors is available for free download [104]. A discussion of the actual NFFT software would go beyond the scope of this survey. Instead, we briefly describe how to use NFFTs and FFTs to evaluate expansions of the form

$$\mathcal{P}f(\boldsymbol{y}_j) = \sum_{k=1}^{N} f(\boldsymbol{x}_k)\Phi(\boldsymbol{y}_j - \boldsymbol{x}_k) \tag{53}$$

simultaneously at many evaluation points $\boldsymbol{y}_j$, $j = 1, \ldots, M$. Direct summation requires $\mathcal{O}(MN)$ operations, while it can be shown that use of the NFFT reduces the cost to $\mathcal{O}(M + N)$ operations. Therefore, as is always the case with fast Fourier transforms, use of the algorithm will pay off for sufficiently many evaluations.

In their papers Nieslony, Potts and Steidl distinguish between kernels Φ that are singular and those that are non-singular. Singular kernels are C^∞ everywhere except at the origin and include examples such as

$$\frac{1}{r},\ \frac{1}{r^2},\ \log r,\ r^2 \log r,$$

where $r = \|\cdot\|$. Non-singular kernels are smooth everywhere such as Gaussians and (inverse) multiquadrics. We will restrict our discussion to this latter class.

The basic idea for the following algorithm is remarkably simple. It relies on the fact that the exponential $e^{-\alpha(\boldsymbol{y}_j - \boldsymbol{x}_k)}$ can be written as $e^{-\alpha \boldsymbol{y}_j} e^{\alpha \boldsymbol{x}_k}$. Moreover, the method applies to arbitrary kernels (which is in strong contrast to the fast multipole type methods for which one requires different expansions for each different kernel). One starts out by approximating the (arbitrary, but smooth) kernel using standard Fourier series, i.e.,

$$\Phi(\boldsymbol{x}) \approx \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell} \boldsymbol{x}}$$

with index set $I_n = \left[-\frac{n}{2}, \frac{n}{2}\right)^s$. The coefficients $b_{\boldsymbol{\ell}}$ are found by the discrete inverse Fourier transform

$$b_{\boldsymbol{\ell}} = \frac{1}{n^s} \sum_{\boldsymbol{k} \in I_n} \Phi\left(\frac{\boldsymbol{k}}{n}\right) e^{-2\pi i \boldsymbol{k} \boldsymbol{\ell}/n}.$$

Numerically, this task is accomplished with software for the standard (inverse) FFT (e.g., [75]).

Remark: Note that this definition of the Fourier transform (as well as the one below) is different from the one used in Section 2. However, in order to stay closer to the software packages, we adopt the notation used there.

Therefore,

$$\begin{aligned} \mathcal{P}f(\boldsymbol{y}_j) &\approx \sum_{k=1}^{N} f(\boldsymbol{x}_k) \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell}(\boldsymbol{y}_j - \boldsymbol{x}_k)} \\ &= \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{2\pi i \boldsymbol{\ell}(\boldsymbol{y}_j - \boldsymbol{x}_k)} \end{aligned}$$

Now, the exponential is split using the above mentioned property, i.e.,

$$\mathcal{P}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\ell} \in I_n} b_{\boldsymbol{\ell}} \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-2\pi i \boldsymbol{\ell} \boldsymbol{x}_k} e^{2\pi i \boldsymbol{\ell} \boldsymbol{y}_j}.$$

This, however, can be viewed as a fast Fourier transform at non-uniformly spaced points, i.e.,

$$\mathcal{P}f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\ell} \in I_n} c_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell} \boldsymbol{y}_j}.$$

where the coefficients $c_{\boldsymbol{\ell}} = b_{\boldsymbol{\ell}} a_{\boldsymbol{\ell}}$ with

$$a_{\boldsymbol{\ell}} = \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-2\pi i \boldsymbol{\ell} \boldsymbol{x}_k}$$

which is nothing but an inverse discrete Fourier transform at non-uniformly spaced points. These latter two transforms are dealt with numerically using the NFFT software.

Together, for the case of non-singular kernels Φ we have the following algorithm.

Algorithm (Fast Fourier transform evaluation)

For $\boldsymbol{\ell} \in I_n$

Compute the coefficients

$$b_{\boldsymbol{\ell}} = \frac{1}{n^s} \sum_{\boldsymbol{k} \in I_n} \Phi\left(\frac{\boldsymbol{k}}{n}\right) e^{-2\pi i \boldsymbol{k}\boldsymbol{\ell}/n}$$

by inverse FFT.

Compute the coefficients

$$a_{\boldsymbol{\ell}} = \sum_{k=1}^{N} f(\boldsymbol{x}_k) e^{-2\pi i \boldsymbol{\ell}\boldsymbol{x}_k}$$

by inverse NFFT.

Compute the coefficients $c_{\boldsymbol{\ell}} = a_{\boldsymbol{\ell}} b_{\boldsymbol{\ell}}$.

end

For $1 \leq j \leq M$

Compute the values

$$\mathcal{P} f(\boldsymbol{y}_j) \approx \sum_{\boldsymbol{\ell} \in I_n} d_{\boldsymbol{\ell}} e^{2\pi i \boldsymbol{\ell}\boldsymbol{y}_j}$$

by NFFT.

end

Remarks:

1. In the papers [105, 152, 154] the authors also suggest a special boundary regularization in case the kernel does not decay fast enough, i.e., the kernel is large near the boundary of the domain.

2. Of course, this method will only provide an approximation of the expansion (53) and error estimates are provided in the literature (see, e.g., [152]).

3. While we only illustrated the use of (N)FFTs for the evaluation of radial sums it should be clear that this method can also be coupled with the algorithms mentioned earlier (such as preconditioned GMRES, the "greedy" algorithm by Schaback and Wendland, or the Faul-Powell algorithm) to efficiently solve radial basis function interpolation systems.

A few examples of the use of fast Fourier transforms for the evaluation of approximate moving least squares approximations (quasi-interpolants) are given in Figures 4–6. The graphs on the left indicate ℓ_∞ approximation errors for a Franke-type function. The graphs on the right show the execution times in seconds for direct summation (solid lines) and FFT summations (dashed lines). The colors correspond to the three different types of kernels listed in Table 9 below. The red curves correspond to the Gaussians (listed in the $\mathcal{O}(h^2)$ column), green curves to the function in the $\mathcal{O}(h^4)$ column (Gaussian multiplied by a linear Laguerre polynomial), and blue curves to those in the $\mathcal{O}(h^6)$ column (Gaussian multiplied by a quadratic Laguerre polynomial).

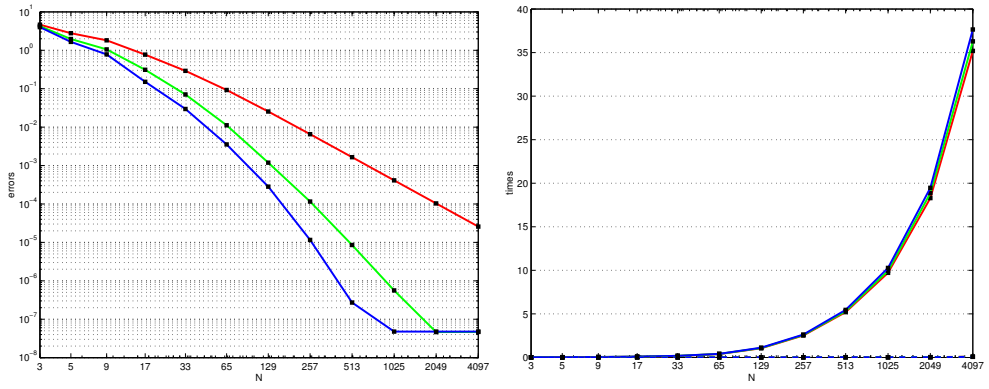

Figure 4: Convergence and execution times for 1D example.

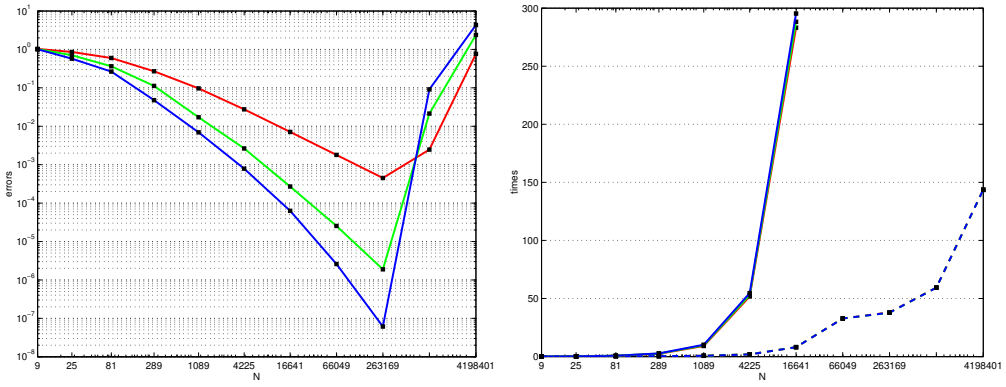

Figure 5: Convergence and execution times for 2D example.

The evaluation of the results listed in Figures 4–6 occurs at 10,001, 16,641, and 2,146,689 randomly distributed points in the unit square, respectively. The 3D experiments show that there is a cross-over value of about 1,000 evaluations at which the FFT approach becomes faster than the direct approach. For the one and two-dimensional experiments this cross-over point occurs much earlier and is not detectable in the figures.

The polynomial terms in Table 9 are given by generalized Laguerre polynomials with radial arguments. In general one can show (see, e.g., [127]) that if $L_d^{s/2}$ is used to denote the generalized Laguerre polynomial of degree d, then the smooth function

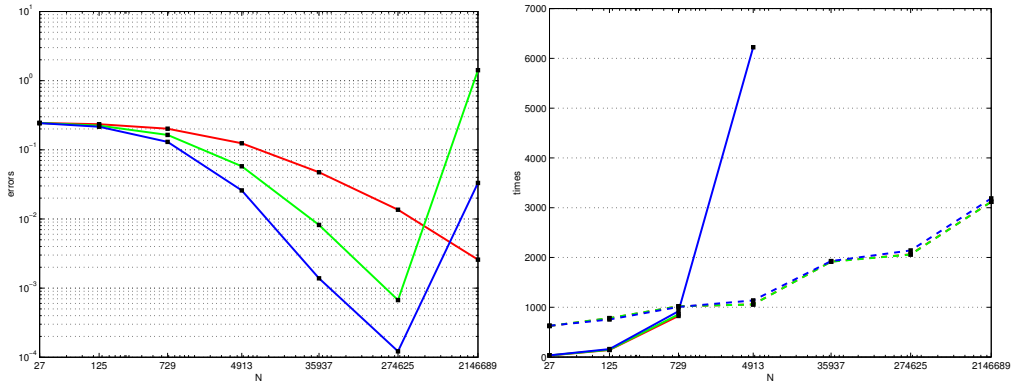

Figure 6: Convergence and execution times for 3D example.

f in $\mathbb{R}^s$ can be approximated with approximate approximation order $\mathcal{O}(h^{2d+2})$ by an expansion of the form

$$\mathcal{P}f(\boldsymbol{x}) = \frac{1}{(\pi\mathcal{D})^{s/2}} \sum_{k=1}^{N} f(\boldsymbol{x}_k) L_d^{s/2} \left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_k\|^2}{\mathcal{D}h^2} \right) \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_k\|^2}{\mathcal{D}h^2} \right).$$

Here $\mathcal{D}$ is a parameter that controls a so-called *saturation error*, i.e., the predicted approximation order is achieved only up to some user-controllable threshold (and therefore referred to as *approximate approximation*). This threshold is clearly visible in the convergence graphs.

s	$\mathcal{O}(h^2)$	$\mathcal{O}(h^4)$	$\mathcal{O}(h^6)$
1	$e^{-\lvert x\rvert^2}$	$\left(\frac{3}{2} - \lvert x\rvert^2\right) e^{-\lvert x\rvert^2}$	$\left(\frac{15}{8} - \frac{5}{2}\lvert x\rvert^2 + \frac{1}{2}\lvert x\rvert^4\right) e^{-\lvert x\rvert^2}$
2	$e^{-\|\boldsymbol{x}\|^2}$	$\left(2 - \|\boldsymbol{x}\|^2\right) e^{-\|\boldsymbol{x}\|^2}$	$\left(3 - 3\|\boldsymbol{x}\|^2 + \frac{1}{2}\|\boldsymbol{x}\|^4\right) e^{-\|\boldsymbol{x}\|^2}$
3	$e^{-\|\boldsymbol{x}\|^2}$	$\left(\frac{5}{2} - \|\boldsymbol{x}\|^2\right) e^{-\|\boldsymbol{x}\|^2}$	$\left(\frac{35}{8} - \frac{7}{2}\|\boldsymbol{x}\|^2 + \frac{1}{2}\|\boldsymbol{x}\|^4\right) e^{-\|\boldsymbol{x}\|^2}$

Table 9: Generating functions for approximate MLS approximation in $\mathbb{R}^s$.

8.6 Domain Decomposition

Finally, another method commonly used to deal with large computational problems is the *domain decomposition* method. The domain decomposition method is frequently implemented on parallel computers in order to speed up the computation even more. A standard reference (based mostly on finite difference and finite element methods) is the book by Smith, Bjørstad and Gropp [188]. For radial basis functions there is a recent paper by Beatson, Light and Billings [13].

The main aim of the paper [13] is to solve the radial basis function interpolation problem discussed multiple times in previous sections. In particular, a so-called *multiplicative Schwarz* algorithm (which is analogous to Gauss-Seidel iteration) is presented,

and linear convergence of the algorithm is proved. A section with numerical experiments reports results for an *additive Schwarz* method (which is analogous to Jacobi iteration).

In particular, the authors implemented polyharmonic radial basis functions, and used the scale invariant basis discussed in Section 8.4.

The classical additive Schwarz algorithm is usually discussed in the context of partial differential equations, and it is known that one should add a coarse level correction in order to ensure convergence and to filter out some of the low-frequency oscillations (see, e.g., [188]).

In [13] a two-level additive algorithm for interpolation problems was presented. One begins by subdividing the set on interpolation point $\mathcal{X}$ into M smaller sets $\mathcal{X}_i$, $i = 1, \ldots, M$, whose pairwise intersection is non-empty. The points that belong to one set $\mathcal{X}_i$ only are called *inner points* of $\mathcal{X}_i$. Those points in the intersection of more than one set need to be assigned in some way as inner points to only one of the subsets $\mathcal{X}_i$ so that the collection of all inner points yields the entire set $\mathcal{X}$. This corresponds to the concept of *overlapping domains*. One also needs to choose a coarse grid $\mathcal{Y}$ that contains points from all of the inner point sets.

In the setup phase of the algorithm the radial basis function interpolation matrices for the smaller problems on each of the subsets $\mathcal{X}_i$, $i = 1, \ldots, M$, are computed and factored. At this point one can use the homogeneous basis of Section 8.4 to ensure numerical stability. Now the algorithm proceeds as follows:

Algorithm:

Input: Data f, point sets $\mathcal{X}_i$ and factored interpolation matrices A_i, $i = 1, \ldots, M$, tolerance ϵ

Initialize $r = f$, $s = 0$

While $\|r\| > \epsilon$ do

For $i = 1$ to m (i.e., for each subset $\mathcal{X}_i$) do

Determine the coefficients $\boldsymbol{c}_i$ of the interpolant to the residual $r|_{\mathcal{X}_i}$ on $\mathcal{X}_i$.

end

Make $\boldsymbol{c}$ orthogonal to Π^s_{m-1}.

Assemble an intermediate approximation $s_1 = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j)$.

Compute the residual on the coarse grid, i.e.,

$$r_1 = r - s_1|_{\mathcal{Y}}.$$

Interpolate to r_1 on the coarse grid $\mathcal{Y}$ using a radial basis function expansion s_2.

Update $s = s + s_1 + s_2$.

Reevaluate the global residual $r = f - s$ on the whole set $\mathcal{X}$

end

Remarks:

1. In [13] it is proved that a multiplicative version of this algorithm converges at least linearly. However, the additive version can be more easily implemented on a parallel computer.

2. If strictly positive definite kernels such as Gaussians are used, then it is not necessary to make the coefficients $\boldsymbol{c}$ orthogonal to polynomials.

3. As in many algorithms before, the evaluation of the residuals needs to be made "fast" using either a fast multipole method or a version of the fast Fourier transform.

4. In the case of very large data sets it may be necessary to use more than two levels so that one ends up with a *multigrid* algorithm.

5. The authors of [13] report having solved interpolation problems with several millions of points using the domain decomposition algorithm above.

6. A number of other papers discussing domain decomposition methods for radial basis functions have recently appeared in the literature (see, e.g., [43, 90, 94, 114, 120, 209]). However, most of these papers contain little theory, focussing mostly on numerical experiments.

9 Applications

In the first two subsections we will focus on the solution of partial differential equations using radial basis functions in either a strong form (collocation) approach or a weak form (Galerkin) approach. A paper that surveys some of these methods and also discusses the concept of multilevel algorithms combined with smoothing techniques which improve the convergence of the multilevel algorithms is [58]. In the recent paper [177] Schaback presents a unified framework for the radial basis function solution of problems both in the strong and weak form. At the end of the section we will briefly mention some of the other meshfree methods available for the solution of differential equations, and then conclude by describing work on meshfree methods by two groups of researchers that is most relevant to the area of computational nanotechnology.

9.1 Solving Partial Differential Equations via Collocation

In the first part of this section we discuss the numerical solution of elliptic partial differential equations using a collocation approach based on radial basis functions. To make the discussion transparent we will focus on the case of a time independent linear elliptic partial differential equation in $\mathbb{R}^2$.

9.1.1 Kansa's Approach

In [100] Kansa suggested a now very popular non-symmetric method for the solution of elliptic PDEs with radial basis functions. In order to be able to clearly point out the differences between Kansa's method and a symmetric approach proposed in [56] we recall some of the basics of scattered data interpolation with radial basis functions in $\mathbb{R}^s$ discussed in the first few sections of this review.

In this context we are given data $\{\boldsymbol{x}_i, f_i\}$, $i = 1, \ldots, N$, $\boldsymbol{x}_i \in \mathbb{R}^s$, where we can think of the values f_i being sampled from a function $f : \mathbb{R}^s \to \mathbb{R}$. The goal is to find an interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{x}_j\|), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{54}$$

such that

$$\mathcal{P}f(\boldsymbol{x}_i) = f_i, \qquad i = 1, \ldots, N.$$

The solution of this problem leads to a linear system $A\boldsymbol{c} = \boldsymbol{f}$ with the entries of A given by

$$A_{ij} = \varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i, j = 1, \ldots, N. \tag{55}$$

As discussed earlier, the matrix A is non-singular for a large class of radial functions including (inverse) multiquadrics, Gaussians, and the strictly positive definite compactly supported functions of Wendland, Wu, or Buhmann. In the case of strictly conditionally positive definite functions such as thin plate splines the problem needs to be augmented by polynomials.

We now switch to the collocation solution of partial differential equations. Assume we are given a domain $\Omega \subset \mathbb{R}^s$, and a linear elliptic partial differential equation of the form

$$L[u](\boldsymbol{x}) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ in } \Omega, \tag{56}$$

with (for simplicity of description) Dirichlet boundary conditions

$$u(\boldsymbol{x}) = g(\boldsymbol{x}), \qquad \boldsymbol{x} \text{ on } \partial\Omega. \tag{57}$$

For Kansa's collocation method we then choose to represent u by a radial basis function expansion analogous to that used for scattered data interpolation, i.e.,

$$u(\boldsymbol{x}) = \sum_{j=1}^{N} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|), \tag{58}$$

where we now introduce the points $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$ as centers for the radial basis functions. They will usually be selected to coincide with the collocation points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$. However, the discussion below is clearer if we formally distinguish between centers $\boldsymbol{\xi}_j$ and collocation points $\boldsymbol{x}_i$. We assume the simplest possible setting here, i.e., no polynomial terms are added to the expansion (58). The collocation matrix which arises when matching the differential equation (56) and the boundary conditions (57) at the collocation points $\mathcal{X}$ will be of the form

$$A = \begin{bmatrix} \Phi \\ L[\Phi] \end{bmatrix}, \tag{59}$$

where the two blocks are generated as follows:

$$\begin{aligned} \Phi_{ij} &= \varphi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i \in B,\ \boldsymbol{\xi}_j \in \mathcal{X}, \\ L[\Phi]_{ij} &= L[\varphi](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i \in I,\ \boldsymbol{\xi}_j \in \mathcal{X}. \end{aligned}$$

Here we have identified (as we will do throughout this section) the set of centers with the set of collocation points. The set $\mathcal{X}$ is split into a set I of interior points, and B of boundary points. The problem is well-posed if the linear system $A\boldsymbol{c} = \boldsymbol{y}$, with $\boldsymbol{y}$ a vector consisting of entries $g(\boldsymbol{x}_i)$, $\boldsymbol{x}_i \in B$, followed by $f(\boldsymbol{x}_i)$, $\boldsymbol{x}_i \in I$, has a unique solution.

We note that a change in the boundary conditions (57) is as simple as changing a few rows in the matrix A in (59) as well as on the right-hand side $\boldsymbol{y}$. We also point out that Kansa only proposed to use multiquadrics in (58), and for that method suggested the use of varying parameters α_j, $j = 1, \ldots, N$, which improves the accuracy of the method when compared to using only one constant value of α (see [100]).

A problem with Kansa's method is that – for a constant multiquadric shape parameter α – the matrix A may for certain configurations of the centers $\boldsymbol{\xi}_j$ be singular. Originally, Kansa assumed that the non-singularity results for interpolation matrices would carry over to the PDE case. However, as the numerical experiments of Hon and Schaback [88] show, this is not so. This is to be expected since the matrix for the collocation problem is composed of rows which are built from *different* functions (which – depending on the differential operator L – might not even be radial). The results for the non-singularity of interpolation matrices, however, are based on the fact that A is generated by a *single* function φ.

An indication of the success of Kansa's method (which has not yet been shown to be well-posed) are the early papers [42, 43, 77, 101, 142] and many more since. In his paper [100] Kansa describes three sets of experiments using his method and comments on the superior performance of multiquadrics in terms of computational complexity and accuracy when compared to finite difference methods. Therefore, it remains an interesting open question whether the well-posedness of Kansa's method can be established at least for certain configurations of centers. Moreover, Kansa's suggestion to use variable shape parameters α_j in order to improve accuracy and stability of the problem has very little theoretical support. Except for one paper by Bozzini, Lenarduzzi and Schaback [22] (which addresses only the interpolation setting) this problem has not been addressed in the literature.

Before we describe an alternate approach which does ensure well-posedness of the resulting collocation matrix and which is based on basis functions suitable for scattered Hermite interpolation we would like to point out that in [142] the authors suggest how Kansa's method can be applied to other types of partial differential equation problems such as non-linear elliptic PDEs, systems of elliptic PDEs, and time-dependent parabolic or hyperbolic PDEs.

9.1.2 An Hermite-based Approach

The following symmetric approach is based on scattered Hermite interpolation (see, e.g., [95, 151, 192, 211]), which we now also quickly review. In this context we are given data $\{\boldsymbol{x}_i, L_i f\}$, $i = 1, \ldots, N$, $\boldsymbol{x}_i \in \mathbb{R}^s$ where $\mathcal{L} = \{L_1, \ldots, L_N\}$ is a linearly

independent set of continuous linear functionals. We try to find an interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j L_j^{\boldsymbol{\xi}} \varphi(\|\boldsymbol{x} - \boldsymbol{\xi}\|), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{60}$$

satisfying

$$L_i \mathcal{P}f = L_i f, \qquad i = 1, \ldots, N.$$

We have used $L^{\boldsymbol{\xi}}$ to indicate that the functional L acts on φ viewed as a function of the second argument $\boldsymbol{\xi}$. The linear system $A\boldsymbol{c} = Lf$ which arises in this case has matrix entries

$$A_{ij} = L_i L_j^{\boldsymbol{\xi}} \varphi, \qquad i, j = 1, \ldots, N. \tag{61}$$

In the references mentioned at the beginning of this subsection it is shown that A is non-singular for the same classes of φ as given for scattered data interpolation in our earlier sections.

Remark: It should be pointed out that this formulation of Hermite interpolation is very general and goes considerably beyond the standard notion of Hermite interpolation (which usually refers to interpolation of successive derivative values). Here any kind of linear functional are allowed as long as the set $\mathcal{L}$ is linearly independent.

We illustrate this approach with a simple example using derivative functionals.

Example: Let data $\{\boldsymbol{x}_i, f(\boldsymbol{x}_i)\}_{i=1}^{n}$ and $\{\boldsymbol{x}_i, \frac{\partial f}{\partial x}(\boldsymbol{x}_i)\}_{i=n+1}^{N}$ with $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$ be given. Then

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{n} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{x}_j\|) - \sum_{j=n+1}^{N} c_j \frac{\partial \varphi}{\partial x}(\|\boldsymbol{x} - \boldsymbol{x}_j\|),$$

and

$$A = \begin{bmatrix} \Phi & -\Phi_x \\ \Phi_x & -\Phi_{xx} \end{bmatrix},$$

with

$$\begin{aligned}
\Phi_{ij} &= \varphi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i, j = 1, \ldots, n, \\
-\Phi_{x,ij} &= -\frac{\partial \varphi}{\partial x}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i = 1, \ldots, n,\ j = n+1, \ldots, N, \\
\Phi_{x,ij} &= \frac{\partial \varphi}{\partial x}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i = n+1, \ldots, N,\ j = 1, \ldots, n, \\
\Phi_{xx,ij} &= \frac{\partial^2 \varphi}{\partial x^2}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \qquad i, j = n+1, \ldots, N.
\end{aligned}$$

Now we describe an alternative collocation method based on the generalized interpolation theory just reviewed. Assume we are given the same PDE (56) with boundary conditions (57) as in the section on Kansa's method. In order to be able to apply the results from scattered Hermite interpolation to ensure the non-singularity of the collocation matrix we propose the following expansion for the unknown function u:

$$u(\boldsymbol{x}) = \sum_{j=1}^{\#B} c_j \varphi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|) + \sum_{j=\#B+1}^{N} c_j L^{\boldsymbol{\xi}}[\varphi](\|\boldsymbol{x} - \boldsymbol{\xi}_j\|), \tag{62}$$

where $\#B$ denotes the number of nodes on the boundary of Ω, and $L^{\boldsymbol{\xi}}$ is the differential operator used in (56), but acting on φ viewed as a function of the second argument, i.e., $L[\varphi]$ is equal to $L^{\boldsymbol{\xi}}[\varphi]$ up to a possible difference in sign. Note the difference in notation. In (60) L is a linear functional, and in (62) a differential operator.

This expansion for u leads to a collocation matrix A which is of the form

$$A = \begin{bmatrix} \Phi & L^{\boldsymbol{\xi}}[\Phi] \\ L[\Phi] & L[L^{\boldsymbol{\xi}}[\Phi]] \end{bmatrix}, \tag{63}$$

where the four blocks are generated as follows:

$$\begin{aligned} \Phi_{ij} &= \varphi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i, \boldsymbol{\xi}_j \in B, \\ L^{\boldsymbol{\xi}}[\Phi]_{ij} &= L^{\boldsymbol{\xi}}[\varphi](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i, \in B,\ \boldsymbol{\xi}_j \in I, \\ L[\Phi]_{ij} &= L[\varphi](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i \in I,\ \boldsymbol{\xi}_j \in B, \\ L[L^{\boldsymbol{\xi}}[\Phi]]_{ij} &= L[L^{\boldsymbol{\xi}}[\varphi]](\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|), \qquad \boldsymbol{x}_i, \boldsymbol{\xi}_j \in I. \end{aligned}$$

The matrix (63) is of the same type as the scattered Hermite interpolation matrices (61), and therefore non-singular as long as φ is chosen appropriately. Thus, viewed using the new expansion (62) for u, the collocation approach is certainly well-posed. Also, note that although A consists of four blocks now, it still is of the same size, namely $N \times N$, as the collocation matrix (59) obtained for Kansa's approach.

Both of the methods described in this section have been implemented for many different applications. A thorough comparison of the two methods was reported in [156]. All in all the Hermite method seems to perform slightly better than Kansa's method. Especially for the cases in which relatively many interior points are used (which is where the methods differ). Also, the matrices for the Hermite method generally have smaller condition numbers. An advantage of the Hermite approach over Kansa's method is that for many differential operators L the collocation matrices resulting from the Hermite approach are symmetric. Therefore the amount of computation can be reduced considerably, which is important for larger problems. Kansa's method has the advantage of being simpler to implement (since less derivatives of the basis functions are required). Another advantage of Kansa's method is that is can be easily adapted for nonlinear elliptic PDEs (see, e.g., [59, 142]).

Since the methods described above were both originally used with globally supported basis functions, the same concerns as for interpolation problems about stability and numerical efficiency apply. Two recent papers by Ling and Kansa [119, 120] address these issues. In particular, they develop a preconditioner in the spirit of the one described in Section 8.3.2, and describe their experience with a domain decomposition algorithm. An attempt to obtain an efficient implementation of the Hermite based collocation method is a version of the greedy algorithm mentioned in Section 8.5 by Hon, Schaback and Zhou [89].

A convergence analysis for the symmetric method was formulated by Franke and Schaback [72, 73]. The error estimates established in [72, 73] require the solution of the PDE to be very smooth. Therefore, meshfree radial basis function collocation techniques are especially well suited for (high-dimensional) PDE problems with smooth solutions on possibly irregular domains. Some numerical evidence for convergence rates of the symmetric collocation method is given in the papers [56, 98, 156]. Due to the

known counterexamples [88] for the non-symmetric method, a convergence analysis is still lacking for that method.

Recently, Miranda [140] has shown that Kansa's method will be well-posed if it is combined with so-called *R-functions*. This idea was also used by Höllig and his co-workers in their development of WEB-splines (see, e.g., [87]).

Other recent papers investigating various aspects of radial basis function collocation are, e.g., [34] by Cheng, Golberg, Kansa and Zammito, [68] by Fedoseyev, Friedman and Kansa, [102] by Kansa and Hon, [110] by Larsson and Fornberg, [112] by Leitão, and [124] by Mai-Duy and Tran-Cong.

For example, in the paper [68] it is suggested that the collocation points on the boundary are also used to satisfy the PDE. However, this adds a set of extra equations to the problem, and therefore one should also use some additional basis functions in the expansion (58). It is suggested in [68] that these centers lie outside the domain Ω. The motivation for this modification is the well-known fact that both for interpolation and collocation with radial basis functions the error is largest near the boundary. In various numerical experiments this strategy is shown to improve the accuracy of Kansa's basic non-symmetric method. It should be noted that there is once more no theoretical foundation for this method.

Larsson and Fornberg [110] compare Kansa's basic collocation method, the modification just described, and the Hermite-based symmetric approach mentioned earlier. Using multiquadric basis functions in a standard implementation they conclude that the symmetric method is the most accurate, followed by the non-symmetric method with boundary collocation. The reason for this is the better conditioning of the system for the symmetric method. Larsson and Fornberg also discuss an implementation of the three methods using the complex Contour-Padé integration method mentioned in Section 8.1. With this technique stability problems are overcome, and it turns out that both the symmetric and the non-symmetric method perform with comparable accuracy. Boundary collocation of the PDE yields an improvement only if these conditions are used as additional equations, i.e., by increasing the problem size. It should also be noted that often the most accurate results were achieved with values of the multiquadric shape parameter α which would lead to severe ill-conditioning using a standard implementation, and therefore these results could be achieved only using the complex integration method. Moreover, in [110] radial basis function collocation is deemed to be far superior in accuracy than standard second-order finite differences or a standard Fourier-Chebyshev pseudospectral method.

Leitão [112] applies the symmetric collocation method to a fourth-order Kirchhoff plate bending problem, and emphasizes the simplicity of the implementation of the radial basis function collocation method. And, finally, Mai-Duy and Tran-Cong [124] suggest a collocation method for which the basis functions are taken to be anti-derivatives of the usual radial basis functions.

9.2 Galerkin Methods

A variational approach to the solution of PDEs with RBFs has so far only been considered by Wendland [202, 203]. In [203] he studies the Helmholtz equation with natural

boundary conditions, i.e.,

$$\begin{aligned} -\Delta u + u &= f \qquad \text{in } \Omega, \\ \frac{\partial}{\partial \nu} u &= 0 \qquad \text{on } \partial\Omega, \end{aligned}$$

where ν denotes the outer unit normal vector. The classical Galerkin formulation then leads to the problem of finding a function $u \in H^1(\Omega)$ such that

$$a(u, v) = (f, v)_{L_2(\Omega)} \qquad \text{for all } v \in H^1(\Omega),$$

where $(f, v)_{L_2(\Omega)}$ is the usual L_2 inner product, and for the Helmholtz equation the bilinear form a is given by

$$a(u, v) = \int_\Omega (\nabla u \cdot \nabla v + uv) dx.$$

In order to obtain a numerical scheme the infinite-dimensional space $H^1(\Omega)$ is replaced by some finite-dimensional subspace $\mathcal{S}_\mathcal{X} \subseteq H^1(\Omega)$, where $\mathcal{X}$ is some computational grid to be used for the solution. In the context of RBFs $\mathcal{S}_\mathcal{X}$ is taken as

$$\mathcal{S}_\mathcal{X} = \text{span}\{\phi(\| \cdot - x_j \|_2),\ x_j \in \mathcal{X}\}.$$

This results in a square system of linear equations for the coefficients of $u_\mathcal{X} \in \mathcal{S}_\mathcal{X}$ determined by

$$a(u_\mathcal{X}, v) = (f, v)_{L_2(\Omega)} \qquad \text{for all } v \in \mathcal{S}_\mathcal{X}.$$

For more on the Galerkin method (in the context of finite elements) see, e.g., [23, 24]. It was shown in [202] that for those RBFs (globally as well as locally supported) whose Fourier transform decays like $(1 + \| \cdot \|_2)^{-2\beta}$ the following convergence estimate holds:

$$\| u - u_\mathcal{X} \|_{H^1(\Omega)} \leq C h^{\sigma - 1} \| u \|_{H^\sigma(\Omega)}, \tag{64}$$

where h is the meshsize of $\mathcal{X}$, the solution satisfies the regularity requirements $u \in H^\sigma(\Omega)$, and where the convergence rate is determined by $\beta \geq \sigma > s/2 + 1$. For Wendland's compactly supported RBFs this implies that functions which are in $C^{2\kappa}$ and strictly positive definite on $\mathbb{R}^s$ satisfying $\kappa \geq \sigma - \frac{s+1}{2}$ will have $\mathcal{O}(h^{\kappa + (s-1)/2})$ convergence order, i.e., the C^0 function $\varphi_{3,0} = (1-r)_+^2$ yields $\mathcal{O}(h)$ and the C^2 function $\varphi_{3,1} = (1-r)_+^4(4r+1)$ delivers $\mathcal{O}(h^2)$ convergence in $\mathbb{R}^3$. As with the convergence estimate for symmetric collocation there is a link between the regularity requirements on the solution and the space dimension s. Also, so far, the theory is only established for PDEs with natural boundary conditions.

The convergence estimate (64) holds for the non-stationary setting, i.e., if we are using compactly supported basis functions, for fixed support radii. By the same argumentation as earlier, one will want to switch to the stationary setting and employ a multilevel algorithm in which the solution at each step is updated by a fit to the most recent residual.

The convergence estimate (64) holds for the non-stationary setting, i.e., if we are using compactly supported basis functions, for fixed support radii. By the same argumentation as used in Section 8, one will want to switch to the stationary setting

and employ a multilevel algorithm in which the solution at each step is updated by a fit to the most recent residual. This should ensure both convergence and numerical efficiency.

Here is the variant of the stationary multilevel interpolation algorithm listed above for the weak form solution of PDEs (see [203]):

Algorithm (Multilevel Galerkin)

$u_0 = 0$.

For k from 1 to K do

Find $u_k \in \mathcal{S}_{\mathcal{X}_k}$ such that $a(u_k, v) = (f, v) - a(u_{k-1}, v)$ for all $v \in \mathcal{S}_{\mathcal{X}_k}$.

Update $u_k \leftarrow u_{k-1} + u_k$.

end

It may come as a little bit of a surprise that this algorithm *does not converge in general* (see Tab. 1 in [203]).

Since the weak formulation can be interpreted as a Hilbert space projection method, Wendland was able to show that a modified version of the multilevel Galerkin algorithm, namely

Algorithm (Nested Multilevel Galerkin)

Fix K and $M \in \mathbb{N}$, and set $v_0 = 0$.

For j from 0 while $residual > tolerance$ to M do

Set $u_0 = v_j$.

Apply the k-loop of the previous algorithm and denote the result with $\hat{u}(v_j)$.

Set $v_{j+1} = \hat{u}(v_j)$.

end

does converge. In fact, using this algorithm Wendland proves, and also observes numerically, convergence which is at least linear (see Theorem 3 and Tab. 2 in [203]). The important difference between the two multilevel Galerkin algorithms is the added outer iteration in the nested version which is a well-known idea from linear algebra introduced in 1937 by Kaczmarz [99]. A proof of the linear convergence for general Hilbert space projection methods coupled with Kaczmarz iteration can be found in [189]. This alternate projection idea is also the fundamental ingredient in the convergence proof of the domain decomposition method of Beatson, Light and Billings [13] described in the previous section. We mention here that in the multigrid literature Kaczmarz' method is frequently used as a smoother (see e.g. [128]).

Remark: Aside from difficulties with Dirichlet (or sometimes called *essential*) boundary conditions, Wendland reports that the numerical evaluation of the weak-form integrals presents a major problem for the radial basis function Galerkin approach. Both of these difficulties are also well-known in many other flavors of meshfree weak-form

methods. An especially promising solution to the issue of Dirichlet boundary conditions seems to be the use of R-functions as proposed by Höllig and Reif in the context of WEB-splines (see, e.g., [87] or our earlier discussion in the context of collocation methods).

Many other meshfree methods for the solution of partial differential equations in the weak form appear in the (mostly engineering) literature. These methods come under such names as smoothed particle hydrodynamics (SPH) (e.g., [141]), reproducing kernel particle method (RKPM) (see, e.g., [115, 122]), point interpolation method (PIM) (see, [121]), element free Galerkin method (EFG) (see, e.g., [17]), meshless local Petrov-Galerkin method (MLPG) [3], h-p-cloud method [41], partition of unity finite element method (PUFEM) [5, 134], or generalized finite element method (GFEM) [4]. Most of these methods are based on the moving least squares approximation method discussed in section 7.

There are two recent books by Atluri [2] and Liu [121] summarizing many of these methods. However, these books focus mostly on a survey of the various methods and related computational and implementation issues with little emphasis on the mathematical foundation of the methods. The recent survey paper [4] by Babuška, Banerjee and Osborn, fills a large part of this void.

9.3 Applications in Computational Nanotechnology and Quantum Chemistry

Applications of meshfree methods to problems at the nanoscale are still at their initial stages. So far, not much has been published on the subject.

The group at Northwestern University around Wing Kam Liu (see, e.g., [157, 158, 164, 215]) has mostly been modelling problems in nanomechanics exploring such properties as Young's modulus, bending stiffness, buckling criteria, and tensile and compressive strenghts. The advantage of the use of meshfree approximation methods is the ability to describe both multi-scale and multi-physics problems. For example, in [158] and [215] the mechanics of carbon nanotubes are studied. In particular the authors study how the mechanical properties of nanotubes are affected if they are filled with fullerenes. This is a phenomenon that has been observed in experiments, but for which no computer model had existed previously. The study is performed by coupling a molecular dynamics model (for the C_{60} fullerenes) to a meshfree continuum model (for the nanotubes). This approach bridges the bonding potential with the continuum property of the material, and additionally serves as a useful tool for evaluating the bonding potentials. The meshfree method used for these simulations is the reproducing kernel particle method (RKPM) which is closely related to the multilevel approximate moving least squares approximation method described in Section 7. Other examples studied in [158] include nanoropes, nanoelectromechanical systems, nanosensors, and nanotube-reinforced polymers.

The Princeton group with Herschel Rabitz [91, 92, 93] uses radial basis functions and reproducing kernel interpolation to solve the Schrödinger equation in quantum fluid dynamics. The authors conclude that for 2D-computations the RBF-based method is comparable to standard split-operator or Chebyshev expansion methods, while for 3D-computations it is superior due to its low storage requirements and uniform accuracy.

Moreover, by constructing the reproducing kernel associated with the differential operator for the Schrödinger equation one obtains an optimal interpolation method (as explained in Section 5) that can be used on arbitrary sets of grid points in arbitrary dimensions. This is seen by the authors as a significant advantage over other methods such as discrete variable representations (DVRs), Fourier spectral methods, the distributed Gaussian basis method (DGB), or distributed approximating functionals (DAFs).

Finally, the group around Michael Griebel at the University of Bonn also are interested in both meshfree approximation methods and applications in nanotechnology (see the contribution of M. Griebel in this handbook).

References

[1] M. G. Armentano, *Error estimates in Sobolev spaces for moving least square approximations*, SIAM J. Numer. Anal. 39 (2001), no. 1, 38–51.

[2] S. N. Atluri and S. Shen, *The Meshless Local Petrov-Galerkin (MLPG) Method*, Tech Science Press, Encino, CA, 2002.

[3] S. N. Atluri and T. Zhu, *A new meshless local Petrov-Galerkin (MLPG) approach in computational mechanics*, Comput. Mech. 22 (1998), 117–127.

[4] I. Babuška, U. Banerjee and J. E. Osborn, Survey of meshless and generalized finite element methods: A unified approach, University of Texas, Austin, TICAM Report 02-03, 2002.

[5] I. Babuška and M. Melenk, *The partition of unity method*, Int. J. Numer. Meths. Eng. (40) 1997, 727–758.

[6] G. Backus and F. Gilbert, *The resolving power of gross earth data*, Geophys. J. R. Astr. Soc. (16) 1968, 169–205.

[7] K. Ball, *Eigenvalues of Euclidean distance matrices*, J. Approx. Theory, 68 (1992), 74–82.

[8] K. Ball, N. Sivakumar, and J. D. Ward, *On the sensitivity of radial basis interpolation to minimal data separation distance*, Constr. Approx., 8 (1992), 401–426.

[9] B. J. C. Baxter, *Norm estimates for inverses of distance matrices*, in Mathematical Methods in Computer Aided Geometric Design II, T. Lyche and L. Schumaker (eds.), Academic Press, New York, 1992, 9–18.

[10] B. J. C. Baxter, *Preconditioned conjugate gradients, radial basis functions, and Toeplitz matrices*, Comput. Math. Appl. 43 (2002), 305–318.

[11] R. K. Beatson, J. B. Cherrie, and C. T. Mouat, *Fast fitting of radial basis functions: methods based on preconditioned GMRES iteration*, Adv. Comput. Math. 11 (1999), 253–270.

[12] R. K. Beatson, G. Goodsell, and M. J. D. Powell, *On multigrid techniques for thin plate spline interpolation in two dimensions*, in The Mathematics of Numerical Analysis, Lectures in Appl. Math., 32, Amer. Math. Soc., Providence, RI, 1996, 77–97.

[13] R. K. Beatson, W. A. Light, and S. Billings, *Fast solution of the radial basis function interpolation equations: domain decomposition methods*, SIAM J. Sci. Comput. 22 (2000), 1717–1740.

[14] R. K. Beatson and M. J. D. Powell, *An iterative method for thin plate spline interpolation that employs approximations to Lagrange functions*, in Numerical Analysis 1993, Dundee, 1993), xx (ed.), Pitman Res. Notes Math. Ser., 303, Longman Sci. Tech., Harlow, 1994, 17–39.

[15] J. Behrens, A. Iske, and M. Käser, *Adaptive meshfree method of backward characteristics for nonlinear transport equations*, in Lecture Notes in Computer Science and Engineering Vol.26: Meshfree Methods for Partial Differential Equations, M. Griebel and M. A. Schweitzer (eds.), Springer Verlag, 2002, 21–36.

[16] A. Bejancu, Jr., *Local accuracy for radial basis function interpolation on finite uniform grids*, J. Approx. Theory 99 (1999), 242–257.

[17] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming and P. Krysl, *Meshless methods: an overview and recent developments*, Comp. Meth. Appl. Mech. Eng. (139) 1996, 3–47.

[18] S. Bochner, *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig, 1932.

[19] S. Bochner, *Monotone Funktionen, Stieltjes Integrale und harmonische Analyse*, Math. Ann., 108 (1933), 378–410.

[20] C. de Boor, *Approximation order without quasi-interpolants*, in Approximation Theory VII, E. W. Cheney, C. Chui, and L. Schumaker (eds.), Academic Press, New York, 1992, 1–18.

[21] L. P. Bos and K. Šalkauskas, *Moving least-squares are Backus-Gilbert optimal*, J. Approx. Theory 59 (1989), 267–275.

[22] M. Bozzini, L. Lenarduzzi, and R. Schaback, *Adaptive interpolation by scaled multiquadrics*, Adv. in Comp. Math. 16 (2002), 375–387.

[23] D. Braess, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press (Cambridge), 1997.

[24] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer Verlag (New York), 1994.

[25] M. D. Buhmann, *Radial functions on compact support*, Proc. Edin. Math. Soc. II 41 (1998), 33-46.

[26] M. D. Buhmann, *Radial basis functions*, in Acta Numerica 2000, Vol. 9, Cambridge University Press (2000), 1–38.

[27] M. D. Buhmann, *Radial Basis Functions : Theory and Implementations*, Cambridge University Press, Cambridge, 2003.

[28] R. E. Carlson and T. A. Foley, *Interpolation of track data with radial basis methods*, Comput. Math. Appl., 24 (1992), 27–34.

[29] R. E. Carlson and B. K. Natarajan, *Sparse approximate multiquadric interpolation*, Comput. Math. Appl., 27 (1994), 99–108.

[30] R. Chan and G. Strang, *Toeplitz equations by conjugate gradients with circulant preconditioners*, SIAM J. Sci. Statist. Comput. 10 (1989), 104–119.

[31] K. F. Chang, *Strictly positive definite functions*, J. Approx. Theory, 87 (1996), 148–158. Addendum: J. Approx. Theory, 88 (1997), 384.

[32] C. S. Chen, M. Ganesh, M. A. Golberg, and A. H.-D. Cheng, *Multilevel compact radial functions based computational schemes for some elliptic problems*, Comput. Math. Appl. 43 (2002), 359–378.

[33] E. W. Cheney and W. A. Light, *A Course in Approximation Theory*, Brooks Cole, 1999.

[34] A. H. D. Cheng, M. A. Golberg, E. J. Kansa, and G. Zammito, *Exponential convergence and H-c multiquadric collocation method for partial differential equations*, Numer. Methods Partial Differential Equations 19 (2003), no. 5, 571–594.

[35] C. K. Chui, *Multivariate Splines*, Reg. Conf. Ser. in Appl. Math. (SIAM), 1988.

[36] K. C. Chung and T. H. Yao, *On lattices admitting unique Lagrange interpolations*, SIAM J. Numer. Anal., 14 (1977), 735–743.

[37] W. S. Cleveland and C. L. Loader, *Smoothing by local regression: Principles and methods*, in Statistical Theory and Computational Aspects of Smoothing, W. Haerdle and M. G. Schimek (eds.), Springer, (1996), 10–49.

[38] E. L. De Forest, *On some methods of interpolation applicable to the graduation of irregular series*, Annual Report of the Board of Regents of the Smithsonian Institution for 1871, (1873) 275–339.

[39] E. L. De Forest, *Additions to a memoir on methods of interpolation applicable to the graduation of irregular series*, Annual Report of the Board of Regents of the Smithsonian Institution for 1873, (1874) 319–353.

[40] T. A. Driscoll and B. Fornberg, *Interpolation in the limit of increasingly flat radial basis functions*, Comput. Math. Appl. 43 (2002), 413–422.

[41] C. A. Duarte and J. T. Oden, *H-p clouds – an h-p meshless method*, Num. Meth. for Part. Diff. Eq. 4, John Wiley & Sons, 1996.

[42] M. R. Dubal, *Construction of three-dimensional black-hole initial data via multiquadrics*, Phys. Rev. D, 45 (1992), 1178–1187.

[43] M. R. Dubal, *Domain decomposition and local refinement for multiquadric approximations. I: Second-order equations in one-dimension*, J. Appl. Sc. Comp., 1 (1994), 146–171.

[44] J. Duchon, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, Rev. Francaise Automat. Informat. Rech. Opér., Anal. Numer., 10 (1976), 5–12.

[45] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in Constructive Theory of Functions of Several Variables, Oberwolfach 1976, W. Schempp and K. Zeller (eds.), Springer Lecture Notes in Math. 571, Springer-Verlag, Berlin, 1977, 85–100.

[46] J. Duchon, *Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines*, Rev. Francaise Automat. Informat. Rech. Opér., Anal. Numer., 12 (1978), 325–334.

[47] J. Duchon, *Fonctions splines homogènes à plusiers variables*, Université de Grenoble, 1980.

[48] N. Dyn, *Interpolation of scattered data by radial functions*, in Topics in Multivariate Approximation, C. K. Chui, L. L. Schumaker, and F. Utreras (eds.), Academic Press, New York, 1987, 47–61.

[49] N. Dyn, *Interpolation and approximation by radial and related functions*, in Approximation Theory VI, C. Chui, L. Schumaker, and J. Ward (eds.), Academic Press, New York, 1989, 211–234.

[50] N. Dyn and D. Levin, *Iterative solution of systems originating from integral equations and surface interpolation*, SIAM J. Numer. Anal., 20 (1983), 377–390.

[51] N. Dyn, D. Levin, and S. Rippa, *Numerical procedures for surface fitting of scattered data by radial functions*, SIAM J. Sci. Statist. Comput., 7 (1986), 639–659.

[52] T. Evgeniou, M. Pontil, and T. Poggio, *Regularization networks and support vector machines*, Adv. Comput. Math. 13 (2000), no. 1, 1–50.

[53] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman & Hall (New York), 1996.

[54] R. Farwig, *Multivariate interpolation of scattered data by moving least squares methods*, Algorithms for Approximation, Oxford Univ. Press, New York, 193–211, 1987.

[55] R. Farwig, *Rate of convergence of moving least squares interpolation methods: the univariate case*, Progress in Approximation Theory, Academic Press, Boston, MA, 313–327, 1991.

[56] G. E. Fasshauer, *Solving partial differential equations by collocation with radial basis functions*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 131–138.

[57] G. E. Fasshauer, *On smoothing for multilevel approximation with radial basis functions*, in Approximation Theory IX, Vol.II: Computational Aspects, Charles K. Chui, and L. L. Schumaker (eds.), Vanderbilt University Press, 1999, 55–62.

[58] G. E. Fasshauer, *Solving differential equations with radial basis functions: multilevel methods and smoothing*, Adv. in Comp. Math. (11) 1999, 139–159.

[59] G. E. Fasshauer, *Newton iteration with multiquadrics for the solution of nonlinear PDEs*, Comput. Math. Applic. (43) 2002, 423–438.

[60] G. E. Fasshauer, *Matrix-free multilevel moving least-squares methods*, in Approximation Theory X: Wavelets, Splines, and Applications, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt University Press, 2002, 271–281.

[61] G. E. Fasshauer, *Approximate moving least-squares approximation with compactly supported weights*, in Lecture Notes in Computer Science and Engineering Vol.26: Meshfree Methods for Partial Differential Equations, M. Griebel and M. A. Schweitzer (eds.), Springer Verlag, 2002, 105-116.

[62] G. E. Fasshauer, *Approximate moving least-squares approximation for time-dependent PDEs*, in WCCM V, Fifth World Congress on Computational Mechanics (http://wccm.tuwien.ac.at), H. A. Mang, F. G. Rammerstorfer, and J. Eberhardsteiner (eds.), Vienna University of Technology, 2002.

[63] G. E. Fasshauer, *Approximate moving least-squares approximation: A fast and accurate multivariate approximation method*, in Curve and Surface Fitting: Saint-Malo 2002, A. Cohen, J.-L. Merrien, and L. L. Schumaker (eds.), Nashboro Press, 2003, 139–148.

[64] G. E. Fasshauer, *Toward approximate moving least squares approximation with irregularly spaced centers*, to appear in Computer Methods in Applied Mechanics & Engineering.

[65] G. E. Fasshauer and J. W. Jerome, *Multistep approximation algorithms: Improved convergence rates through postconditioning with smoothing kernels*, Adv. in Comput. Math. (10) 1999, 1–27.

[66] A. C. Faul and M. J. D. Powell, *Proof of convergence of an iterative technique for thin plate spline interpolation in two dimensions*, Adv. Comput. Math. 11 (1999), 183–192.

[67] A. C. Faul and M. J. D. Powell, *Krylov subspace methods for radial basis function interpolation*, Numerical Analysis 1999 (Dundee), Chapman & Hall/CRC, Boca Raton, FL, 2000, 115–141,

[68] A. I. Fedoseyev, M. J. Friedman and E. J. Kansa, *Improved multiquadric method for elliptic partial differential equations via PDE collocation on the boundary*, Comput. Math. Applic. 43 (2002), 439–455.

[69] M. S. Floater and A. Iske, *Multistep scattered data interpolation using compactly supported radial basis functions*, J. Comput. Applied Math., 73 (1996), 65–78.

[70] T. A. Foley, *Near optimal parameter selection for multiquadric interpolation*, J. Appl. Sc. Comp., 1 (1994), 54–69.

[71] B. Fornberg and G. Wright, *Stable computation of multiquadric interpolants for all values of the shape parameter*, preprint, University of Colorado.

[72] C. Franke and R. Schaback, *Solving partial differential equations by collocation using radial basis functions*, Appl. Math. Comp. (93) 1998, 73–82.

[73] C. Franke and R. Schaback, *Convergence orders of meshless collocation methods using radial basis functions*, Adv. in Comput. Math. (8) 1998, 381–399.

[74] R. Franke, *Scattered data interpolation: tests of some methods*, Math. Comp., 48 (1982), 181–200.

[75] M. Frigo and S. G. Johnson, *FFTW: The fastest Fourier transform in the West (C library)*, `http://www.fftw.org/`.

[76] F. Girosi, *An equivalence between sparse approximation and support vector machines*, MIT Artificial Intelligence Lab, Report 1606, 1997.

[77] M. A. Golberg, C. S. Chen, and S. R. Karur, *Improved multiquadric approximation for partial differential equations*, Eng. Anal. with Bound. Elem. (18) 1996, 9–17.

[78] M. von Golitschek and L. L. Schumaker, *Data fitting by penalized least squares*, in Algorithms for Approximation II, M. G. Cax and J. C. Mason (eds.), Chapman & Hall, London, 1990, 210–227.

[79] M. Golomb and H. F. Weinberger, *Optimal approximation and error bounds*, in On numerical approximation, R. E. Langer(ed.), University of Wisconsin Press, 1959, 117–190.

[80] J. P. Gram, *Über Entwicklung reeler Functionen in Reihen mittelst der Methode der kleinsten Quadrate*, J. Math. 94 (1883), 41–73.

[81] K. Guo, S. Hu, and X. Sun, *Conditionally positive definite functions and Laplace-Stieltjes integrals*, J. Approx. Theory, 74 (1993), 249–265.

[82] S. J. Hales and J. Levesley, *Error estimates for multilevel approximation using polyharmonic splines*, Numer. Algorithms 30 (2002), 1–10.

[83] R. L. Hardy, *Multiquadric equations of topography and other irregular surfaces*, J. Geophys. Res., 76 (1971), 1905–1915.

[84] R. L. Hardy, *Geodetic applications of multiquadric analysis*, AVN Allg. Vermess. Nachr., 79 (1972), 389–406.

[85] R. L. Hardy, *Research results in the application of multiquadric equations to surveying and mapping problems*, Survg. Mapp., 35 (1975), 321–332.

[86] S. Hartmann, *Multilevel-Fehlerabschätzung bei der Interpoaltion mit radialen Basisfunktionen*, Ph.D. dissertation, Universität Göttingen, 1998.

[87] K. Höllig, *Finite Element Methods With B-Splines*, SIAM Frontiers in Applied Mathematics no. 26, Philadelphia, 2003.

[88] Y. C. Hon and R. Schaback, *On nonsymmetric collocation by radial basis functions*, Appl. Math. Comput. 119 (2001), 177–186.

[89] Y. C. Hon, R. Schaback, and X. Zhou, *An adaptive greedy algorithm for solving large RBF collocation problems*, Numer. Algorithms 32 (2003), 13–25.

[90] Y. C. Hon and Z. Wu, *Additive Schwarz domain decomposition with a radial basis approximation*, Int. J. Appl. Math. 4 (2000), 81–98.

[91] X.-G. Hu, T.-S. Ho and H. Rabitz, *Solving the bound-state Schrödinger equation by reproducing kernel interpolation*, Phys. Rev. E 61 (2000), 2074–2085.

[92] X.-G. Hu, T.-S. Ho and H. Rabitz, *Solution of the quantum fluid dynamical equations with radial basis function interpolation*, Phys. Rev. E 61 (2000), 5967–5976.

[93] X.-G. Hu, T.-S. Ho and H. Rabitz, *Multivariate radial basis interpolation for solving quantum fluid dynamical equations*, Comput. Math. Appl.43 (2002), 525–538.

[94] M. S. Ingber, C. S. Chen, and J. A. Tanski, *A mesh free approach using radial basis functions and parallel domain decomposition for solving three dimensional diffusion equations*, University of Nevada, preprint 2003.

[95] A. Iske, *Reconstruction of functions from generalized Hermite-Birkhoff data*, in Approximation Theory VIII, Vol. 1: Approximation and Interpolation, C. Chui, and L. Schumaker (eds.), World Scientific Publishing, Singapore, 1995, 257–264.

[96] A. Iske, *Hierarchical scattered data filtering for multilevel interpolation schemes*, in Mathematical Methods for Curves and Surfaces: Oslo 2000, T. Lyche and L. L. Schumaker (eds.), Vanderbilt University Press (2001), 211–220.

[97] A. Iske, *On the approximation order and numerical stability of local Lagrange interpolation by polyharmonic splines*, in Modern Developments in Multivariate Approximation, W. Haussmann, K. Jetter, M. Reimer, and J. Stöckler (eds.), ISNM 145, Birkhäuser Verlag (2003), 153–165.

[98] B. Jumarhon, S. Amini and K. Chen, *Hermite type collocation method using radial basis functions*, preprint, 1998.

[99] S. Kaczmarz, *Approximate solution of systems of linear equations* (originally published as: Angenäherte Auflösung von Systemen linearer Gleichungen, Bulletin International de l'Academie Polonaise des Sciences, Lett. A, 1937, 355–357), Int. J. Control 57 (1993) 1269–1271.

[100] 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., 19 (1990), 147–161.

[101] E. J. Kansa, *A strictly conservative spatial approximation scheme for the governing engineering and physics equations over irregular regions and inhomogeneous scattered nodes*, Comput. Math. Appl., 24 (1992), 169–190.

[102] E. J. Kansa and Y. C. Hon, *Circumventing the ill-conditioning problem with multiquadric radial basis functions: Applications to elliptic partial differential equations*, Comput. Math. Applic. 39 (2000), 123–137.

[103] D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing* (2nd ed.), Brooks/Cole, Pacific Grove, CA, 1996.

[104] S. Kunis and D. Potts, *NFFT, Softwarepackage (C-library)*, Universität Lübeck, `http://www.math.uni-luebeck.de/potts/nfft/`, 2002.

[105] S. Kunis, D. Potts, and G. Steidl, *Fast Fourier transforms at nonequispaced knots: A user's guide to a C-library*, Universität Lübeck, `http://www.math.uni-luebeck.de/potts/nfft/`, 2002.

[106] J. Kybic, T. Blu, and M. Unser, *Generalized sampling: A variational approach – Part I: Theory*, IEEE Trans. Signal Proc., 50 (2002), 1965–1976.

[107] J. Kybic, T. Blu, and M. Unser, *Generalized sampling: A variational approach – Part II: Applications*, IEEE Trans. Signal Proc., 50 (2002), 1977–1985.

[108] P. Lancaster and K. Šalkauskas, *Surfaces generated by moving least squares methods*, Math. Comp. (37) 1981, 141–158.

[109] P. Lancaster and K. Šalkauskas, *Curve and Surface Fitting*, Academic Press, New York, 1986.

[110] E. Larsson and B. Fornberg, *A numerical study of some radial basis function based solution methods for elliptic PDEs*, Comput. Math. Appl., to appear.

[111] E. Larsson and B. Fornberg, *Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions*, preprint, University of Colorado.

[112] V. M. A. Leitão, *A meshless method for Kirchhoff plate bending problems*, Int. J. Numer. Meth. Engng. 52 (2001), 1107–1130.

[113] D. Levin, *The approximation power of moving least-squares*, Math. Comp. 67 (1998), 1517–1531.

[114] J. Li and Y. C. Hon, *Domain decomposition for radial basis meshless methods*, preprint, University of Nevada.

[115] S. Li and W. K. Liu, *Moving least-square reproducing kernel method Part II: Fourier analysis*, Comp. Meth. Appl. Mech. Eng. (139) 1996, 159–193.

[116] S. Li and W. K. Liu, *Meshfree and particle methods and their applications*, Applied Mechanics Review (55) 2002, 1–34.

[117] W. A. Light and H. Wayne, *Error estimates for approximation by radial basis functions*, in Approximation Theory, Wavelets and Applications, S. P. Singh (ed.), Kluwer, Dordrecht, Netherlands, 1995, 215–246.

[118] W. A. Light and H. Wayne, *On power functions and error estimates for radial basis function interpolation*, J. Approx. Theory 92 (1998), 245–266.

[119] L. Ling and E. J. Kansa, *A least-squares preconditioner for radial basis functions collocation methods*, preprint, Simon Fraser University, 2002.

[120] L. Ling and E. J. Kansa, *Preconditioning for radial basis functions with domain decomposition methods*, preprint, Simon Fraser University, 2003.

[121] G. R. Liu, *Mesh Free Methods: Moving beyond the Finite Element Method*, CRC Press, Boca Raton, FL, 2002.

[122] W. K. Liu, S. Li, and T. Belytschko, *Moving least square reproducing kernel methods (I) Methodology and convergence*, Comp. Meth. Appl. Mech. Eng. (143) 1997, 113–154.

[123] W. R. Madych and S. A. Nelson, *Multivariate interpolation: a variational theory*, manuscript, 1983.

[124] N. Mai-Duy and T. Tran-Cong, *Numerical solution of differential equations using multiquadric radial basis function networks*, Neural Networks 14 (2001), 185–199.

[125] J. C. Mairhuber, *On Haar's theorem concerning Chebyshev approximation problems having unique solutions*, Proc. Am. Math. Soc. (7) 1956, 609–615.

[126] M. Mathias, *Über positive Fourier-Integrale*, Math. Zeit., 16 (1923), 103–125.

[127] V. Maz'ya and G. Schmidt, *On quasi-interpolation with non-uniformly distributed centers on domains and manifolds*, J. Approx. Theory (110) 2001, 125–145.

[128] S. F. McCormick, *Multilevel Projection Methods for Partial Differential Equations*, CBMS-NSF Regional Conference Series in Applied Mathematics 62, SIAM (Philadelphia), 1992.

[129] D. H. McLain, *Drawing contours from arbitrary data points*, Comput. J. (17) 1974, 318–324.

[130] J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. Phys., 30 (1979), 292–304.

[131] J. Meinguet, *An intrinsic approach to multivariate spline interpolation at arbitrary points*, in Polynomial and Spline Approximations, N. B. Sahney (ed.), Reidel, Dordrecht, 1979, 163–190.

[132] J. Meinguet, *Basic mathematical aspects of surface spline interpolation*, in Numerische Integration, G. Hämmerlin (ed.), Birkhäuser, Basel, 1979, 211–220.

[133] J. Meinguet, *Surface spline interpolation: basic theory and computational aspects*, in Approximation Theory and Spline Functions, S. P. Singh, J. H. W. Burry, and B. Watson (eds.), Reidel, Dordrecht, 1984, 127–142.

[134] J. M. Melenk and I. Babuška, *The partition of unity finite element method: basic theory and applications*, Comput. Methods Appl. Mech. Engrg. 139 (1996), 289–314.

[135] C. A. Micchelli, *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx., 2 (1986), 11–22.

[136] C. A. Micchelli and T. J. Rivlin, *A survey of optimal recovery*, in Optimal estimation in approximation theory (Proc. Internat. Sympos., Freudenstadt, 1976), C. A. Micchelli and T. J. Rivlin (eds.), Plenum Press, New York, 1977, 1–54.

[137] C. A. Micchelli and T. J. Rivlin, *Optimal recovery of best approximations*, Resultate Math. 3 (1980), no. 1, 25–32.

[138] C. A. Micchelli and T. J. Rivlin, *Lectures on optimal recovery*, in Numerical analysis, Lancaster 1984, P. R. Turner (ed.), Lecture Notes in Math., 1129, Springer, Berlin, 1985, 21–93.

[139] C. A. Micchelli, T. J. Rivlin, and S. Winograd, *The optimal recovery of smooth functions*, Numer. Math. 26 (1976), no. 2, 191–200.

[140] J. Miranda, *Incorporating R-functions to the Theory of Positive Definite Functions to Solve Partial Differential Equations*, Ph.D. Dissertation, Illinois Institute of Technology, 2004.

[141] J. J. Monaghan, *An introduction to SPH*, Comp. Phys. Comm. 48 (1988), 89–96.

[142] G. J. Moridis and E. J. Kansa, *The Laplace transform multiquadric method: A highly accurate scheme for the numerical solution of linear partial differential equations*, J. Appl. Sc. Comp., 1 (1994), 375–407.

[143] C. T. Mouat, *Fast algorithms and preconditioning techniques for fitting radial basis functions*, Ph.D. Dissertation, University of Canterbury, NZ, 2001.

[144] D. E. Myers, *Kriging, co-Kriging, radial basis functions and the role of positive definiteness*, Comput. Math. Appl., 24 (1992), 139–148.

[145] E. A. Nadaraya, *On estimating regression*, Theor. Probab. Appl. 9 (1964), 141–142.

[146] F. J. Narcowich, R. Schaback and J. D. Ward, *Multilevel interpolation and approximation*, Appl. Comput. Harmon. Anal. 7 (1999), 243–261.

[147] F. J. Narcowich, N. Sivakumar, and J. D. Ward, *On condition numbers associated with radial-function interpolation*, J. Math. Anal. Appl., 186 (1994), 457–485.

[148] F. J. Narcowich and J. D. Ward, *Norms of inverses and condition numbers for matrices associated with scattered data*, J. Approx. Theory, 64 (1991), 69–94.

[149] F. J. Narcowich and J. D. Ward, *Norms of inverses for matrices associated with scattered data*, in Curves and Surfaces, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), Academic Press, New York, 1991, 341–348.

[150] F. J. Narcowich and J. D. Ward, *Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices*, J. Approx. Theory, 69 (1992), 84–109.

[151] F. J. Narcowich and J. D. Ward, *Generalized Hermite interpolation via matrix-valued conditionally positive definite functions*, Math. Comp., 63 (1994), 661–687.

[152] A. Nieslony, D. Potts, and G. Steidl, *Rapid evaluation of radial functions by fast Fourier transforms at nonequispaced knots*, Universität Mannheim, preprint 2003.

[153] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. 33 (1962), 1065–1076.

[154] D. Potts and G. Steidl, *Fast summation at nonequispaced knots by NFFTs*, SIAM J. Sci. Comput. 24 (2003), 2013 - 2037.

[155] M. J. D. Powell, *A new iterative algorithm for thin plate spline interpolation in two dimensions*, Annals of Numerical Mathematics, 4 (1997), 519–528.

[156] H. Power and V. Barraco, *A comparison analysis between unsymmetric and symmetric radial basis function collocation methods for the numerical solution of partial differential equations*, Comput. Math. Appl. 43 (2002), 551–583.

[157] D. Qian, W. K. Liu and R. S. Ruoff , *Mechanics of C60 in nanotubes*, J. Phys. Chem. B 105 (2001), 10753–10758.

[158] D. Qian, G. J. Wagner, W. K. Liu, M.-F. Yu and R. S. Ruoff , *Mechanics of carbon nanotubes*, Appl. Mech. Rev. 55(6) (2002), 495–553.

[159] E. Quak, N. Sivakumar, and J. D. Ward, *Least squares approximation by radial functions*, SIAM J. Numer. Anal., 24 (1993), 1043-1066.

[160] C. H. Reinsch, *Smoothing by spline functions*, Numer. Math. 10 (1967), 177–183.

[161] S. Rippa, *An algorithm for selecting a good value for the parameter c in radial basis function interpolation*, Adv. Comput. Math. 11 (1999), 193–210.

[162] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. 27 (1956), 832–837.

[163] W. Rudin, *Functional Analysis*, McGraw-Hill, New York, 1973.

[164] R. S. Ruoff, D. Qian, W. K. Liu, *Mechanical properties of carbon nanotubes: theoretical predictions and experimental measurements*, C. R. Physique 4 (2003), 993–1008.

[165] Y. Saad and M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear equations*, SIAM J. Sci. Statist. Comput. 7 (1986), 856–869.

[166] S. A. Sarra, *Accurate derivative approximations using radial basis functions*, preprint, Marshall University, 2003.

[167] R. Schaback, *Comparison of radial basis function interpolants*, in Multivariate Approximation: From CAGD to Wavelets, Kurt Jetter and Florencio Utreras (eds.), World Scientific Publishing, Singapore, 1993, 293–305.

[168] R. Schaback, *Lower bounds for norms of inverses of interpolation matrices for radial basis functions*, J. Approx. Theory, 79 (1994), 287–306.

[169] R. Schaback, *Error estimates and condition numbers for radial basis function interpolation*, Adv. in Comput. Math., 3 (1995), 251–264.

[170] R. Schaback, *Optimal recovery in translation-invariant spaces of functions*, Annals of Numerical Mathematics, 4 (1997), 547–556.

[171] R. Schaback, *On the efficiency of interpolation by radial basis functions*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 309–318.

[172] R. Schaback, *Native Hilbert spaces for radial basis functions I*, in New Developments in Approximation Theory, M. W. Müller, M. D. Buhmann, D. H. Mache and M. Felten (eds.), Birkhäuser, Basel, 1999, 255–282.

[173] R. Schaback, *A unified theory of radial basis functions. Native Hilbert spaces for radial basis functions II*, J. Comput. Appl. Math. 121 (2000), 165–177.

[174] R. Schaback, *Remarks on meshless local construction of surfaces*, in The Mathematics of Surfaces, IX, R. Cipolla and R. Martin (eds.), Springer, 2000, 34–58.

[175] R. Schaback, *Stability of radial basis function interpolants*, Approximation Theory X, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt Univ. Press, Nashville, TN, 2002, 433–440.

[176] R. Schaback, *Multivariate interpolation by polynomials and radial basis functions*, preprint, Universität Göttingen, 2002.

[177] R. Schaback, *On the scope of meshless kernel methods*, preprint, Universität Göttingen, 2003.

[178] R. Schaback and H. Wendland, *Numerical techniques based on radial basis functions*, Curve and Surface Fitting: Saint-Malo 1999, A. Cohen, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville, TN, 2000, 359-374.

[179] R. Schaback and H. Wendland, *Adaptive greedy techniques for approximate solution of large RBF systems*, Numer. Algorithms 24 (2000), 239–254.

[180] R. Schaback and Z. Wu, *Operators on radial functions*, J. Comput. Appl. Math. 73 (1996), 257–270.

[181] I. J. Schoenberg, *Metric spaces and completely monotone functions*, Ann. of Math., 39 (1938), 811–841.

[182] I. J. Schoenberg, *Spline functions and the problem of graduation*, Proc. Nat. Acad. Sci. 52 (1964), 947–950.

[183] L. L. Schumaker, *Spline Functions: Basic Theory*, John Wiley & Sons, New York, 1981 (reprinted by Krieger Publishing 1993).

[184] C. Shannon, *Communication in the presence of noise*, Proc. IRE 37 (1949), 10–21.

[185] D. Shepard, *A two dimensional interpolation function for irregularly spaced data*, Proc. 23rd Nat. Conf. ACM 1968, 517–523.

[186] R. Sibson and G. Stone, *Computation of thin-plate splines*, SIAM J. Sci. Statist. Comput., 12 (1991), 1304–1313.

[187] N. Sivakumar and J. D. Ward, *On the least squares fit by radial functions to multidimensional scattered data*, Numer. Math., 65 (1993), 219–243.

[188] B. F. Smith, P. Bjørstad, and W. D. Gropp, *Domain Decomposition. Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, 1996.

[189] K. T. Smith, D. C. Solmon and S. L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. (83) 1977, 1227–1270.

[190] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis in Euclidean Spaces*, Princeton University Press, Princeton, 1971.

[191] J. Stewart, *Positive definite functions and generalizations, an historical survey*, Rocky Mountain J. Math., 6 (1976), 409–434.

[192] X. Sun, *Scattered Hermite interpolation using radial basis functions*, Linear Algebra Appl., 207 (1994), 135–146.

[193] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.

[194] M. Unser, *Sampling – 50 years after Shannon*, Proc. IEEE 88 (2000), 569–587.

[195] G. Wahba, *Convergence rate of "thin plate" smoothing splines when the data are noisy (preliminary report)*, Springer Lecture Notes in Math., 757, 1979, 233–245.

[196] G. Wahba and Z. Luo, *Smoothing spline ANOVA fits for very large, nearly regular data sets, with application to historical global climate data*, in The Heritage of P. L. Chebyshev: a Festschrift in honor of the 70th birthday of T. J. Rivlin, Ann. Numer. Math. 4 (1997), no. 1-4, 579–597.

[197] A. Wald, *Statistical Decision Functions* John Wiley & Sons, New York, 1950 (reprinted by Chelsea Publishing 1971).

[198] G. S. Watson, *Smooth regression analysis*, Sankhya, Ser. A 26 (1964), 359–372.

[199] H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Adv. in Comput. Math., 4 (1995), 389–396.

[200] H. Wendland, *Error estimates for interpolation by compactly supported radial basis functions of minimal degree*, J. Approx. Theory (93) 1998, 258–272.

[201] H. Wendland, *Sobolev-type error estimates for interpolation by radial basis functions*, in Surface Fitting and Multiresolution Methods, A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), Vanderbilt University Press, Nashville TN, 1997, 337–344.

[202] H. Wendland, *Meshless Galerkin approximation using radial basis functions*, Math. Comp. 68 (1999), 1521–1531.

[203] H. Wendland, *Numerical solution of variational problems by radial basis functions*, in Approximation Theory IX, Vol.II: Computational Aspects, Charles K. Chui, and L. L. Schumaker (eds.), Vanderbilt University Press, 1999, 361–368.

[204] H. Wendland, *Local polynomial reproduction and moving least squares approximation*, IMA J. Numer. Anal. 21 (2001), no. 1, 285–300.

[205] H. Wendland, *Scattered Data Modelling by Radial and Related Functions*, Cambridge University Press, Cambridge, to appear.

[206] J. M. Whittaker, *On the functions which are represented by expansions of the interpolation theory*, Proc. Roy. Soc. Edinburgh 35 (1915), 181–194.

[207] E. T. Whittaker, *On a new method of graduation*, Proc. Edinburgh Math. Soc. 41 (1923), 63–75.

[208] D. V. Widder, *The Laplace Transform*, Princeton University Press, Princeton, 1941.

[209] S. M. Wong, Y. C. Hon and T. S. Li, *Radial basis functions with compactly support and multizone decomposition: applications to environmental modelling*, in Boundary Element Technology XIII, C. S. Chen, C. A. Brebbia, and D. W. Pepper (eds.), WIT Press, 1999, 355–364.

[210] W. S. B. Woolhouse, *Explanation of a new method of adjusting mortality tables, with some observations upon Mr. Makeham's modification of Gompertz's theory*, J. Inst. Act. 15 (1870), 389–410.

[211] Z. Wu, *Hermite-Birkhoff interpolation of scattered data by radial basis functions*, Approx. Theory Appl., 8 (1992), 1–10.

[212] Z. Wu, *Compactly supported positive definite radial functions*, Adv. in Comput. Math., 4 (1995), 283–292.

[213] Z. Wu and R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal., 13 (1993), 13–27.

[214] J. Yoon, *L_p-error estimates for "shifted" surface spline interpolation on Sobolev space*, Math. Comp. 72 (2003), 1349–1367.

[215] L. Zhang, W. K. Liu, S. F. Li, D. Qian and S. Hao, *Survey of Multi-scale Meshfree Particle Methods*, in Lecture Notes in Computer Science and Engineering Vol.26: Meshfree Methods for Partial Differential Equations, M. Griebel and M. A. Schweitzer (eds.), Springer Verlag, 2002, 441–458.